Assessing Science Training Programs: Structured Undergraduate Research Programs Make a Difference

ALAN E. WILSON, JENNA L. POLLOCK, IAN BILLICK, CARMEN DOMINGO, EDNA G. FERNANDEZ-FIGUEROA, ERIC S. NAGY, TODD D. STEURY, AND ADAM SUMMERS

Training in science, technology, engineering, and mathematics (STEM) is a top priority for driving economic growth and maintaining technological competitiveness. We propose that exposure to a rigorous research program as an undergraduate leads to success in a research STEM career. We compared the scientific outcomes of 88 participants from five National Science Foundation Research Experiences for Undergraduates (REU) Site programs with demographically similar applicants to assess the impact that formal, organized, and funded undergraduate summer research experiences have on participants. Our study demonstrates that REU participants are more likely to pursue a PhD program and generate significantly more valued products, including presentations, publications, and awards, when compared with applicants. We believe that key components of the program include funding for personal and professional needs; access to diverse intellectual, analytical, and field resources; and the presence of other undergraduate researchers who support each other and share their goals and interests.

Keywords: REU, researcher, assessments, field biology, ecology

Scientific, technological, and economic competitiveness is motivating greater interest and investment in science, technology, engineering, and mathematics (STEM) training around the world (Gentile et al. 2017), with an emphasis on addressing the current shortage of STEM PhDs (NAS et al. 2010, Brewer et al. 2011). With annual spending on STEM training well over $14 billion in the United States (DOL 2007), guiding future investments in STEM training demands a good understanding of effective approaches (Lopatto 2004, Wei and Woodin 2011, NSB 2015, Hanauer et al. 2017). For example, undergraduate research experience is often credited with preparing students for success in STEM careers (Kolb and Kolb 2005, DOL 2007, NAS et al. 2010, Graham et al. 2013, Hernandez et al. 2018). However, quantitative assessments of STEM training are rare (Linn et al. 2015, Hanauer et al. 2017) because of a variety of problems, including the difficulty of tracking long-term scientific outcomes in a controlled fashion.

Considering the need to identify effective models for STEM training (Barney 2017), we quantitatively analyzed data from demographically matched students who participated (hereafter *participants*) or applied but did not participate (*applicants*) to the same US National Science Foundation (NSF) Research Experiences for Undergraduates (REU) Site summer program held from 2009 to 2011. These independent and geographically dispersed training programs fully support REU participants to conduct independent research projects. Participants are awarded an NSF-defined “take-home” stipend and travel and housing support. During the fiscal years 2015–2017, NSF REU Site programs across the entire foundation spent more than $185 million on more than 500 grants and trained over 150,000 REU participants (grant data available at *www.nsf.gov/funding/pgm_summ.jsp?pims_id=5517*).

In this study, we used a matched pair sampling design (Faresjö and Faresjö 2010) of 88 participant–applicant pairs of undergraduate students associated from five field-biology- and ecology-based REU Sites supported by the NSF to determine the impact of structured research experience on future STEM productivity (measured as the number of scientific presentations, publications, and merit-based academic awards, as well as the highest degree pursued at the time of tracking; see data available in the supplemental material). Given our interest to best

BioScience XX: 1–6.
doi:10.1093/biosci/biy052

match REU participants and applicants prior to outcome data collection, we used as much provided demographic information as possible. Prior research experience was not considered when matching REU participants and applicants. Some applicants (32%) participated in other research experiences, including other REU programs (8%). Therefore, our results may be viewed as conservative and actually underestimate the impact and value of undergraduate research programs for participants.

Participant and applicant information

Participant and applicant data from five field-based NSF REU Site programs held during summer 2009 (three Sites, 23 student pairs), 2010 (three Sites, 22 student pairs), and 2011 (five Sites, 43 student pairs) were collected from REU grant principal investigators (see supplemental material). A *participant* is defined as a student who was admitted and who successfully completed the program. An *applicant* is defined as a qualified undergraduate student who applied for admission to one of the participating REU Site programs but was not admitted. Each student (participant and applicant) was tracked between 5 and 7 years after their REU experience. To account for this variation in time, paired (participant and applicant) data were treated as a random factor in our statistical analyses. For each REU Site, a demographically similar (e.g., gender, race, ethnicity, age, home institution type—private or public—and size, major, focus area, and grade point average) applicant was paired with an REU participant (see supplemental data set S1).

