Developing Instruments to Assess Learning Outcomes from Undergraduate Research Experiences that Align with Faculty Goals

Abstract:
This study has focused on developing and testing a series of assessment instruments to measure undergraduate student researchers' ability to integrate their understanding of the scientific practices and content knowledge involved in their research project. The assessment tools include rubrics for evaluating poster presentations and responses to a series of open-ended prompts on experimentation. In parallel, we have interviewed faculty about their goals and mentoring approaches for undergraduate researchers in their labs. Results have shown that many of our measures distinguish between undergraduates at three different levels of prior research experience. Interviews showed that these instruments are aligned with the most highly valued outcomes for faculty mentors of URE students.

INTRODUCTION

Undergraduate research experiences (UREs) are widely recognized as valuable to the professional growth of science, technology, engineering, and mathematics (STEM) undergraduates, contributing to increased student retention in STEM, promotion of discipline-specific knowledge, and integration into the larger scientific community (National Academies of Sciences, Engineering, and Medicine, 2017). Many empirical investigations of UREs rely primarily on self-report data to document learning outcomes (NASEM, 2017). URE participants believe they learned scientific practices such as lab techniques, the ability to analyze data, and skill in presentations (Adedokun et al., 2014; Khoukhi, 2013; Lopatto, 2004). However, few studies validate self-reports with analysis of research products, direct measures of scientific content or practice, or observations of student activities (Linn, Palmer, Baranger, Gerard, & Stone, 2015).

We have developed instruments to assess the ways in which scientific practices and content are discussed by undergraduates as they engage in their own research projects. Scientific practices include those formulated in such educational policy documents as Next Generation Science Standards and Vision and Change in Undergraduate Biology Education (Brewer & Smith, 2011; NGSS Lead States, 2013). These instruments will allow for assessment of student learning in UREs, both in terms of the acquisition of conceptual knowledge and gains in skills important to STEM professionals (NASEM, 2017) as part of a larger study investigating learning outcomes for UREs and developing mentor training programs to improve these outcomes (see Figure 1).

Theoretical Framework

We used the Knowledge Integration (KI) framework to guide our design of assessment tools for research on UREs (National Research Council, 2000; Linn, 1995; Linn & Eylon, 2011; Pellegrino, Chudowsky, & Glaser, 2001; Wilson, Howitt, Roberts, Åkerlind, & Wilson, 2013). KI calls for eliciting students' initial ideas by engaging in research practices such as predicting and hypothesizing (White & Gunstone, 2014). KI documents the value of adding ideas by participating in research experiences and guiding students to reflect on their research observations to sort and consolidate their ideas (Brown, Collins, & Duguid, 1989). The KI framework specifies that coherent understanding about science arises when students make deep connections between new and existing ideas (Linn & Eylon, 2011). In the context of UREs, the KI framework suggests that students more fully integrate their knowledge of scientific practices and content as they develop into effective researchers.

The KI framework guides our expectation that, as undergraduates progress in research, they will integrate scientific practices with their understanding of the content of their projects in the following ways (termed Indicators for Integration of Scientific Practices and Content; ISPC):

1. Communicate the significance of their specific project to the overarching research questions of the laboratory and the broader scientific field.
2. Justify their experimental design as appropriate for their research question.
3. Analyze and interpret data in order to construct explanations and models that are relevant to their research question.
4. Generate hypotheses and plan future experiments relevant to their research question in response to their analysis and interpretation of data.

The extent to which these practices and relevant content knowledge are integrated provides the main underlying metric for our instruments.

Research Questions

Our data collection and subsequent analysis was guided by two primary research questions. First, can our instruments detect differences between the scientific content and practices discussed by novice versus advanced undergraduates? Second, what goals do UC Berkeley faculty have for undergraduate students engaged in research in their own laboratories, and does it align with what we are attempting to measure with our instruments?

METHODOLOGY

Participants and Sampling

Participants consisted of undergraduate students majoring in STEM fields at a large research university who were conducting research in faculty laboratories or were enrolled in a Course-based Undergraduate Research Experience (CURE). These individuals ranged from having 0 to 5 or more semesters of research experience prior to study participation. Graduate student researchers also participated in this study to ensure that our instruments are able to capture “expert-like” responses. Gender, race, ethnic, and first language identities were mixed in these study populations.

All students enrolled in one of two different CUREs were invited to be part of this study, and data was collected from everyone who agreed to participate. We attended the poster sessions at the end of each of these courses and talked with a subset of the students who had consented to be part of our study. Students were stratified by major and prior research experience, and a random sample was chosen, intentionally oversampling URM students. To recruit URE participants as well, all presenters at two different annual undergraduate poster sessions were invited to be part of our study. One of these sessions is specific to chemistry and chemical biology students, while the other is more general across STEM disciplines. Everyone who agreed to participate and provided us with written data was interviewed at these poster sessions.

STEM faculty at our institution who are mentoring undergraduate researchers in their laboratories were also interviewed. All chemistry professors and associate professors from biology, physics, and engineering that have undergraduates in their groups were emailed and invited to participate in our study. A total of 21 of 39 chemistry professors and 12 other professors from various subfields of biology, physics, and engineering consented to being interviewed. This study was authorized by the University of California, Berkeley institutional review board, approval #2016-02-8360. All participants completed informed consent.

