Discovery and broad relevance may be insignificant components of course-based undergraduate research experiences (CUREs) for non-biology majors

Authors: Cissy J. Ballen[1], Seth K. Thompson[1,2], Jessamina E. Blum[1], Nicholas P. Newstrom[3], Sehoya Cotner[1]
[1]Department of Biology Teaching and Learning, University of Minnesota, Minneapolis, MN 55455
[2]Department of Water Resources, University of Minnesota, Minneapolis, MN 55455
[3]Program in Human Sexuality, University of Minnesota, Minneapolis, MN 55455
*Balle027@umn.edu

Manuscript published as Research Article in the *Journal of Microbiology & Biology Education (JMBE)*

Running title: CUREs for non-biology majors

Abstract

Course-based undergraduate research experiences (CUREs) are a type of laboratory learning environment associated with a science course in which undergraduates participate in novel research. According to Auchinchloss *et al.* (2104), CUREs are distinct from other laboratory learning environments because they possess five core design components, and while national calls to improve STEM education have led to an increase in CURE programs nationally, less work has specifically focused on which core components are critical to achieving desired student outcomes. Here we use a backward elimination experimental design in order to test the importance of two CURE components for a population of non-biology majors: the experience of discovery and the production of data broadly relevant to the scientific or local community. We found nonsignificant impacts of either laboratory component on students' academic performance, science self-efficacy, sense of project ownership, and perceived value of the laboratory experience. Our results challenge the assumption that all core components of CUREs are essential to achieve positive student outcomes when applied at scale.

Introduction

Engaging undergraduate science students in research experiences has a number of important benefits (1, 2). However, the traditional "apprenticeship" model of undergraduate research, in which a highly motivated student works as part of a faculty member's research team, is typically restricted to a subset of developing scientists. A relatively recent approach for providing undergraduate students with opportunities to conduct research is the *course-based undergraduate research experience* (CURE) (3, 4). CUREs are scalable research experiences capable of reaching large numbers of students by involving entire courses in a research question within the context of the course itself. This structure provides research experiences to students who would not otherwise participate in more traditional research, such as students in non-biology majors (hereafter 'nonmajors'). For these students, a one-semester laboratory course may be the only formal scientific or research training they experience in college. CUREs offer opportunities for these students to gain valuable experience while also meeting course requirements (5, 6). CUREs can vary in their duration, setting, extent of mentoring, and cost depending upon the logistical restrains of the institution (7).

According to Auchincloss *et al.* (2014), several core components define a CURE. These core principles include (1) cycles of *iterative experimentation* followed by critical evaluation of data; (2) *collaborative work* with other students and/or the course instructor in order to address complex problems; (3) use of *scientific practice* through engagement with science investigations; (4) experience of *discovery* as students work on a novel question to arrive at a conclusion unknown to the student, instructor, and broader scientific community; (5) production of data *broadly relevant* to the scientific or local community (Figure 1).

While identifying these five core components provides a useful framework for thinking about the design and implementation of CUREs, there is little empirical evaluation of the importance of individual core components. In other words, we lack an understanding of whether each of the components relates to positive student outcomes, e.g., competencies, student attitudes, or retention in the discipline. Implementing all components simultaneously can be resource-intensive or difficult to facilitate and maintain in a classroom or laboratory setting over time, limiting the scalability of

CUREs for some institutions. Therefore, it is essential that we justify the utility of these design features in a variety of contexts. Empirical validation of each component will allow for more efficient course design that maximizes the impact of course-based research for all students, and contribute new scientific knowledge to the scientific community (8).

Some of the core components highlighted in L. C. Auchincloss et al. (4) are fairly easy to understand—if not implement. For example, the way students can experience the use of "scientific practice" has been articulated by D. Lopatto (1) and echoed by E. Seymour et al. (9). They include understanding primary literature, designing experiments, collecting and interpreting data, and writing scientifically. "Collaboration" and "iteration" are likewise unambiguous concepts. While the relevance of these hallmarks should also be critically explored, we chose to examine the importance of "Discovery" and "broad relevance" because we believe these are less tangible CURE components.

What is discovery? Discovery in science is the process of obtaining new knowledge, leading to new understanding of the natural world. In many laboratory experiments, students participate in a discovery exercise because the outcome of their investigation is new to them, but within a CURE, the outcome is unknown to both the student and the instructor. This 'discovery with novelty' implies that students have the potential to contribute new knowledge to the field. Thus, establishing this potential, via a careful understanding of the status of the field, is imperative. Arriving at this understanding may require a course facilitator (instructor or teaching assistant) with a solid grasp of the discipline, and an awareness of the boundaries of knowledge. Developing scientific novelty of proposed work is relatively simple when the work involves an area in which the instructor is an expert. However, large courses with multiple lab sections are often taught by graduate students or undergraduates who are not experts in the discipline, and may be unable to judge the novelty of student research proposals. If in fact novelty is critical for obtaining the proclaimed benefits of a CURE, instructors will need to think seriously about creative ways to incorporate this aspect into their courses. In the current study, we collaborated with an expert in the discipline who could steer students towards novel questions.