Measured outcomes

Five individual-specific outcomes were considered in this study, including (1) general field of study (STEM or non-STEM; figure 1a); (2) highest degree pursued (doctorate, PhD; healthcare, e.g., doctor of medicine, doctor of veterinary medicine, doctor of pharmacy; master of science, MS; master of business administration, MBA; bachelor of science, BS; associate of arts, AA; and high school, HS; figure 1b); (3) number of scientific conference presentations; (4) number of publications; and (5) number of academic awards (figure 2; see supplemental material). Information for outcomes was collected using a combination of REU Site principal investigator (PI) tracking data, social media (i.e., LinkedIn and Facebook), scientific databases (i.e., Google Scholar and PubMed), and Internet searches. The identities of each student were confirmed by name, undergraduate institution, and year of graduation. Pairs were included in our analyses only if all data were available for both members of the pair. Publications were carefully matched with REU participants or applicants and counted only if they were (a) published, (b) scientific, and (c) available in searched databases, social-media profiles, or PI reports. In most cases, publications were peer-reviewed journal articles. Undergraduate research-based honors theses were also included. Nonscientific publications, such as fashion blog articles, were not counted. Scientific awards (including grants) associated with merit and related to scientific contributions were counted. Grade point average–related awards, such as dean's list, were not included. Finally, presentations (including oral and poster formats) were counted if they were scientific in nature but did not include formal presentations directly associated with the REU program (such as a final poster symposium).

Statistical analyses

Chi-square and Fisher Exact tests were used to compare the number of REU participants and applicants according to their discipline (STEM or non-STEM; figure 1a) and highest degree pursued when tracked (figure 1b). A generalized linear mixed model fit by maximum likelihood (Laplace approximation) with Poisson distribution and "Pair" as a random factor was used to compare the scientific outcomes (i.e., presentations, publications, or awards) of de-identified, demographically paired NSF REU participants and applicants (the main effect was REU experience). The REU experience effect is interpreted as the multiplicative increase in scientific productivity an REU participant exhibits relative to the demographically similar applicant. Thus, an "REU effect" in which the 95% confidence intervals include 1 indicates that paired students had statistically similar outcomes. An REU effect of more than 1 (lower 95% confidence interval > 1) indicates that an REU applicant was more productive than a paired applicant. Chi-square and generalized linear mixed model statistics were conducted using the base package of R and the *glmer* function in the *lme4* package of R (Bates et al. 2015), respectively.

De-identified data associated with REU participant and applicant pairs are available in the supplemental material. Requests for additional information about this study can be made directly to the corresponding author, Alan Wilson, at wilson@auburn.edu.

Results

As a group, applicants to and participants in NSF REU Site programs are similarly biased toward selecting STEM field careers (figure 1a; chi-square $p = .214$). When we considered all degrees, no statistical difference was observed between the REU participants' and applicants' highest degree types pursued at the time of tracking (figure 1b; Fisher's Exact Test $p = .10$). Given our interest in determining whether REU experiences encourage greater interest in and pursuit of advanced STEM degrees, we conducted a chi-square test using only PhD and MS degree data. In this analysis, we found that REU participants pursued significantly more PhD (+48%) and fewer MS (–45%) degrees than applicants within 6 years after completing their baccalaureate degree (figure 1b, chi-square $p = .018$). The matched paired design means that the positive effect of the REU experience on the pursuit of a PhD is not a function of self-selecting populations because REU participants were matched with demographically similar applicants to the same REU Site. This result alone supports the hypothesis that structured independent

