# Insights into Student Gains from Undergraduate Research Using Pre- and Post-Assessments

ANDREW L. MCDEVITT, MANISHA V. PATEL, BRAD ROSE, AND AARON M. ELLISON

*Undergraduate research experiences in science, technology, engineering, and mathematics (STEM) fields expose students to scientific research and are thought to increase student retention in STEM. We developed a pre- and post-survey and administered it to the participants of the Harvard Forest Summer Research Program in Ecology (HF-SRPE) to evaluate the effectiveness of these programmatic goals. Between 2005 and 2015, the survey was sent to all 263 HF-SRPE participants; 79% completed it. Results, controlled for prior experiences, revealed significant improvements across all learning goals. Prior laboratory research experience and perception of being a respected member of a research team were positively associated with gains in research skills and abilities to do and present research. Although the pre- and post-surveys did not indicate changes in the students' goals of pursuing STEM and/or environmental careers, the positive learning gains suggest that students with prior interests in STEM fields take advantage of UREs to solidify further their aspirations in STEM.*

*Keywords: Harvard Forest, retention, STEM, survey, undergraduate research experience (URE)*

**Undergraduate research experiences (UREs) in** science, technology, engineering, and mathematics (STEM) fields (e.g., Research Experience for Undergraduates, REU, Sites) provide students with hands-on experiences in scientific research. For more than a quarter-century, the National Science Foundation (NSF), the Howard Hughes Medical Institute (HHMI), and the National Academy of Sciences (NAS) have promoted UREs as a way of increasing the retention of students and encouraging their pursuit of STEM careers (Lopatto 2004, Harsh et al. 2011). These experiences also are thought to provide a wide range of transferable skills, with underrepresented groups showing the greatest increase in learning gains (Lopatto 2004, 2007). In the biological sciences, a wide range of UREs are available as classroom-based fieldwork (Maw et al. 2011, Scott et al. 2012), research apprenticeships (Sadler et al. 2010), and structured summer research programs (Lopatto 2004, 2007).

The NAS (NRC 2014) promoted summer UREs at biological field stations for a number of reasons. The summer often is the most intensive time for data collection, and many sites rely on undergraduates to collect large quantities of field data (Hodder 2009). Because faculty and other senior investigators in ecology (*sensu lato*) also are focused on research in the summer, UREs at field stations provide students with intensive mentor–mentee interactions, focused research experiences, and work in interdisciplinary research communities (Lopatto 2007, Hodder 2009).

Since its inception in 1985, when a single undergraduate worked on a study of old-growth forests, the Harvard Forest Summer Research Program in Ecology (HF-SRPE) has developed into a thriving and well-coordinated program that is central to the educational and research mission of this combined research department and biological field station. With core support since 1993 from a succession of NSF REU Site awards and NSF REU supplements, and with additional funding from Harvard University, the HF-SRPE has grown to support 20–30 undergraduate students annually (supplementary figure S1). The participating students do research in ecology, soil science, paleoecology, wildlife biology, conservation biology, and atmospheric sciences while being mentored by principal investigators, senior scientists, postdoctoral fellows, and graduate students.

The HF-SRPE includes a range of projects both within and among years. Although individual students' experiences are completed under the umbrella of the HF-SRPE, each one requires a select set of skills and knowledge base that is relevant to a student's particular project. The hiring of summer research students, the establishment of research goals, and project supervision all are overseen by an individual mentor or research group. At the same time, the HF-SRPE overall introduces students to a broad interdisciplinary research

*BioScience* 66: 1070–1078. 
doi:10.1093/biosci/biw141
Advance Access publication 26 October 2016

community and illustrates where their individual projects fit into a broader ecological context.

The HF-SRPE also provides across-the-board support to students and mentors that promotes five programmatic goals: (1) enhancing the ability of students to undertake high-quality interdisciplinary research, (2) building teams of researchers in which students bring different strengths to the table, (3) facilitating collaboration on cutting-edge projects while students find their own intellectual "voices," (4) encouraging students to link fundamental and applied issues in their research, (5) and cultivating the next generation of ecological scientists and educators that reflects the diversity of backgrounds and experiences of students in the United States. Student research projects are structured both to work toward meeting these five student-centered goals and to make substantial contributions to the broad range of ongoing and long-term scientific research at Harvard Forest.

Since 2005, the NSF has emphasized the use of project evaluations to measure, both qualitatively and quantitatively, the success of REU programs (NSF 2005); participant tracking for STEM employment and matriculation has been required for UREs supported by the all NSF directorates since the implementation of the America COMPETES Act of 2010 (42 USC 6621: Coordination of Federal STEM Education). The initial objective of these project evaluations was to determine whether student learning and other measurable outcomes were aligned with specific programmatic goals of individual UREs and of NSF. Lopatto (2004) had previously examined the ability of summer UREs to attract and retain students—especially those from groups otherwise underrepresented in STEM—in STEM careers. However, virtually all of the UREs assessed by Lopatto (2004) using the Survey of Undergraduate Research Experiences (SURE) study were focused on biomedical research and funded by HHMI.

A comprehensive assessment tool for NSF-supported REU programs in biology—the Undergraduate Research Student Self-Assessment (URSSA) survey—was implemented in 2010 (Hunter et al. 2009). The standard implementation of URSSA provides data to NSF on how well REU programs meet national programmatic benchmarks, but it is limited to a single post-program assessment and cannot measure changes in student learning or other programmatic goals resulting from a student's participation in a URE (Frechtling 2002, Hunter et al. 2009). Because the students who participate in UREs have a range of different backgrounds and prior skills in scientific research, it is also important to determine how these factors can influence the success of any URE.