What is broad relevance? Creating the opportunity for students' work to be broadly relevant requires the involvement of one or more interested parties who exist beyond the classroom. Examples of interested parties include a research laboratory conducting work on a topic relevant to the CURE, a local community who benefits from the results of a CURE, or a publically available database of student results that could further research in the field. Of the CURE elements suggested by L. C. Auchincloss et al. (4), we found that discovery and broad relevance require more logistical considerations than the other elements and are especially difficult to successfully execute in a large-enrollment nonmajors' course.

This work is motivated by the overriding question: Do discovery and broad relevance matter in a laboratory experience geared toward nonmajors? In other words, do students who are working at the edge of scientific knowledge benefit from the novelty aspect of their work, or the fact that someone is interested in their findings?

We addressed these questions via a backward elimination experimental design, which involves some sections of a nonmajors' biology course engaged in the course's capstone CURE with all five core components, and then testing the impact of eliminating one component at a time (Figure 1). We hypothesized that experimental treatments would not influence students' course performance, reported science self-efficacy and project ownership, and qualitative perceptions of the lab experience. We chose to examine self-efficacy because of its power to predict actual performance among students (10-12). We chose to measure self-reported project ownership because of prior demonstrated positive outcomes associated with independent research experiences for undergraduates (9, 13-15). Finally, we wanted to provide students with an opportunity to describe in open ended responses their perceptions of the value of each laboratory experience, on which we performed qualitative analyses. We hypothesized that experimental treatments would not influence students' course performance, reported science self-efficacy and project ownership, and qualitative perceptions of the lab experience. Our results have broad implications for the development of scalable CUREs in university curricula.

Methods

Student population. Our student population included 412 students enrolled in an introductory biology course for nonmajors at the University of Minnesota in Minneapolis, MN. This course, *Biology 1003: The Evolution and Biology of Sex*, has the dubious distinction of being the favored course of the most science-phobic subset of the University's student population (Cotner, unpublished data). Students come from a variety of different academic backgrounds, range from incoming freshmen to graduating seniors, and are diverse with respect to age and racial/ethnic identity (Table 1). To control for the influence of instructor gender on any of the student outcome variables (e.g., 16), the two instructors involved in the courses were both women. The gender of teaching assistants who guided labs varied across treatment groups (Inquiry treatment TAs 75% women; discovery treatment TAs 100% women, CURE treatment TAs 50% women).

Experimental manipulation. This experiment included 18 laboratory sections across three large lecture sections of Biology 1003 in fall 2016. A significant portion of a student's lab grade involved their work on a multi-week, collaborative research project examining an authentic dataset used in collaboration with the University of Minnesota's Program in Human Sexuality (PHS). The laboratory activity, entitled 'Testing Hypotheses about Adolescent Sexual Behavior', occurs over five lab periods that take place once a week, and has students reading and discussing the literature about adolescent sexual behavior. For the learning exercise, we had students (1) observing and interpreting a real, anonymized dataset, (2) developing a hypothesis to test using the dataset, and (3) analyzing the data to test their hypothesis, and (4) presenting the results of their research. The full exercise can be found in University of Minnesota's 'The Evolution and Biology of Sex: Laboratory Investigations' (17). Undergraduate or graduate-student teaching assistants lead the lab sessions of 20-24 students. We split students into one of three treatment groups and trained TAs to guide students through a CURE, discovery-based inquiry, or an inquiry lab as defined below (Figure 1; see also (5)):

> (1) The **CURE treatment group** (*N* = 115 students from 5 laboratory sections in lecture section 01) possessed all core components of a CURE as defined by Auchincloss et al. (4): cycles of iterative experimentation, collaborative work, use of scientific practice, experience of discovery, and dissemination of data broadly relevant to the science community. Specifically, we required that students ask questions not previously addressed in the published literature after reviewing research already conducted with the PHS dataset, and after students presented their findings to the lab section, they emailed their presentations to a researcher at the Program in Human Sexuality (Newstrom). Prior to the onset of the CURE, Newstrom attended lecture section 01 to explain the importance of the research to students, and express his interest in student findings.
>
> (2) In the **discovery-based inquiry treatment group** (*N* = 115 students from 5 laboratory sections in lecture section 02), we required that students undertake *four* out of the five

defining features of a CURE: cycles of iterative experimentation, collaborative work, use of scientific practice, and experience of discovery; we did not require they disseminate data broadly relevant to the science community. Specifically, students asked original questions after reviewing previous research conducted with the PHS dataset. Students presented these results to their lab section, but did *not* work with or disseminate their results to the researcher from the PHS.