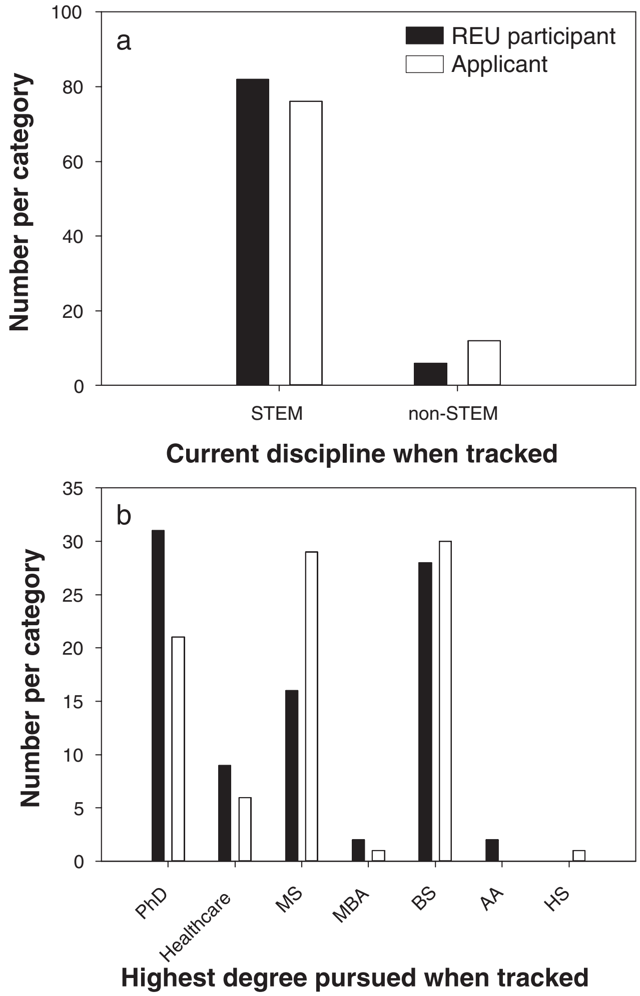

Figure 1. Interests and pursued degrees of REU participants and applicants. Number of students (black bar, participant; white bar, applicant) included in the REU assessment associated with their (a) current discipline (chi-square* p *= .214) and (b) highest degree pursued at the time of being tracked for this study (all degrees Fisher's Exact Test* p *= .10; PhD and MS only chi-square* p *= .018). Degrees included doctor of philosophy (PhD), master of science (MS), and bachelor of science (BS). Healthcare includes all health-related advanced degrees, such as medical doctor.

undergraduate research experience is an important stepping stone to a STEM terminal degree.

In addition, of the REU participants included in our analyses who provided demographic information (gender, 72%; race, 75%), females and underrepresented minorities (including African Americans, Hispanic Americans, and Native Americans) accounted for 59% and 42%, respectively. These demographics were similar to the matched applicant pool based on available data (female, 64%, and/or underrepresented minorities, 44%). However, across the five REU Sites, the most common applicants, including REU participants, were female (68%) and/or Caucasian (84%) based on available data. Therefore, further diversification is generated from the NSF's expectation that REU participants be selected from a broad range of schools, especially minority-serving institutions and institutions with limited research opportunities. Typically, PhD students come from research-intensive public universities or private liberal arts colleges (Fiegener and Proudfoot 2013). Therefore, our results suggest an important broader impact of REU programs: namely, that they serve as a powerful tool for supporting the economic and cultural diversification of PhD-level scientists.

Potentially as a result of an increase in advanced degrees pursued by REU participants, we found that participation in REU Site programs was also effective at boosting research productivity (figure 2). For example, REU participants produced 2.14 times and 1.58 times as many scientific presentations and publications, respectively, and earned 1.37 times as many academic awards than applicants (figure 2d; generalized linear mixed model all $p \leq .012$). Considering that these outcomes are central forms of intellectual currency and indicators of future success in STEM (Laurance et al. 2013, Morales et al. 2017), our findings suggest that there are both short-term (products) and long-term (careers) benefits to participating in NSF REU Site programs. We observed a greater range of products for applicants than REU participants (see supplemental data set S2); however, variance did not scale with the observed data ranges because many applicants had 0 products for a specific scientific outcome.

Although we were not interested in evaluating differences across the five REU Sites included in this study, we conducted additional generalized linear mixed model analyses including REU experience, Site, and REU experience × Site interaction for the three primary outcomes (presentations, publications, and awards) for thoroughness (see supplemental data set S3). In general, our findings from these additional analyses were consistent with our primary analyses (presented in figure 1b). For example, all three of the analyses showed significant REU experience effects (all $p \leq .024$). Moreover, we did find a significant interaction (all $p \leq .0074$) between the REU experience and REU Site program location (any of the five participating programs) for each scientific outcome. However, for all Sites and outcomes except one (publications at Site C; within-Site comparisons results not shown), the estimated effect of the REU experience was positive (albeit not always significantly so). Therefore, the effect of REU experience was estimated to be positive across almost all REU Sites and outcomes, suggesting that the REU experience, in general, drives the patterns we observed despite variation in program location, management, and implementation.