Since 2005, the HF-SRPE has used a pre- and post-survey to measure changes in student learning, skills, and attainment of its programmatic goals. We also have used data from initial surveys to determine how the students' backgrounds and prior research experiences influence their self-reported changes in meeting our programmatic goals and in their educational and career goals. With the pre- and post-data conditioned on their background information as revealed by initial surveys, we addressed three specific questions: (1) To what level is HF-SRPE reaching its educational goals? (2) Which prior experiences predict the greatest gains in the students' perceptions of their research ability? (3) Is the HF-SRPE increasing student interest in STEM (including environmental) careers?

## Questionnaire

In 2005, we developed a set of surveys (the "instrument") to evaluate systematically the experiences and persistence in STEM or environmental education and the careers of the participants in the HF-SRPE (questions and de-identified or anonymized data available in Ellison 2016). We deployed this locally designed instrument to assess critical changes in the students' attitudes toward science; the students' identification with the norms and professional practices of scientific research; the specific skills associated with conducting and disseminating scientific research; and post-program career and educational plans. The questions were reviewed by program administrators for face validity, a subjective confirmation that the measurements are appropriate, and alignment with NSF-REU objectives. Annual pre- and post-program evaluations invited the student participants to report significant changes in these domains and offered the researchers indicators of the changes that the HF-SRPE program designers sought to affect in the program participants. The instrument was designed deliberately for rapid application, ease of program-participant use, and economy of data analysis, and it was administered with a minimum of obtrusiveness.

Data were collected from the students three times during the summer program. First, the students completed a short survey on their arrival at HF-SRPE to determine how they were recruited, their expectations of the program, and their educational and occupational aspirations. Second, the students were surveyed in midsummer. The questions on the midsummer survey probed whether the program was meeting their expectations; their satisfaction with their independent research, their mentors, and field trips; their interactions with scientists, staff, and other student participants; and changes that could improve their experiences. At the conclusion of the summer, the students completed a third survey containing follow-ups to the questions in the first and second surveys. These three surveys were supplemented with individual, semistructured interviews to examine the students' survey responses and to provide them with an opportunity to discuss in detail their experiences and specific aspects of the program. This design allowed the evaluator to explore new topics that arose during the interviews and to follow up on compelling responses (Neuman 2003). The interviews explored in more detail the students' relationships with their mentors; how, if at all, their educational and occupational aspirations had changed; and their perceptions of science in general and the field of ecology in particular.

After reviewing the 2005 pilot study, we reduced the yearly assessment of participant experiences to a single

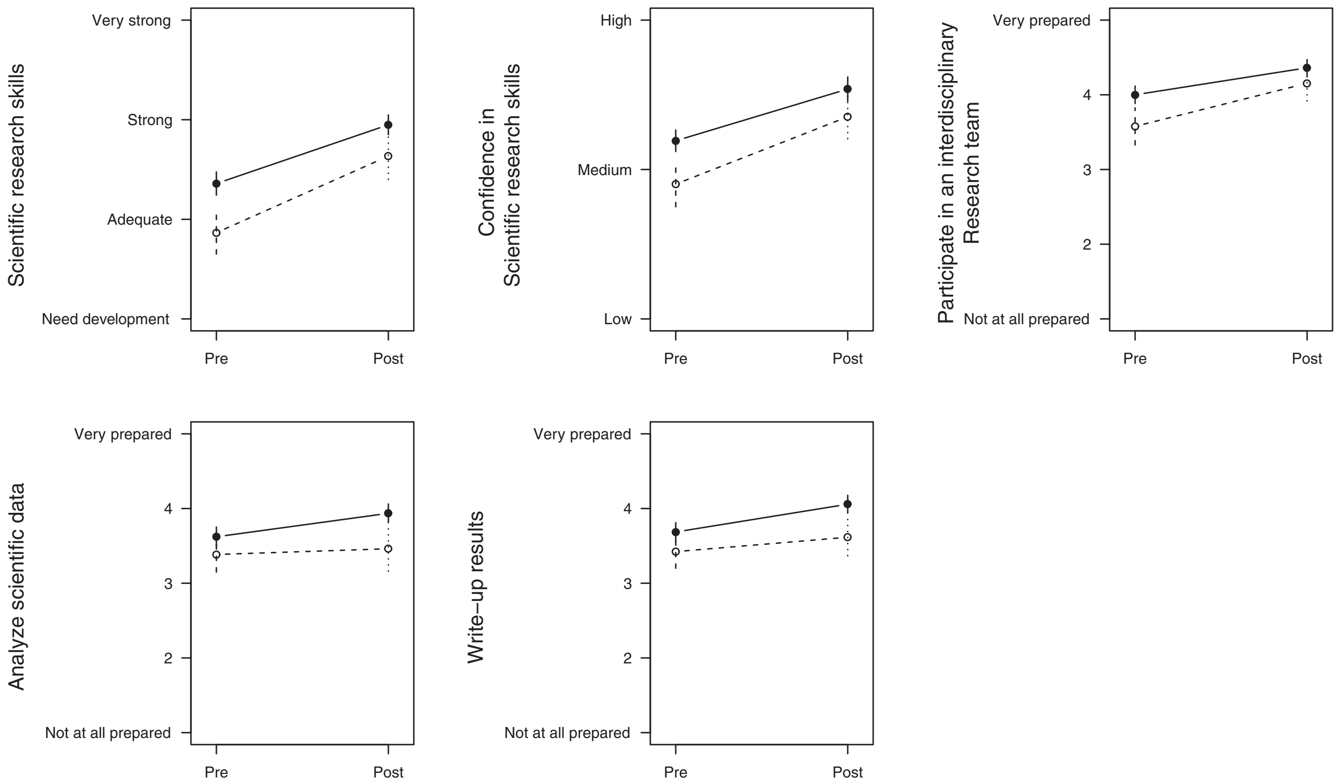


*Figure 1. Mean changes in learning gains based on the presence (closed symbols) or absence (open symbols) of prior laboratory experience. The error bars illustrate 95% confidence intervals.*