(3) In the **inquiry treatment group** (*N* = 182 students from 8 laboratory sections in lecture section 03), we required that students undertake *three* out of the five defining features of a CURE: cycles of iterative experimentation, collaborative work, and use of scientific practice; we did not facilitate an experience of discovery, nor require they disseminate data broadly relevant to the science community. Specifically, students developed and pursued their own research questions about the PHS dataset (without any requirement to ask a novel question); furthermore, students did not interact with or disseminate their results to the researcher from the PHS. Students were assigned readings from the primary literature that highlighted research similar to that conducted on the PHS dataset.

Across treatments, students worked in groups to develop a hypothesis, learn basic statistical analyses (e.g., one-way ANOVA) using the statistical software JMP Pro 12 (SAS Institute Inc., Cary, NC, USA), analyze data, interpret their results, graphically depict their results, and prepare a written and oral presentation of their work. In all treatment groups, students presented their research to their lab section with a powerpoint presentation, but we only required that the CURE treatment group send their presentations to Nicholas Newstrom at the PHS. The laboratory schedule and point allocation within laboratory and lecture can be found in Supplemental Table 1 and Supplemental Table 2.

Data collection and analysis. To test the importance of discovery and broad relevance in nonmajors' CUREs, we conducted quantitative and qualitative analyses. First, we examined the effects of treatments on student performance (lab grade), self-reported confidence in the ability to do science ("science self-efficacy"), and sense of project ownership. We addressed the following questions: 1) Does discovery or broad relevance improve student performance in lab, as compared to an inquiry laboratory that lacks these components? 2) How does discovery or broad relevance impact student science self-efficacy and sense of project ownership?

We performed all statistical analyses using SPSS software version 24 (SPSS Inc., Chicago, IL, USA). We first ran a post-hoc ANOVA to compare incoming student academic preparation among treatments. These only included students who finished the course. Then, we used general linear mixed models to compare student lab achievement (lab grades) and two affective metrics (science self-efficacy and project ownership) across the three treatment groups: CURE, discovery-based inquiry, and inquiry. We evaluated student performance based on total grade in lab because all laboratory sections are evaluated using the same manual and grading rubric. We did not use the research laboratory reports as a performance measure because laboratory teaching assistants graded them out of 8 points, and most students received full credit. All models include the same fixed and random variables, and we included a covariate in the performance model to address the variation in incoming academic preparation (student cumulative GPA). Fixed factors included laboratory treatment group, gender, underrepresented minority (URM) status, and age. We included the laboratory section as a random effect in all analyses. To fit the assumptions of the general linear model, we transformed students' lab grades by taking the linear log of [120 - student grade]. For all Likert scale analyses we treated the dependent variables as continuous for ease of interpretation, given that non-parametric tests have yielded very similar results to the ones reported in this paper (18, 19). Prior to the analysis, we decided that it is unlikely that student characteristics (e.g., age, gender, race/ethnicity) would interact with treatments to influence the statistical outcomes, but included them in all analyses because of their demonstrable effect on some performance outcomes (e.g. for confidence: 16, 20).

Using post-course surveys, we asked to what extent students felt confident comprehending, critically assessing, and communicating scientific concepts. Following Bandura's (12) work on self-efficacy, we modified survey questions from an existing instrument (21) in which students rated confidence in their ability to complete course-relevant tasks. Responses were quantified on a 4-point Likert scale [Assessment 1, supplementary materials (SM)]: 1 = not confident; 2 = slightly confident; 3 = mostly confident; 4 = very confident.

We conducted exploratory factor reduction analyses on the eleven science self-efficacy survey items. We had adequate sampling to produce reliable results according to the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy for the whole dataset (KMO > 0.8). In order to test the presence of relationships between variables we used Bartlett's test of sphericity, which was significant ($P < 0.001$). Post-course surveys generated a single component that explained 58% of the total variance. We tested for internal consistency using Cronbach's alpha, and found these survey items to be correlated (Cronbach's alpha > 0.9). We then generated a single science self-efficacy response variable for each student by combining scores using an additive scale for use in the statistical model. We re-ran the analysis after excluding seven outliers and the results were the same so we include them here.

To examine student sense of ownership and perceptions of the laboratory experiences, we used five survey questions modified from D. I. Hanauer and E. L. Dolan (22); these responses were also quantified on a 5-point Likert-type scale (Assessment 2, SM). To test whether these data were suitable for factor reduction we conducted an exploratory factor analysis. For project ownership, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy for the whole dataset was KMO = 0.833 and Bartlett's test of sphericity was $P < 0.001$. The five survey items generated a single component that explained 61% of the total variance. We tested for internal consistency using Cronbach's alpha, and found them to be highly correlated (Cronbach's alpha > 0.8). In response to these results, we combined measures using an additive scale that represented a comprehensive project ownership score for analyses.