Conclusions

Our quantitative results showing the potential effectiveness of undergraduate research experiences (figure 2) are

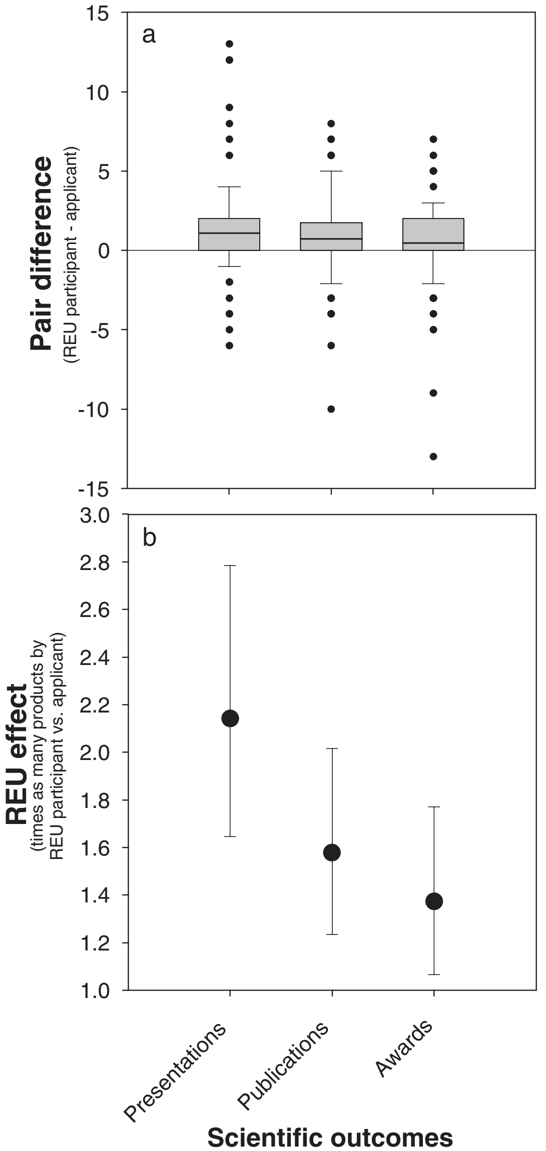

Figure 2. Scientific outcome comparisons of Research Experiences for Undergraduates (REU) participants versus demographically similar applicants. Box plots showing paired response differences (n = 88 pairs; calculated as REU participant–applicant) for three assessed scientific outcomes, including (a) presentations, (b) publications, and (c) awards. A value of 0 would mean that the paired students have the same outcome (e.g., 0–0, 1–1, or 2–2). Box plots show 10% and 90% of the paired difference data (black whisker caps), 25% to 75% of data (gray box), mean (black line in gray box; the median for each outcome was 0), and outliers (black circles outside whisker caps). (d) REU effects (estimate ± 95% CI) for the three scientific outcomes is the multiplicative increase in scientific productivity by an REU picipant relative to a demographically similar applicant (presentations p = .0000000068, publications p = .0002, and awards p = .012). P-values are from a generalized linear mixed model fit by maximum likelihood (Laplace approximation) with Poisson distribution and "Pair" as a random factor.

consistent with earlier qualitative (Lopatto 2004, Linn et al. 2015, Hernandez et al. 2018) and quantitative findings (Hanauer et al. 2017): Structured independent research training is effective at cultivating future scientists. However, an important question remains: Why do these experiences work? (Gentile et al. 2017).