pre–post survey according to NSF guidelines for REU Site evaluation (Frechtling 2002). This survey consisted of 22 multiple-choice questions and 2 open-response questions. The pre-program survey was sent to a total of 263 HF-SRPE participants, with a 91.6% response rate. A similar post-program survey was sent to the participants at the end of HF-SRPE, with 79.5% of the individuals responding to both surveys.

## Data analyses

We used repeated-measures analysis of variance (rm-ANOVA) to test for changes in the students' perceived research skills and their confidence in them; the student responses to nine prior experiences (table 1) entered the rm-ANOVA as fixed factors. Multicollinearity (Graham 2003) among the nine prior experiences was assessed using multiple correspondence analysis (MCA) in the *FactoMineR* packages, and subsequent analyses were done using the *car* package, both in R version 3.2.3 (R Development Core Team 2015). Because responses were ordinal, quasi-Poisson linear models were used to model responses; the significance of each term was assessed using Type III sum of squares. *Post hoc* Tukey tests were done only on statistically significant ($\alpha = 0.05$) terms. See Gotelli and Ellison (2012) for additional details on ANOVA, Poisson error terms, and different types of sums of squares.

We used correspondence analysis, using the *ca* package in R, to examine whether the HF-SRPE influenced the participants' long-term career goals. One-sided paired *t*-tests were used to evaluate the self-assessed likelihood that the participants persisted in environmental or STEM research fields.

Anonymized raw data and associated R code are available from the Harvard Forest Data Archive, data set no. HF-279 (Ellison 2016)

## Student background

The students responding to the pre-program survey between 2006 and 2015 ($n = 241$) came into the program with varied backgrounds. Most reported prior experience in laboratory (75%) or field (71%) research, often (52%) on research teams outside of a class. The majority of these students (72%) had worked with a more experienced researcher, but fewer had presented their research to peers (48%) or (co-)authored a scientific paper (7%). Although only 24% of the participants felt that they had contributed previously to the production of a scientific paper, 54% felt that they had been a respected member of their scientific research team. We report key findings in the main text and in figures 1–4. The complete results for all the questions are provided in supplementary figures S2–S9.

## Research skills

The students' perception of their research skills was higher after they had completed the HF-SRPE (table 1, questions 3 and 4). Both prior experience in laboratory research

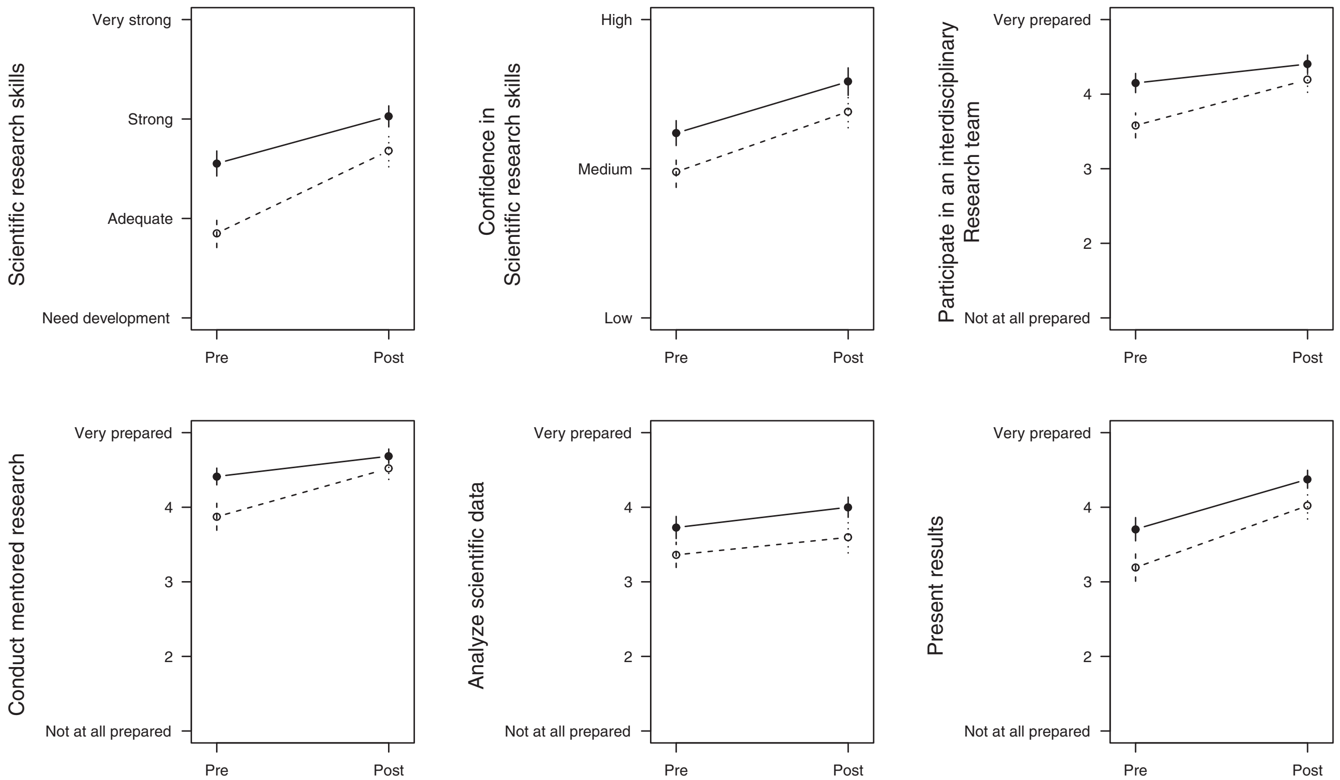


***Figure 2. Mean changes in learning gains based on the presence (closed symbols) or absence (open symbols) of previously having been respected as a member of a research team. The error bars illustrate 95% confidence intervals.***