Students took the project ownership survey only once at the end of the course because it was designed to gauge student perceptions over the course of the laboratory experience. Figure 2 displays the student responses to the project ownership survey separately in order to illustrate nuanced results from the survey rather than a broader construct (which is more suitable for analyses with variable reduction).

Students were assured of anonymity during the course, confidentiality after the course, and the ability to omit any of the survey items. The surveys were approved by the University of Minnesota's Institutional Review Board (#1405E50826). Of the 412 possible respondents, we secured post-course surveys from 302 students (73% response rate).

Qualitative analyses. Our second objective was to qualitatively explore, through open ended responses submitted by students, perceptions of the value of each laboratory experience (CURE, discovery-based inquiry, or inquiry; $N = 78$ student responses). After asking a third party to collate and randomize student responses, two of the authors used inductive coding to generate six recurring themes. These included 1) real world applications, 2) choice/ownership/discovery, 3) learned science process skills, 4) learned something new, 5) general interest in the topic, 6) required more guidance (Table 2). From the 78 student responses, we coded 98 different statements that were coded into one of the six constructed themes. We excluded six responses because they were unclear, addressed difficulty with the statistical software, or only expressed their feelings about the lab TA. Any coding disputes were discussed and consensus reached before analysis was done. We analyzed coding data by comparing the relative ratios of each coding theme as a percentage of the total number of coded responses within each treatment group (Cohen's kappa > 70%).

Results

Quantitative results. We tested the effect of different laboratory environments on student performance, student science self-efficacy, and sense of project ownership. First, an ANOVA comparing incoming student academic preparation among treatments revealed nonsignificant differences between student populations (cumulative GPA $P = 0.155$). Within our mixed models, we found no significant effect of laboratory treatment on laboratory grade ($F_{2,14.7} = 2.155$, $P = 0.151$, $N = 411$), while cumulative GPA ($F_{1,397.5} = 168.350$, $P < 0.001$) positively predicted laboratory grade, along with gender ($F_{1,402.9} = 4.950$, $P = 0.027$), with males outperforming female students. These results suggest that there are not statistically significant differences in laboratory performance among students who participate in an inquiry lab, discovery-based inquiry lab, or a CURE.

Next, students' post-course reported science self-efficacy did not differ based on laboratory treatment group ($F_{2,13.2} = 0.008$, $P = 0.992$; $N = 289$; Table 3), suggesting that different treatment groups did not impact students' confidence in their skills related to conducting, communicating, and interpreting science. Note that we only measured science self-efficacy in one post-course survey, and did not examine its change over the course of the semester. We assume that students in the three different treatment groups had roughly equivalent incoming measures. Third, we carried out a similar analysis of student project ownership responses (Figure 2). We found that laboratory treatment group did not significantly affect students' responses to the construct ($F_{2,15.2} = 0.023$, $P = 0.977$; $N = 302$). All other factors in the analyses were also non-significant ($P > 0.15$).

Qualitative results. We categorized 98 themes from 78 open-ended post-course survey responses to questions in which the students were asked to reflect on their laboratory experience. Student comments were categorized into one or more themes based on whether they mentioned the following in their open-ended response: real world application, choice/ownership/discovery, learning science process skills, learning something new, or needing more guidance. Coding showed that students in the inquiry treatment commented on all six themes, whereas students in discovery and CURE treatment groups did not comment on learning science process skills or needing more guidance (Figure 3; Table 4).

Overall, student comments from all the treatment groups were remarkably similar. The similarity of these student comments suggests that there were not large differences in the overall student perceptions of their laboratory experience regardless of the treatment group they were assigned to. It is also interesting that students assigned to the inquiry laboratory commented on the need for more guidance in their labs. This feeling could be due to the limited role of the TA and primary literature in guiding their question and hypothesis creation, which may have made the inquiry exercise feel artificial to the students. However, we are cautious to draw firm conclusions based on the percentages generated from these data because the total number of responses for the inquiry treatment were approximately double the number of responses for the discovery and CURE treatments (Table 4). Therefore, the lack of comments related to science process skills and guidance could be due to limited sampling within the discovery and CURE groups. Nonetheless, based on this analysis we found little evidence that suggests predominant themes emerged from student comments that were unique to any one laboratory treatment group. However, the finding that only students in the inquiry treatment mentioned learning science process skills in their comments is consistent with other research in which students describe the inquiry lab as a ‘skill-building’ opportunity (23).