In general, NSF REU Sites provide structured and fully funded research experiences for student cohorts (approximately 8–10 students) for several weeks (approximately 8–10 weeks per summer), during which REU participants collaboratively work with a senior scientist with an active research program and peer scientists while participating in a variety of professional development enrichment activities, such as learning to read research articles, presenting oral or poster presentations, preparing applications for graduate school, and networking with other scientists. Although all or some of these training characteristics could explain our findings that compared demographically matched participants and applicants (Abudayyeh 2003, Taraban and Logue 2012, Auchincloss et al. 2014, Linn et al. 2015, Shanahan et al. 2015, Rocchi et al. 2016, Fox et al. 2017, Morales et al. 2017), it is impossible to completely eliminate potential confounding factors without a controlled, replicated experiment. Therefore, alternative factors beyond the REU experience itself may explain our findings. For example, the participant selection process could bias toward students who are more likely to be successful in science. Although demographic matching does not completely eliminate the possibility that the selection process introduced confounding factors that explained student outcomes rather than the outcomes being generated by the program, that outcome is highly unlikely. REU PIs independently use a variety of data to choose REU participants, including essays, transcripts, letters of recommendation, fit for program and mentors, future aspirations, interview experience, prior research experience, current institution type, and/or demographics. Because each program has its own selection process, it is unlikely that the broad range of selection processes used across Sites would generate a consistent effect of the REU program on participants. Furthermore, given that PIs typically lack any formal human resources training, we would not expect that REU PIs would be more effective at picking more successful participants than groups with formally trained human resources staff. In fact, REU PIs are required to review their REU Site each summer, and part of this process includes reflecting on the students selected to participate in the program. Despite REU PI selection efforts, participants always range widely in performance (see data sets S1 and S2).

Although we cannot completely discount the influence of any confounding factors associated with the participant selection process for an REU Site, NSF REU Sites are not prescribed. Instead, REU PIs have significant flexibility in leveraging available laboratory, analytical, field, and human infrastructure to create the most impactful experiential learning for their REU participants. Each REU PI approaches the selection of their REU participants independently, considering the nature of their REU program, and with NSF's guidance in mind regarding creating opportunities for underrepresented students, students with disabilities, and students from limited research opportunities. Considering the latter (and its influence on future funding for an REU Site), many REU PIs tend to recruit students with limited to no prior research experience. Therefore, despite variation across NSF REU Sites in their research, professional development, and networking activities, we found strong effects of NSF REU experiences (figure 2d). These findings are even more impressive considering that our comparison of paired participants and applicants to the same REU Site did not exclude applicants who conducted other undergraduate research, including participating in other NSF REU Sites or similar programs. Therefore, our analyses may actually underestimate the impact of participating in undergraduate research programs in general. Given the positive impacts of undergraduate research, we argue for continued investment in such programs, including making certain that they are available to all demographic groups (NAS et al. 2010, MacLachlan 2012, Economy et al. 2014, Linn et al. 2015, Hernandez et al. 2018), as a way of maintaining a strong, global STEM workforce.

Acknowledgments
We sincerely appreciate the support of Sally O'Connor, who designed the project and supervised Jenna Pollock, and the Directorate of Biological Sciences' Division of Biological Infrastructure staff at the National Science Foundation for sharing ideas and data associated with our NSF REU impact assessment, as well as the REU Site PIs who shared applicant and participant data and their associated institutions. We would also like to thank the four anonymous reviewers for their suggestions, which substantially improved the final version of this manuscript. This project was supported by the National Science Foundation Internship Program, the Alabama Agricultural Experiment Station, and the Hatch program of the National Institute of Food and Agriculture, US Department of Agriculture.

Supplemental material
Supplementary data are available at *BIOSCI* online.

References cited
Abudayyeh O. 2003. Undergraduate research mentoring model in construction engineering and management. Journal of Construction Engineering and Management 129: 65–69.

Auchincloss LC, et al. 2014. Assessment of course-based undergraduate research experiences: A meeting report. CBE–Life Sciences Education 13: 29–40.

Barney C. 2017. An analysis of funding for the NSF REU Site program in biology from 1987 to 2014. Journal of the Council on Undergraduate Research 1: 11–19.

Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using *lme4*. Journal of Statistical Software 67: 1–48.

Brewer C, et al. 2011. Vision and Change in Undergraduate Biology Education: A Call to Action. American Association for the Advancement of Science.

[DOL] Department of Labor. 2007. The STEM Workforce Challenge: The Role of the Public Workforce System in a National Solution for a Competitive Science, Technology, Engineering, and Mathematics (STEM) Workforce. DOL.

Economy DR, Sharp JL, Martin JP, Kennedy MS. 2014. Factors associated with student decision-making for participation in the Research Experiences for Undergraduates program. International Journal of Engineering Education 30: 1395–1404.

Faresjö, T., Faresjö, Å. 2010. To match or not to match in epidemiological studies—Same outcome but less power. International Journal of Environmental Research and Public Health 7: 325–332.

Fiegener MK, Proudfoot SL. 2013. Baccalaureate Origins of US-Trained S&E Doctorate Recipients in National Center for Science and Engineering Statistics. National Science Foundation.