($F_{1,396}$ = 10.52, $p$ = 0.001 for question 3; $F_{1,396}$ = 10.73, $p < 0.001$ for question 4) and prior respect as a member of a scientific research team ($F_{1,396}$ = 19.45, $p < 0.001$; $F_{1,396}$ = 19.61, $p < 0.001$ for question 4) contributed significantly to these perceptual gains; the students without prior lab experience in research showed greater percentage gains (figures 1 and 2).

## Doing and presenting research

The students' perception of their ability to participate in interdisciplinary research in teams (question 5a); work with research mentors (question 5b); and analyze, write up, and present research data (questions 5c–5e) all increased following participation in the HF-SRPE (table 1). As with basic research skills, the students without previous experience with laboratory research showed greater improvement in these areas (figures 1 and 2), except for ability to analyze data. For that skill, the students with prior research experience had greater gains (figure 1).

## Future aspirations in science

The students showed little change in their conceptual domains of educational and career aspirations (figure 3). The first axis of the correspondence analysis separated environmental from nonenvironmental professions and accounted for 39.8% of the variation in the data. The second axis (27.9% of the variation) separated responses along a postgraduate education versus employment in fields outside of environmental science. There were no significant changes in expressed long-term educational or employment goals (question 8) or interest in environmental or STEM fields (questions 10a and 10b) among the students participating in the HF-SRPE (table 1). Long-term goals were generally uncertain to clear, whereas the likelihood of pursuing a career in environmental or STEM fields was generally likely or quite likely.

## Interesting interactions

For the two questions in which the students experienced the largest gains—scientific research skills (table 1, questions 3 and 4) and presenting scientific results (question 5e)—we also ran rm-ANOVAs in which the students' clarity of long-term goals prior to entering HF-SRPE entered the model a predictor variable. For the students' perceptions of their scientific research skills, there were significant differences between pre- and post-participation responses ($F_{1,407}$ = 41.19, $p < 0.001$), the level of clarity in their long-term goals ($F_{2,407}$ = 8.00, $p$ = 0.001), and their interaction ($F_{2,407}$ = 4.83, $p$ = 0.008). Even though the students entering the program with a lower clarity of their post-graduation goals also had a lower perception of their research skills, the interaction plot (figure 4) illustrated that these students increased their perception of these skills to an equivalent level after the completion of the HF-SRPE. Similarly, for the students' self-reported preparedness to present scientific