Discussion

Our data show that discovery and broad relevance have insignificant effects on student performance, science self-efficacy, and sense of project ownership in our population of nonmajors. Instead, students across all laboratory treatment types found personal reliance to be important for determining the value of a research experience. We demonstrate that, for nonmajor students, a course utilizing inquiry approaches may be sufficient to achieve the measured outcomes for a laboratory learning environment. These findings should be relevant to instructors whose desired course outcomes mirror those we document here—course performance, science self-efficacy, and project ownership.

These results highlight the need to empirically evaluate the other design elements of CUREs in order to establish those which contribute to positive student outcomes for both nonmajor and major student populations. For example, if students value the opportunity to choose their own research questions, such autonomy may be in direct conflict with instructors who seek high quality data through directed undergraduate CURE collaborations. Future research will profit from a comparison of the influence of these on various metrics of success in laboratory learning environments. It will also be important to test the generality of our results; for instance, a similar study examining an undergraduate biology majors' population, or a graduate student population, may find stronger preferences for elements of discovery and broad relevance.

The limitations of this research also warrants consideration. CUREs are structured in a variety of ways, the most common of which are described in the CURE Network website (http://curenet.cns.utexas.edu/) and the National Research Council convocation report on discovery-based research courses (http://serc.carleton.edu/NAGTWorkshops/undergraduate_research/strategies.html). These describe more 'traditional' wet-bench research, while students in our lab worked with an established database. We also used data that related to human sexuality, which is a unique topic that might influence students' responses. Finally, we chose three surveys that measure affective qualities and one measure of performance, but a number of different assessment tools would allow us to ask other specific questions to deliberately align teaching goals with practical outcomes. For example, the Test of Scientific Literacy Skills (TOSLS) quantifies student proficiency in using scientific concepts beyond the laboratory setting (24). In addition, one can use the Test of Science-Related Attitudes (TOSRA) to quantify favorable attitudes towards science and scientists (25). However, in the absence of results from these other measures, we conclude that there were no observable differences between laboratory treatment groups. Finally, our attempt to create treatment groups that reflected truly 'broadly relevant' questions and provided students with a sense of 'discovery with novelty' (investigating questions that are new to a field), while controlling for other factors that might influence student outcomes other than these features, were limited by logistical restraints inherent to working within three large introductory classrooms. For instance, we would ideally provide each novice undergraduate with the opportunity to work together and directly with a principal investigator to experience discovery as it is generated by the work conducted within the context of the researcher's agenda. Students who experience this type of hands-on research may have reported different attitudes at the end of the semester. These logistical restraints to our research are the same restraints that will limit sustained efforts to implement research experiences in large introductory courses, particularly at a large public institution.

Critically, it would be naïve to assume that all expert guidance is equal. Few would claim that all primary instructors or teaching assistants are similarly skilled at facilitating inquiry, thus we assume that there exists a range of expert types that vary in communicating scientific facts and enthusiasm for the investigation, and providing helpful feedback. In our CURE treatment, the expert involvement was minimal, meeting once with the students at the beginning of the project to introduce the research and convey interest in the students' work, and then being available to review ideas and final products. An expert who was involved with student projects on a weekly basis may have contributed differently to student outcomes. However, given our interest in scalability, the involvement we document is more practical for multiple sections of nonmajors introductory biology.

Our results may come as a relief to some instructors designing research experiences for nonmajors. Fully implanting the current CURE model; which requires incorporating expert input, providing students with unexplored data, and finding an audience who cares about the results can be time consuming and impractical, especially for large courses with several laboratory sections. Furthermore, it may be difficult to find experts who are as enthused about working with nonmajors as they might be to work with developing scientists and potential future colleagues. Additionally, it is reasonable for instructors to have very different desired learning outcomes for nonmajor students compared to major students and when designing a laboratory experience a reverse design framework where the instructor uses the learning outcomes to determine appropriate student experiences should be applied (19). Instructors must think critically about if the full CURE laboratory is the most appropriate way to achieve the desired learning outcomes for their students. Overall, these findings indicate that instructor efforts to incorporate research into the curricula may not require the additional – and often logistically difficult – steps of providing students with a sense of authentic discovery and broad relevance. Our results challenge the value of CUREs as they are currently defined, and support a call for a deeper understanding of why different laboratory environments are effective for both major and nonmajor student populations.

Acknowledgements

We thank Daniel Baltz and Sadie Hebert for help with data organization and interpretation; Deena Wassenburg for help with data collection from students; Kristina Prescott for TA support and training; J.D. Walker for statistical support; Michael Miller for coordinating the PHS collaboration. This work was supported in part by a National Science Foundation IUSE grant (Integrated Science Education for Discovery in Introductory Biology, Proposal number 1432414), awarded to Sehoya Cotner and Catherine Kirkpatrick. This work was approved by IRB protocol number 1405E50826, and has complied with all relevant institutional guidelines and policies.