Fox GA, Kuster EL, Fox AK. 2017. The importance of scientific publishing: Teaching an undergraduate how to swim the entire length of the pool. Journal of Contemporary Water Research and Education 160: 1–4.

Gentile J, Brenner K, Stephens A. 2017. Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities. National Academies Press.

Graham MJ, Frederick J, Byars-Winston A, Hunter AB, Handelsman J. 2013. Increasing persistence of college students in STEM. Science 341: 1455–1456.

Hanauer DI, et al. 2017. An inclusive Research Education Community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. Proceedings of the National Academy of Sciences 114: 13531–13536.

Hernandez PR, Woodcock A, Estrada M, Schultz PW. 2018. Undergraduate research experiences broaden diversity in the scientific workforce. BioScience 68: 204–211.

Kolb AY, Kolb DA. 2005. Learning styles and learning spaces: Enhancing experiential learning in higher education. Academy of Management Learning and Education 4: 193–212.

Laurance WF, Useche DC, Laurance SG, Bradshaw CJA. 2013. Predicting publication success for biologists. BioScience 63: 817–823.

Linn MC, Palmer E, Baranger A, Gerard E, Stone E. 2015. Undergraduate research experiences: Impacts and opportunities. Science 347 (art. 1261757).

Lopatto D. 2004. Survey of undergraduate research experiences (SURE): First findings. Cell Biology Education 3: 270–277.

MacLachlan AJ. 2012. Minority undergraduate programs intended to increase participation in biomedical careers. Mount Sinai Journal of Medicine 79: 769–781.

Morales DX, Grineski SE, Collins TW. 2017. Increasing research productivity in undergraduate research experiences: Exploring predictors of collaborative faculty–student publications. CBE–Life Sciences Education 16 (art. ar42).

[NAS] National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. 2010. Rising above the Gathering Storm, Revisited: Rapidly Approaching Category 5. NAS.

[NSB] National Science Board. 2015. Revisiting the STEM Workforce: A Companion to Science and Engineering Indicators 2014. National Science Foundation.

Rocchi M, Beaudry SG, Anderson C, Pelletier LG. 2016. The perspective of undergraduate research participant pool nonparticipants. Teaching of Psychology 43: 285–293.

Shanahan JO, Ackley-Holbrook E, Hall E, Stewart K, Walkington H. 2015. Ten salient practices of undergraduate research mentors: A review of the literature. Mentoring and Tutoring 23: 359–376.

Taraban R, Logue E. 2012. Academic factors that affect undergraduate research experiences. Journal of Educational Psychology 104: 499–514.

Wei CA, Woodin T. 2011. Undergraduate research experiences in biology: Alternatives to the apprenticeship model. CBE–Life Sciences Education 10: 123–131.

Alan E. Wilson (wilson@auburn.edu) and Todd Steury are associate professors and Edna G. Fernandez-Figueroa is a master's student at Auburn University, in Alabama. AEW is an aquatic ecologist who studies the biotic and abiotic factors that regulate aquatic food webs. TS is a wildlife ecologist who studies carnivore ecology and conservation. EF-F is a National Science Foundation Fellow and past Research Experiences for Undergraduates participant currently investigating consumer adaptation trade-offs to multiple stressors in aquatic ecosystems. Jenna L. Pollock is a Post-Baccalaureate Intramural Research Training Awardee (IRTA) in the National Institutes of Health's Rehabilitation Medicine Department, in Bethesda, Maryland. She studies work-related decisions in patients with Erdheim-Chester Disease. Ian Billick is the executive director of the Rocky Mountain Biological Laboratory, in Crested Butte, Colorado. His focus is place-based research and education. Carmen Domingo is a professor in biology and the interim dean of the College of Science and Engineering at San Francisco State University, in California. CD is a vertebrate developmental biologist as well as a program director for several student training programs focused on increasing the diversity of the STEM workforce. Eric S. Nagy is associate director of Mountain Lake Biological Station at the University of Virginia, in Charlottesville. He is an evolutionary ecologist and directs student training programs, research, education, and outreach at MLBS. Adam Summers is a professor in biology and aquatic and fisheries science at the University of Washington's Friday Harbor Laboratories marine station. The authors spend a lot of time mentoring undergraduate students in research. The authors declare no competing financial interests.