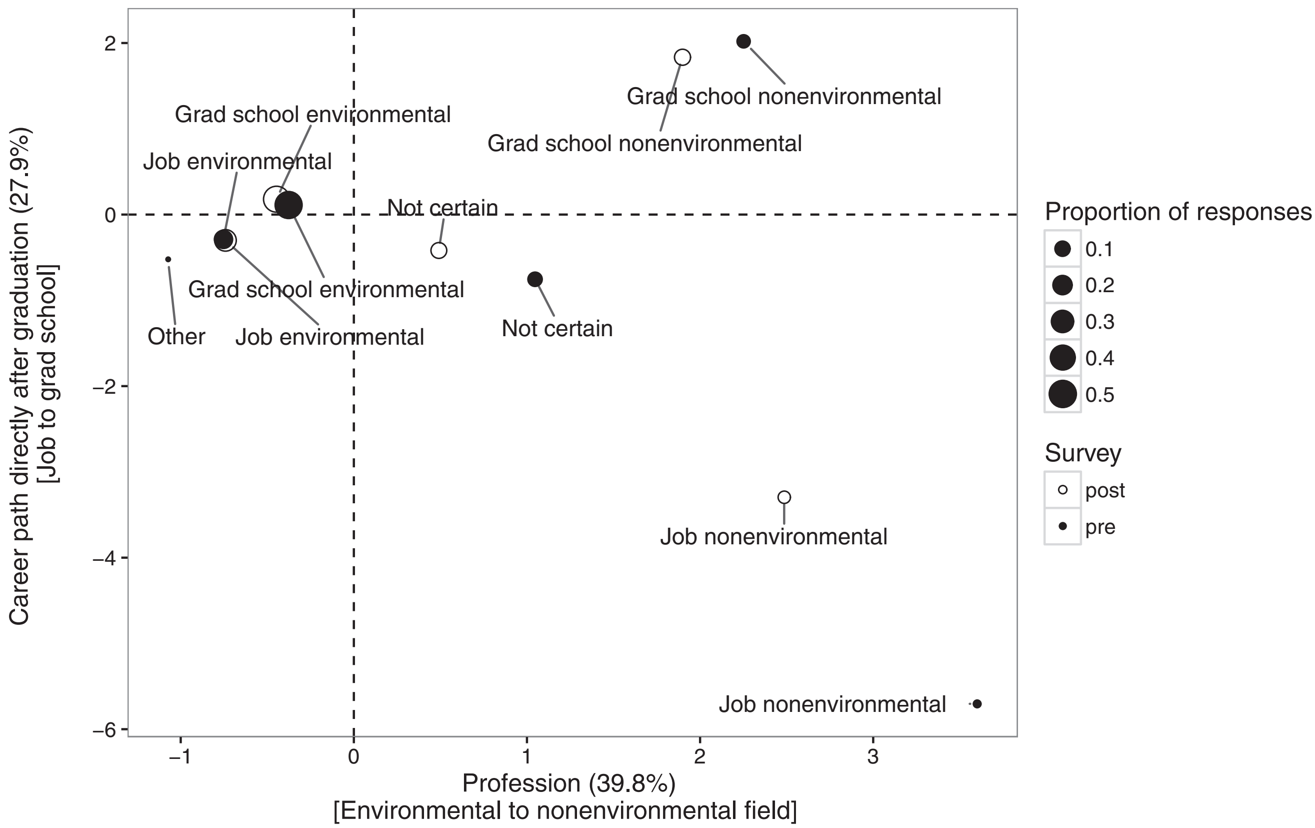


***Figure 3. Correspondence analysis of data on the students' plans after graduation. The pre-test responses (treated as rows) are indicated by the blue circles, and the post-test responses (treated as columns) are indicated by the red triangles. The size of the symbols indicates the relative proportion of individuals responding to a given category. $\chi^2$-distance approximations are only valid among their respective profiles.***

results (figure 4), there were significant differences between pre- and post-participation responses ($F_{1,403} = 25.67$, $p < 0.001$) and the level of clarity in their long-term goals ($F_{2,403} = 4.01$, $p = 0.019$). However, there was no interaction between these two factors ($F_{2,403} = 1.96$, $p = 0.143$).

## Discussion

Undergraduate research experiences (UREs) have been widely touted as providing valuable research experiences that provide valuable skills for future scientists (Russell et al. 2007, Linn et al. 2015). Not surprisingly, the HF-SRPE participants experienced significant increases in the various skills- or self-assessments measured by our survey, but gains were greater for the students without prior research experience (figure 1). This result suggests that UREs interested in promoting these various learning gains should emphasize the recruitment of students without prior research experiences so that these students can begin to expand on these skillsets through future experiences (see also Lopatto 2004, Hunter et al. 2009, Maw et al. 2011).

One limitation of all self-assessment surveys is that students' reporting of their own perceptions of their learning gains could be inflating our impression of success. The internal validity of self-assessment surveys is a known concern (Linn et al. 2015) and supports the use of pre- and post-designs to evaluate educational impacts (Pascarella 2001). Indeed, our conclusions are supported by an ethnographic study of similar URE programs that revealed a strong correlation between student and faculty perceptions of learning gains, especially with regards to constructs such as scientific identify and professional development (Hunter et al. 2007). As long as we remember to treat these data as perceptions, not objective measurements, of cognitive or psychomotor domains (see also Turner et al. 2008), we can begin to assess where the students are experiencing growth as a result of the HF-SRPE.

Although our results suggest positive effects of the HF-SRPE on student learning and skills development, it was more difficult to tease out specific details about the student experiences that contributed to these perceived differences. The instrument was designed to quickly gauge a broad range of skills and experiences that students from different disciplines might experience or have in common and therefore lacked the granularity to describe fully the unique research experience of each student. This is a challenge not only for the HF-SRPE but also for assessing student experiences at other UREs. There are numerous validated concept inventories (e.g., macroevolution, Nadelson and Southerland