References

1. Lopatto D. 2007. Undergraduate research experiences support science career decisions and active learning. CBE-Life Sciences Education 6:297-306.
2. Russell SH, Hancock MP, McCullough J. 2007. Benefits of undergraduate research experiences. Science(Washington) 316:548-549.
3. Bangera G, Brownell SE. 2014. Course-based undergraduate research experiences can make scientific research more inclusive. CBE-Life Sciences Education 13:602-606.
4. Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Lawrie G, McLinn CM, Pelaez N, Rowland S. 2014. Assessment of course-based undergraduate research experiences: a meeting report. CBE-Life Sciences Education 13:29-40.
5. Ballen CJ, Blum JE, Brownell S, Hebert S, Hewlett J, Klein JR, McDonald EA, Monti DL, Nold SC, Slemmons KE. 2017. A Call to Develop Course-Based Undergraduate Research Experiences (CUREs) for Nonmajors Courses. Am Soc Cell Biol.
6. Wei CA, Woodin T. 2011. Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE-Life Sciences Education 10:123-131.
7. Linn MC, Palmer E, Baranger A, Gerard E, Stone E. 2015. Undergraduate research experiences: impacts and opportunities. Science 347:1261757.
8. Cooper KM, Soneral PA, Brownell SE. 2017. Define Your Goals Before You Design a CURE: A Call to Use Backward Design in Planning Course-Based Undergraduate Research Experiences. Journal of microbiology & biology education 18.
9. Seymour E, Hunter AB, Laursen SL, DeAntoni T. 2004. Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-

year study. Science education 88:493-534.

10. Ballen CJ, Wieman C, Salehi S, Searle JB, Zamudio KR. 2017. Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. CBE-Life Sciences Education 16:ar56.
11. Stajkovic AD, Luthans F. 1998. Self-efficacy and work-related performance: A meta-analysis. Psychological bulletin 124:240.
12. Bandura A. 1997. Self-efficacy: The exercise of control. Macmillan.
13. Lopatto D. 2004. Survey of undergraduate research experiences (SURE): First findings. Cell biology education 3:270-277.
14. Hathaway RS, Nagda BA, Gregerman SR. 2002. The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study. Journal of College Student Development 43:614-631.
15. Kremer JF, Bringle RG. 1990. The effects of an intensive research experience on the careers of talented undergraduates. Journal of Research & Development in Education.
16. Cotner S, Ballen C, Brooks DC, Moore R. 2011. Instructor gender and student confidence in the sciences: a need for more role models. Journal of College Science Teaching 40:96-101.
17. Ballen CJ, Newstrom NP. 2017. Testing hypotheses about sexual violence among adolescents. *In* S C, P N (ed), Evolution and Biology of Sex: Laboratory Investigations, 4th ed. Bluedoor, Minnesota.
18. Norman G. 2010. Likert scales, levels of measurement and the “laws” of statistics. Advances in health sciences education 15:625-632.
19. Murray J. 2013. Likert data: what to use, parametric or non-parametric? International Journal of Business and Social Science 4.
20. Fox MF, Firebaugh G. 1992. Confidence in science: The gender gap. Social Science Quarterly.
21. Robnett RD, Chemers MM, Zurbriggen EL. 2015. Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. Journal of Research in Science Teaching 52:847-867.
22. Hanauer DI, Dolan EL. 2014. The Project Ownership Survey: measuring differences in scientific inquiry experiences. CBE-Life Sciences Education 13:149-158.
23. Rowland S, Pedwell R, Lawrie G, Lovie-Toon J, Hung Y. 2016. Do We Need to Design Course-Based Undergraduate Research Experiences for Authenticity? CBE-Life Sciences Education 15:ar79.
24. 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 11:364-377.
25. Fraser BJ. 1978. Development of a test of science-related attitudes. Science Education 62:509-515.

	Iteration	Collaboration	Science Process	Discovery	Broad Relevance
CURE	★	★	★	★	★
Discovery-based Inquiry	★	★	★	★	
Inquiry	★	★	★		

Figure 1. Summary of differences and similarities among three laboratory learning environments (described in 5). Specifically, CUREs possesses all five core components; discovery-based labs lack broad relevance; inquiry labs lack discovery and broad relevance. We used a backward elimination experimental design in order to test the importance of one or more CURE components for student success.

