To the Editor,

We read with interest the recently published paper from Ballen and colleagues (1), who indicated that discovery and relevance may be insignificant components of course-based undergraduate research experiences (CUREs) for nonbiology majors. We appreciate the authors for addressing this important topic because there is a need for empirical evidence to shed light on design features that are hypothesized to make CUREs distinctive as learning experiences (2). We write this letter in this spirit of collegial exchange and with the aim of promoting further investigation on how CURE instruction influences the learning and development of diverse students. In doing so, we also respectfully submit our opinion that the claims made by Ballen and colleagues overstate the evidence they present and potentially mislead readers in judging the importance of discovery and relevance to CUREs.

Of most concern is what we consider to be a basic validity flaw in the design of Ballen et al.’s comparison groups. The study is described as a “backward elimination experimental design,” which utilizes what are assumed to be critical differences in three educational experiences, termed: 1. CURE treatment group, 2. Discovery-based inquiry group, and 3. Inquiry-based treatment group. According to the authors, the difference between these groups is the presence of components hypothesized by Auchincloss et al. (2), specifically, the inclusion or exclusion of the “experience of discovery” and the inclusion or exclusion of “dissemination of data broadly relevant to the scientific community” (i.e., relevance). According to Ballen et al., the CURE group had all the components specified by Auchincloss et al. and the Discovery group had all the components except relevance. The third group was a form of control and did not have either the experience of discovery or relevance. As such, the validity of the design and all subsequent results and conclusions is based on the discovery and relevance variables.

The basic validity problem is in the definition of discovery and relevance and thus the definition of a CURE. Ballen and colleagues specified that the experience of discovery consisted of students constructing and addressing their own research questions “not asked before in the literature” during five lab sessions, based on the data they received from the Program in Human Sexuality on campus. We are skeptical that five sessions worth of work with limited understanding of or access to the scientific field actually constitutes scientific discovery. Ballen and colleagues further specified that relevance consisted of students e-mailing their final “presentations to a researcher at the Program in Human Sexuality.” We are similarly skeptical that e-mailing a presentation to a professor at one’s own institution constitutes relevance to a larger community. Other CUREs have engaged students in making discoveries that are documented in products relevant to stakeholders in the research, such as journal articles, conference presentations, database entries, and community reports (3–9). Furthermore, Shaffer and colleagues have demonstrated that sustained engagement in research (>36 hours) is necessary for students to fully realize certain benefits of these experiences (10).

The question as to whether or not the three groups varied in their levels of discovery or relevance and whether or not the CURE group included both discovery and relevance could have been addressed empirically. There are existing measures that Ballen and colleagues could have used to demonstrate that the experiences of the three groups are indeed distinct. For example, we (LAC, ELD) have developed and validated a survey measure of relevance and discovery in undergraduate lab courses (11), and responses to this instrument have been shown to distinguish between different types of lab courses in multiple studies (11, 12). Ballen and colleagues could have measured these variables in other ways, but such measures would need to be accompanied by evidence of their validity and reliability in order to draw meaningful conclusions (13–15). Regardless of the measures used, Ballen and colleagues should also have measured the relevance and discovery variables in each laboratory section to demonstrate that their interventions were implemented with fidelity and differed as intended (16). Without these data, and in light of the limitations inherent in the study design, we have no idea what the manipulations of the groups actually mean. Further evidence of this validity problem is apparent in Figure 2 of their article, in which students in all three treatments rate the level of discovery and relevance similarly. This result suggests the treatments may not have been meaningfully different, which is what we would predict from the descriptions of the groups.

Another substantive validity concern is that the study relies only on single items to make inferences about the extent to which students develop a sense of ownership over their coursework. This is problematic from a measurement
perspective (13, 14), similar to measuring the quality of “fun” at an amusement park based on a single ride. In order to measure a latent variable such as students’ level of ownership, there is widespread agreement in the education measurement literature that multiple items—worded slightly differently yet each getting at “ownership”—should be used in order to best assess the construct of interest (6). Furthermore, while the survey items were adapted from an established measure, the Project Ownership Survey (POS) (17), some of the items used by Ballen and colleagues are unrelated to the items in the POS, which raises questions about what the responses on those items actually mean. In general, any selective or modified uses of items from existing measures should be accompanied by validity evidence (13), which Ballen and colleagues did not provide. Because of these validity issues, it seems impossible to examine how student engagement in discovery and dissemination relates to their project ownership, as the study purports to do. We (LAC, ELD) have examined these relationships in our work and find that, at least for science majors, students who have opportunities to make relevant discoveries in their lab courses are more likely to develop a sense of project ownership, but that opportunities to engage in iterative work, such as repeating experiments, troubleshooting, and problem-solving, may be more influential than opportunities to make discoveries (12).

Our final concern is more theoretical in nature and relates to the overall study design. CURE instruction has grown rapidly in response to national calls for transforming undergraduate STEM education (18, 19). This growth has the potential to outpace careful, local-level thinking about whether, why, and how to teach CUREs. For instance, CUREs could be integrated into introductory biology courses to introduce students to research and help them decide whether to pursue additional research experiences or a career involving research; existing theories related to career development are useful for examining this (e.g., 20). Alternatively, CUREs could be integrated into general education courses with the aim of familiarizing students with the practice and nature of science, and for this examining this situation, existing theories related to student understanding of the nature and practice of science are useful (e.g., 21, 22). Similarly, we and others have offered guidance on how to align CURE instruction with educational goals and explained different theories that may help to elucidate how CUREs function (23, 24). The paper by Ballen and colleagues did not present their rationale (i.e., a theory) for the design of the courses in the study. It was unclear to us why they expected the outcomes they studied (e.g., course grades) to differ between the different treatments, or why a lack of effect on course grades would lead to the conclusion that relevance and discovery are “insignificant” elements of course design for the population in question. The study from Ballen et al. was also limited to one course at one institution with one particular population of students (i.e., nonbiology majors).

Therefore, we respectfully disagree with the authors’ conclusion that the “results have broad implications for the development of scalable CUREs in university curricula.” Nevertheless, the authors are asking an important and timely question: whether and how the course design features hypothesized to make CUREs distinctive as learning environments relate to desirable student outcomes. We encourage the community to test, refine, and even refute the framework we have proposed (2). To be informative, however, such studies must address both related and contradictory research, and they must attend to the theory behind CUREs as an intervention, measure the implementation of any interventions, and demonstrate the validity and reliability of the measures.

Sincerely,

Lisa A. Corwin
Ecology & Evolutionary Biology
University of Colorado Boulder
Boulder, CO
lisa.corwin@colorado.edu

Erin L. Dolan
Biochemistry & Molecular Biology
University of Georgia
Athens, GA
eldolan@uga.edu

Mark J. Graham
STEM Program Evaluation and Research Lab
(STEM-PERL)
Ecology & Evolutionary Biology
Yale University
New Haven, CT
mark.graham@yale.edu

David I. Hanauer
Professor of Applied Linguistics/English
Indiana University of Pennsylvania
Indiana, PA
hanauer@iup.edu

Nancy Pelaez
Department of Biological Sciences
Purdue University
West Lafayette, IN
npelaez@purdue.edu

REFERENCES


Volume 19, Number 3

LETTERS TO THE EDITOR