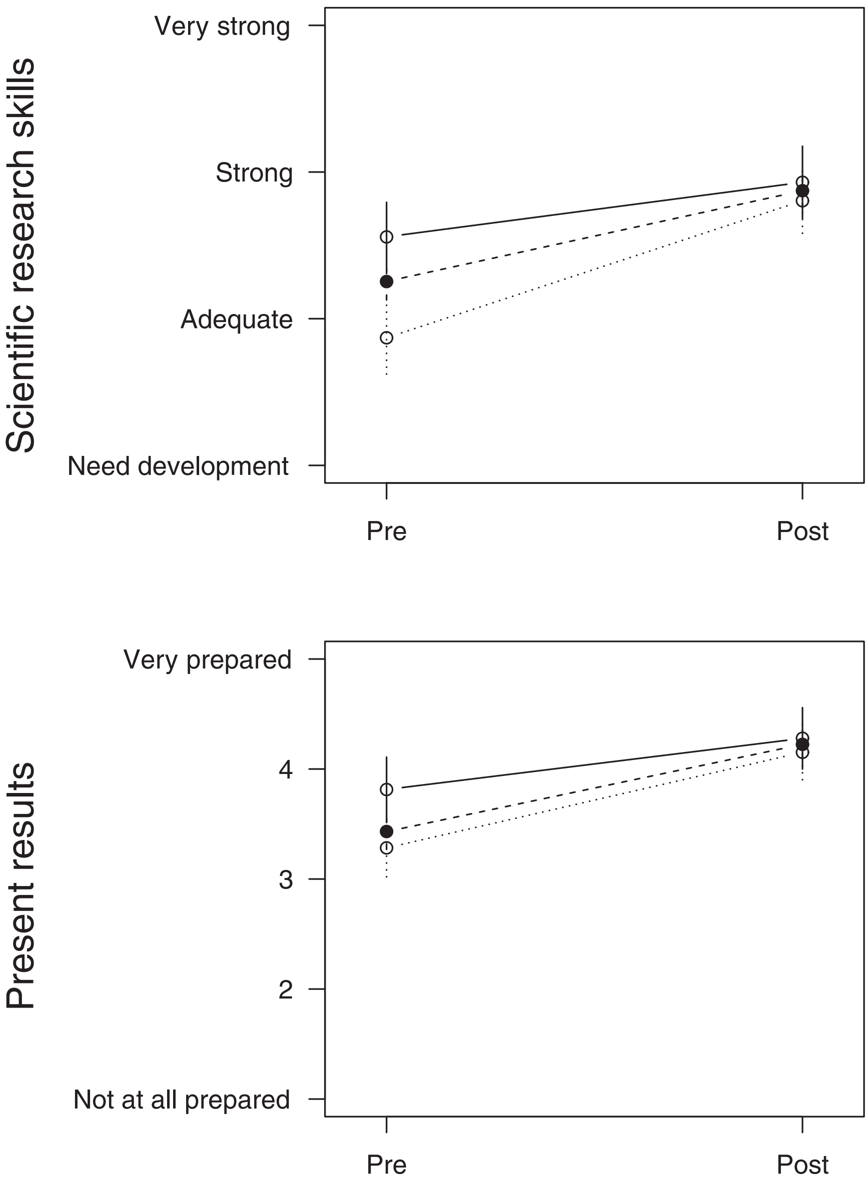


***Figure 4. Mean changes in learning gains based on the clarity of post-graduation plans prior to starting the HF-SRPE. The participants identified clarity as*** **low** ***(triangles),*** **medium** ***(squares), or*** **high** ***(circles). The error bars illustrate 95% confidence intervals.***

2010; natural selection, Anderson et al. 2002; carbon cycles, Hartley et al. 2011; scientific literacy, Nuhfer et al. 2016) that evaluate the working knowledge of an individual and can help demonstrate the value of an intervention such as a URE or course. However, the programmatic or disciplinary specificity of such assessments makes it difficult to compare individual student experiences across programs, especially interdisciplinary ones such as the HF-SRPE.

Compared with our assessment, the URSSA does a more robust job of measuring cognitive and affective domains such as thinking like a scientist, research skills, personal gains, and attitudes and behaviors (Weston and Laursen 2015). Because it is a post-program-only instrument, however, the URSSA cannot differentiate the effects of a URE itself or any selection bias for the type of programs (Linn et al. 2015). This one-time survey also fails to account for differences in scientific ability among individuals prior to the URE, limiting our ability to use it for comparing among programs. Although we advocate the continued use of URSSA, especially because of the nearly complete participation in it of all BIO-REU programs, we encourage other programs to consider adding a pre- and post-instrument so that individual program directors and funding agencies can learn more about the factors underlying the effects of UREs.

**The importance of laboratory experiences.** The students with or without prior laboratory experiences displayed increases in their strength and confidence in scientific skills, participation in interdisciplinary research, analysis of scientific data, and the writing of results (figure 1). The HF-SRPE's nationwide recruitment and breadth of research projects for which students are specifically selected suggest that the prior laboratory experiences reported by our participants are representative of students in other UREs. In another study of comparable summer UREs at liberal-arts colleges, Hunter and colleagues (2007) found that broader and more confident laboratory skills increased the students' sense of independence and helped facilitate other gains beyond their current research projects. Furthermore, they found that the reinforcement of these skills also aided in shaping the students' self-efficacy and scientific identity. In future studies, additional qualitative inquiries into prior laboratory experiences could help reveal which components of these prior experience are related consistently to higher perceptions of various skillsets (table 1). For example, student understanding of the nature of science can be facilitated through the use of *instructional scaffolding*, a process through which progressive activities and experiences can guide a student toward more autonomous learning and stronger conceptual understanding and can help them resolve their misconceptions toward authentic scientific research (Clough 2006).

**Respected member of a scientific team.** In addition to scientific knowledge and skills, the affective domain also played a role

*Table 1. A summary of the students' skills and aspirations.*

| Survey question[a] | Pairs (n) | Pre | Standard error of the mean (SEM) | Post | SEM | *t*-value | *p*-value |
|---|---|---|---|---|---|---|---|
| Q3 | 209 | 2.23 | 0.05 | 2.87 | 0.05 | –11.28 | <.001 |
| Q4 | 208 | 2.12 | 0.03 | 2.49 | 0.04 | –9.42 | <.001 |
| Q5a | 208 | 3.90 | 0.06 | 4.31 | 0.05 | –6.50 | <.001 |
| Q5b | 209 | 4.17 | 0.05 | 4.61 | 0.04 | –7.56 | <.001 |
| Q5c | 209 | 3.56 | 0.06 | 3.81 | 0.06 | –3.70 | <.001 |
| Q5d | 209 | 3.62 | 0.06 | 3.95 | 0.06 | –5.19 | <.001 |
| Q5e | 207 | 3.47 | 0.06 | 4.21 | 0.05 | –11.44 | <.001 |
| Q8 | 205 | 1.99 | 0.05 | 2.04 | 0.05 | –1.02 | .154 |
| Q9 | 179 | – | – | – | – | – | – |
| Q10a | 209 | 3.53 | 0.05 | 3.43 | 0.06 | 2.45 | .993 |
| Q10b | 188 | 3.14 | 0.06 | 2.93 | 0.07 | 3.28 | .999 |

*Note: Directional paired t-tests were used to examine educational gains due to participation in the HF-SRPE. The questions for which post-surveys were not significantly greater than pre-surveys exhibited a decrease.*

[a] Complete question text from survey:

Q3: Would you say that your scientific research skills are 1, *need development*; 2, *adequate*; 3, *strong*; or 4, *very strong*?