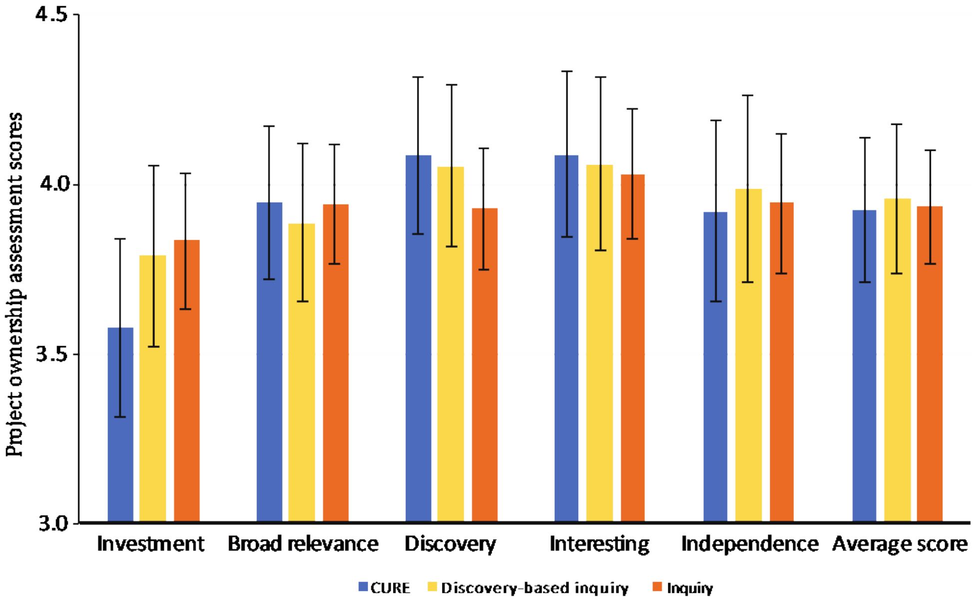

Figure 2. Mean scores (95% C. I.) reported by students on project ownership survey items (Assessment 2) do not significantly differ across CURE (blue), discovery-based inquiry (yellow), and inquiry (orange) laboratory treatment groups (*N* = 302). The survey gauged to what extent students felt invested in the project ('Investment'), agreed that work on their project was broadly relevant beyond the classroom ('Broad relevance'), that there was the potential to discover something new ('Discovery'), that their research project was interesting ('Interesting'), and that they were responsible for the outcomes of the project ('Independence'). For all posthoc analyses of individual survey items, P > 0.15.

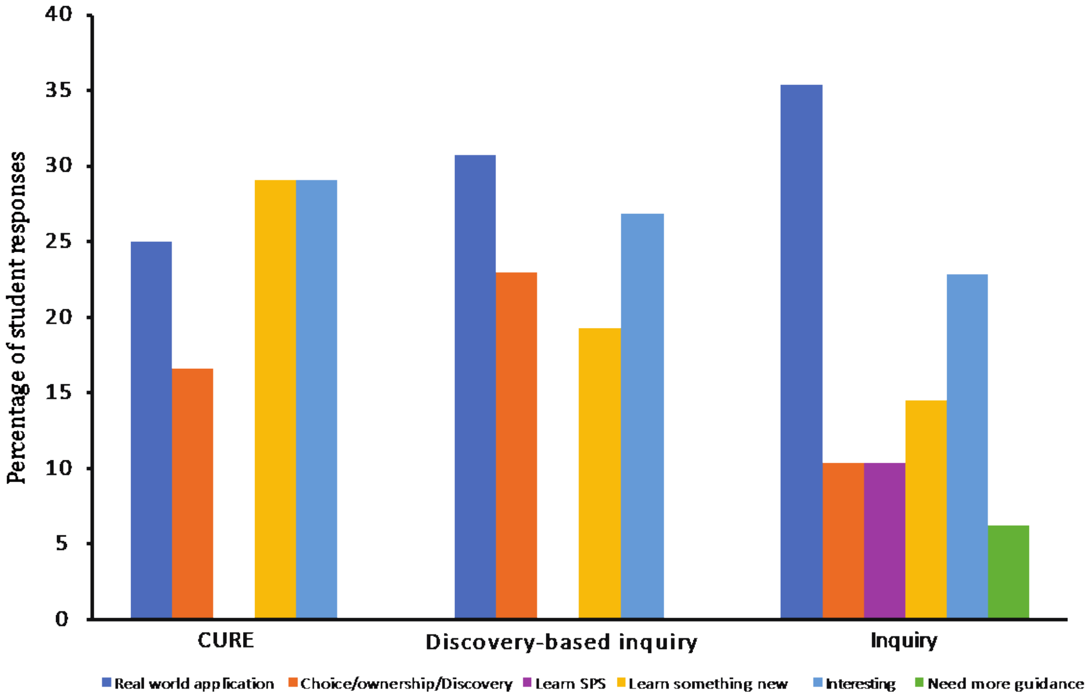

Figure 3. Percentages of binned themes from open-ended responses by students about one of three laboratory experiences (CURE, discovery-based inquiry, or inquiry). We categorized responses based on whether students emphasized real world application (dark blue), choice/ownership/discovery (orange), learning science process skills (SPS; purple), learning something new (yellow), or needing more guidance (green) in their answers.