Q4: Would you say that your confidence in your scientific research skills is 1, *low*; 2, *medium*; or 3, *high*?

Q5a–e: How prepared are you to (1, *not at all prepared*; 5, *very prepared*)

Q5a: participate in interdisciplinary research with a team of researcher?

Q5b: conduct research supervised by a research mentor?

Q5c: analyze scientific data?

Q5d: write up scientific results?

Q5e: present scientific results?

Q8: Would you say that your long-term post-college goals, either for education or employment, are 1, *uncertain*; 2, *clear*; or 3, *very clear*?

Q9: What are your plans immediately after graduating from college? *graduate school environmental, graduate school nonenvironmental, job environmental, job nonenvironmental, or not certain*?

Q10a–b: What is the likelihood that you will pursue a career in (1, *not at all likely*; 4, *quite likely*)

Q10a: an environmental field?

Q10b: a STEM research field?

in the expression of these learning goals and development of scientific identity. Both before and after participation in the HF-SRPE, having felt respected as a member of a scientific research team was a reliable predictor of higher self-assessment of research skills (figure 2). As with learning gains, the students who reported not having previously been a respected member of a research team displayed a greater degree of change. Respect in the context of UREs can aid in the development of a student's scientific competence and individual identity as researcher (Hunter et al. 2007). Facilitating a culture of respect in an inclusive, collaborative learning community reinforces students' interest and empowers them as active learners (Walsh et al. 2014). Positive interactions with other members of a research community can help foster students' understanding how they construct scientific knowledge and derive meaning from their experiences through "self-authorship" (Baxter-Magolda 1999a). This process of applying their contextual knowledge, a component within the constructive–developmental framework, is especially important for college students as they begin to identify and shape their career paths (Baxter-Magolda 1999b).

**STEM retention.** A common critique of UREs is that they tend to favor students who already have a high probability of persistence in STEM fields (Linn et al. 2015). Our competitive selection process, in which 600–900 applicants are competing for 25–30 positions, may reinforce similar biases. Our data showing that the HF-SRPE has not changed the participants' short-term career paths (table 1, questions 8, 10a, and 10b; figure 3) lend support to this characterization of URE programs. Recent efforts to try to minimize this bias within the HF-SRPE include increasing recruitment at community colleges, removing GPA requirements, and supporting more interdisciplinary projects.

Sadler and colleagues (2010) argued that one of the greatest insights gained through research apprenticeship is a sophisticated understanding of the nature of science. For our participants, prior laboratory research experience resulted in a higher clarity of long-term goals of remaining in STEM fields. This may suggest that the students without previous laboratory experience had an unclear image of research or at least the types of interdisciplinary research conducted at the HF-SRPE. We intentionally recruit students who we think would benefit the most from a URE at a major research institution (figure S1), and we will continue do so, especially in light of the results presented here. Such students express in their application essays a strong interest in ecological research or have demonstrated a potential as an environmental researcher but have not yet had experience with independent research. The HF-SRPE therefore provides students with an opportunity to evaluate their true preparedness for environmental or STEM research disciplines. The absence of a change in their expressed long-term plans may result only from a lack of time to reflect on their summer experience. The long-term evaluation of student career paths will help us differentiate among these alternative hypotheses.

The HF-SRPE provides students already interested in environmental and STEM research with an opportunity to expand their skills and become part of the next generation of research scientists. Program-level tracking of our participants provided annually since 2001 to agencies supporting our URE show that a consistent 15% of each year's participants have published their summer work in peer-reviewed journals, 10% (with rates rising up to 45% within the past 5 years) have presented posters at regional or national conferences, and a consistent 10% of the students have developed their summer projects into senior theses. These data cannot be linked directly to the individual survey responses reported here, but they do lend support to the idea that research skills gained both from prior experiences and from the HF-SRPE have led to the production of professional-level research products. We note that the production of research products, often used to demonstrate the value of professional reasearchers to universities and funding agencies, may not serve as informative indicators of undergraduate learning and growth (Hunter et al. 2007). However, identity theory argues that a collaborative and respectful learning environment helps students apply skills learned through the creation of these research products, increases the salience of their scientific identity, and further strengthens their likelihood of pursuing and remaining in of STEM careers (Merolla and Serpe 2013).

## Conclusions

Our data suggest that to maximize gains in the learning of scientific skills, UREs should emphasize the recruitment of students without certain prior experiences within both cognitive and affective domains. The intellectual, social, professional, and financial support of young students by UREs increases the access to these valuable learning opportunities so that more students have a stronger research foundation to build on in the future. Long-term assessments will illuminate further whether the short-term gains of the HF-SRPE and other UREs have persistent effects (Linn et al. 2015). There is still much more to examine about the relationships between summer UREs and STEM retention; increased focus on recruitment methods and the implementation of repeated-measures designs would help align program-level evaluations with NSF objectives to provide meaningful research experiences for a broader range of undergraduate students (NSF 2013). In addition, the use of more domain-specific measures would indicate changes in the participants' noncognitive skills (e.g., teamwork, professionalism, and work ethic) that are increasingly recognized by employers as crucial traits for workplace and post-educational success. Augmenting the toolkit used for measuring participant outcomes will strengthen the evidence-based decisions made by URE administrators and provide researchers with more fine-grained data about participant outcomes.

## Acknowledgments

Funding for evaluation of the HF-SRPE has been provided by NSF (REU Site award nos. 0452254, 1003938, and 1459519) and NASA Innovations in Climate Education (award no. NNX10AT52A). This research with human subjects was approved by the Harvard University IRB, protocol nos. F14015-102 [2007], F16874-102 [2009], 14-2580 [2014], and 15-2719 [2015].