Table 1. Student demographic information (%) across three laboratory treatments in introductory biology at the University of Minnesota (*N* = 412).

		CURE (*N* = 115)	Discovery-based (*N* = 115)	Inquiry (*N* = 182)
Year in school	1st year	7.0	5.2	5.5
	2nd year	43.5	47.8	54.9
	3rd year	25.2	27.0	21.4
	4th year	24.3	20.0	18.2
Race/ethnicity	American Indian	0.9	0.0	2.2
	Asian American	6.1	10.4	7.1
	African American	6.1	4.3	2.3
	Hawaiian	1.7	0.0	0.5
	Hispanic	3.5	4.3	4.4
	International	16.5	14.0	14.3
	White	65.2	67.0	69.2
Gender	Female	61.7	61.7	59.9
	Male	38.3	38.3	40.1
College	Other STEM	8.7	6.9	7.1
	Non-STEM	91.3	93.1	92.9

Table 2. Student-reported views in response to the open-ended question, "Please comment on any aspect of you research project. Was it a valuable experience? What could your instructor or TA have done differently to help you make the most of your research experience?" We categorized 98 comments into one of the six constructed themes, and provide example statements below.

Response category	**Guide to coding responses**	**Example**
Real World Applications	Words like "useful", "outside connections", "relevant", "real world", "relate", "helpful", makes connections from project to the outside world	"This was my favorite experiment because it could be related to the real world and an overall big picture. It was also a topic that isn't usually covered in a classroom setting so it was a new topic for almost everyone."
Choice, ownership, or discovery	Ability to choose question/project/topic, expresses ownership of project/direction, or discovery of something for themselves	"I thought it was really cool to find our own relationship and think about the factors that contribute. I liked being able to pick what I wanted."
Learn science process skills	Learning science process skills, or "how science is done"	"I think the final research project was an incredible way to cap off the semester, and we were able to use the things we learned throughout the course to come up with a hypothesis, test it, and make educated conclusions."
Learn something new	Learning something new; not related to science process skills	"It was fun looking through all of the information and learning about different aspects that affect adolescent sexual behavior."
Learn something interesting	Mentions that the project was "interesting", or wanted to "know the answer" to their question	"It was an excellent learning experience and we discovered a lot of interesting data."
More guidance	Mentions needing more guidance on question/topic selection	"I should have chosen a more interesting subject. My subject we predicted and got it right easily."

Table 3. Itemized means (SD) of science self-efficacy measures reveal no significant differences between treatment groups (all P > 0.15).

Please rate your level of confidence:	**CURE (N = 84)**	**DISCOVERY (N = 71)**	**INQUIRY (N = 140)**
Understand and evaluate scientific literature	3.04 (0.783)	3.06 (0.735)	2.99 (0.715)
Analyze a set of observations tables, or graphs to identify possible patterns	2.63 (0.788)	2.55 (0.713)	2.65 (0.719)
Pose questions about the observations that can be answered with an experiment	2.99 (0.799)	2.96 (0.726)	2.90 (0.733)
Develop a hypothesis related to a question that has been posed	2.96 (0.719)	3.01 (0.707)	2.94 (0.702)
Design a well-controlled experiment to test a hypothesis	3.05 (0.731)	3.14 (0.723)	3.06 (0.702)
Make predictions about the results I could get from an experiment	2.71 (0.769)	2.80 (0.786)	2.80 (0.741)
Collect, organize, and display the results of an experiment	3.11 (0.712)	3.13 (0.716)	3.06 (0.679)
Use statistics or other appropriate methods to analyze data	3.18 (0.779)	3.23 (0.778)	3.10 (0.720)
Draw conclusions about a hypothesis based on the results of the experiment	2.89 (0.870)	2.89 (0.854)	2.92 (0.720)
Explain an experiment, the results, and analysis orally	3.00 (0.760)	3.11 (0.747)	3.14 (0.637)
Explain an experiment, the results, and analysis in writing	3.01 (0.829)	3.07 (0.743)	3.03 (0.739)

Table 4. Percentages of binned themes from open-ended responses by students about one of three laboratory experiences (CURE, discovery-based inquiry, or inquiry).

	Real World Applications	**Choice/ownership/ discovery**	**Learn SPS**	**Learning something new**	**Interesting**	**More guidance**	**N themes from student responses**	**N Students**
CURE	25.0%	16.7%	---	29.2%	29.2%	---	24	21
Discovery-based	30.8%	23.1%	---	19.2%	26.9%	---	26	19
Inquiry	35.4%	10.4%	10.4%	14.6%	22.9%	6.3%	48	38