## Supplemental material

Supplementary data are available at *BIOSCI* online.

## References cited

Anderson DL, Fisher KM, Norman JG. 2002. Development and validation of the conceptual inventory of natural selection. Journal of Research in Science Teaching 39: 952–978.

Baxter-Magolda MB. 1999a. Creating Contexts for Learning and Self-Authorship: Constructive–Developmental Pedagogy. Vanderbilt University Press.

Baxter-Magolda MB. 1999b. The evolution of epistemology: Refining contextual knowledge at twentysomething. Journal of College Student Development 40: 333–344.

Clough MP. 2006. Learners' responses to the demands of conceptual change: Considerations for effective nature of science instruction. Science Education 15: 463–494.

Ellison A. 2016. Harvard Forest Summer Research Program in Ecology Pre/Post Survey 2006–2015. Harvard Forest Data Archive: HF279. Harvard University. doi:10.6073/pasta/41c0bc4d1f2afb4f9b1b577e7beb7600

Frechtling J. 2002. The 2002 User-Friendly Handbook for Project Evaluation. National Science Foundation, Directorate for Education and Human Resources.

Harsh JA, Maltese AV, Tai RH. 2011. Undergraduate research experiences from a longitudinal perspective. Journal of College Science Teaching 41: 84–91.

Hartley LM, Wilke BJ, Schramm JW, D'Avanzo C, Anderson CW. 2011. College students' understanding of the carbon cycle: Contrasting principle-based and informal reasoning. BioScience 61: 65–75.

Hodder J. 2009. What are undergraduates doing at biological field stations and marine laboratories? BioScience 59: 666–672.

Hunter AB, Laursen SL, Seymour E. 2007. Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. Science Education 91: 36–74.

Hunter AB, Weston TJ, Laursen SL, Thiry H. 2009. URSSA: Evaluating student gains from undergraduate research in the sciences. CUR Quarterly 29: 15–19.

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.

———. 2007. Undergraduate research experiences support science career decisions and active learning. CBE–Life Sciences Education 6: 297–306.

Maw SJ, Mauchline AL, Park JR. 2011. Biological fieldwork provision in higher education. Bioscience Education 17: 1–14.

Merolla D, Serpe R. 2013. STEM enrichment programs and graduate school matriculation: The role of science identity salience. Social Psychology of Education 16: 575–597.

Nadelson LS, Southerland SA. 2010. Development and preliminary evaluation of the Measure of Understanding of Macroevolution: Introducing the MUM. Journal of Experimental Education 78: 151–190.

Neuman L. 2003. Social Research Methods: Qualitative Approaches. Allyn and Bacon.

[NRC] National Research Council. 2014. Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the Twenty-First Century. National Academies Press.

[NSF] National Science Foundation. 2005. Research Experiences for Undergraduates (REU): Supplements and Sites: Program Solicitation no. NSF 05-592. NSF.

[NSF] National Science Foundation. 2013. Research Experiences for Undergraduates (REU): Supplements and Sites: Program Solicitation no. NSF 13-542. NSF.

Nuhfer EB, Cogan CB, Kloock C, Wood GG, Goodman A, Delgado NZ, Wheeler CW. 2016. Using a concept inventory to assess the reasoning component of citizen-level science literacy: Results from a 17,000-student study. Journal of Microbiology and Biology Education 17: 143–155.

Pascarella ET. 2001. Using student self-reported gains to estimate college impact: A cautionary tale. Journal of College Student Development 42: 488–492.

R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. (3 October 2016; *www.R-project.org*)

Russell SH, Hancock MP, McCullough J. 2007. Benefits of undergraduate research experiences. Science 316: 548–549.

Sadler TD, Burgin S, McKinney L, Ponjuan L. 2010. Learning science through research apprenticeships: A critical review of the literature. Journal of Research in Science Teaching 47: 235–256.

Scott GW, Goulder R, Wheeler P, Scott LJ, Tobin ML, Marsham S. 2012. The value of fieldwork in life and environmental sciences in the context of higher education: A case study in learning about biodiversity. Journal of Science Education and Technology 21: 11–21.

Turner N, Wuetherick B, Healey M. 2008. International perspectives on student awareness, experiences, and perceptions of research: Implications for academic developers in implementing research-based teaching and learning. International Journal for Academic Development 13: 199–211.

Walsh C, Larsen C, Parry D. 2014. Building a community of learning through early residential fieldwork. Journal of Geography in Higher Education 38: 373–382.

Weston TJ, Laursen SL. 2015. The Undergraduate Research Student Self-Assessment (URSSA): Validation for use in program evaluation. CBE–Life Sciences Education 14: 1–10.

*Andrew L. McDevitt (andrew_mcdevitt@fas.harvard.edu) is a graduate student at the School of Biological Sciences at Illinois State University, in Normal. Andrew studies undergraduate research experiences as it relates to motivational theory, STEM retention, and program assessment. Brad Rose, PhD, (bradrose1@comcast.net) is the president of Brad Rose Consulting, Inc., a program-evaluation firm based in Wellesley, Massachusetts. Aaron M. Ellison (aellison@fas.harvard.edu) is a senior research fellow in ecology and Manisha V. Patel is a research assistant at the Harvard Forest, Harvard University, in Petersham, Massachusetts. Since 2004, Aaron has directed the Summer Undergraduate Research Program in Ecology (an NSF REU site) at the Harvard Forest. His research focuses on the disassembly and reassembly of ecological systems. Manisha (manishapatel@fas.harvard.edu) is interested in ecosystem ecology, undergraduate research education, and outreach.*