Instrument Development: Measures of URE Learning Outcomes

A central source of data for our instruments consisted of observing student presentations at **poster sessions**. The researchers attended normally scheduled poster sessions in which the study participants were presenting. After each student presented their research without interruption, 7 follow-up questions based on the ISPC indicators were asked. Assessment was based entirely on transcripts of audio recordings for the oral presentation and answers; photos or observations of the poster itself were not included in the coding process. In addition to attending poster sessions, researchers gave student participants two prompting questions about recent data analysis and future directions for their projects. The **journal reflections** produced in response to these questions provide another primary data source for this work.

Using a small subset of the data, rubrics were developed based on the KI framework and the observed student responses to each of the instruments. These rubrics include a number of different items, many of which are shared across instruments, such as measures of knowledge integration in the areas of rationalizing design choices, interpreting data, and proposing next steps for the project (see caption to Figure 2 for a full list of items). After the initial development of our instruments, data from 62 undergraduate students at various levels of research experience were analyzed to determine whether our instruments could detect differences between novice and more experienced undergraduate researchers. We completed analyses of both the reflective prompt responses and poster presentation transcripts for each study participant to make comparisons between different levels of experience. Each transcript or response was coded independently by at least two people using the previously developed rubrics. Any discrepancies between coders were discussed and resolved.

Faculty Interviews

Faculty members were interviewed in a semi-structured format regarding their goals for their undergraduate researchers and the organization of UREs in their laboratory. Many of these questions were modified from a recent multi-institutional study of UREs (Laursen, Hunter, Seymour, Thiry, & Melton, 2010). In addition to open-ended inquiries about faculty goals, interviewees were probed with specific URE outcomes reported in the literature and asked for their assessment of the Indicators for ISPC. Additional discussion included questions regarding how undergraduate student researchers are selected, the formal and informal guidance provided to them by their mentors, and whether graduate student mentors are given any specific instruction on how to interact with undergraduate researchers in the group.

Interviews were transcribed and initially coded for answers to the following interview questions, regardless of where they appeared during the interview:

1. What overall goals do you have for your undergraduates while they are in your lab?
2. How do you choose which undergraduates will work in your lab?
3. Do you give any advice or expectations to graduate student mentors about working with an undergraduate?

Excerpts that provided faculty answers to the above questions were collected and organized into common themes. Interview transcripts were then recoded using these themes to get a quantitative count of the most commonly expressed views.

FINDINGS

To distinguish between more and less experienced undergraduate researchers, participants ($n = 62$) were split into three groups based on their prior research experience, and their prompt response scores were compared. An average score across 7 different items showed significant differences between each of the three groups; the most experienced students scored higher than the intermediate students ($p < 0.05$), who in turn scored higher than the novice students ($p < 0.01$). Although our sample size did not permit statistical analysis for individual rubric items, the overall trend for all items appears to be a gradual increase with more research experience (see Figure 2a).

Our poster rubric was used to evaluate poster presentations for an overlapping set of students ($n = 61$). As with the prompt responses, an average score across all 11 items on the poster rubric showed significant differences between each of the three groups; the most experienced students scored higher than the intermediate students ($p < 0.05$), who in turn scored higher than the novice students ($p < 0.05$). All data were coded in pairs, and a weighted Cohen's kappa value for interrater reliability between pairs was calculated to be 0.82. Again, the overall trend for nearly all items appears to be a gradual increase with more research experience (see Figure 2b). However, one exception to this trend is student ability to discuss the societal context of their project, which was a point of strength for novice researchers and did not appear to increase with additional research experience.

Faculty interviews revealed that exposing undergraduates to "real research" is a priority for nearly all interviewees. There is a particular focus on helping students decide whether they want to pursue a career in research and, if so, preparing them for graduate school. Other common responses included making meaningful contributions to the lab's research, gaining technical skills, and prioritizing safety (Figure 3). Faculty overwhelmingly responded that the central indicators for ISPC are important goals for them and a good description of the overall research process. No significant differences were observed between chemistry and other STEM faculty.

Additionally, we found that when faculty consider offering a student a position in their group, they tend to look at one or more of the following characteristics of the undergraduate: academic interests and career goals, academic performance, stage of undergraduate career, and general temperament. In particular, faculty are often hesitant to give positions to students who cannot make at least a one-year commitment or have not expressed at least some interest in graduate school and a research career. Within faculty labs, graduate students or postdocs serve as the direct mentors for undergraduate researchers in nearly all cases. Faculty guidance for graduate students on how to mentor undergraduates is very limited and highly variable.

CONCLUSIONS

We have developed a set of instruments to assess the extent to which undergraduate researchers develop an integrated understanding of scientific practices and content over time. Our analysis revealed that these instruments distinguish between undergraduates at three different levels of prior research experience, with the poster and prompt instruments providing complementary sets of data on individual participants. The validity of our instruments derives partly from the fact that they assess authentic practices that researchers often participate in, particularly poster presentations. Additionally, the high agreement among faculty of the importance of our indicators for ISPC supports the validity of our underlying constructs. Finally, an examination of the data shows that a broad range of the target skills can be detected by these instruments, including the ability to capture features of expert-like presentations by advanced graduate students. These instruments will be useful for monitoring undergraduate researchers longitudinally to determine which features of UREs, particularly with respect to mentoring, correlate with more integrated knowledge around student projects.

Interviews showed that STEM faculty goals at our institution are highly consistent with the Indicators for ISPC on which our instruments are based. Given the primary emphasis on having undergraduates be exposed to “real research,” our instruments measure the outcomes valued by faculty mentors of URE students. Interviews also showed a substantial gap in faculty guidance for graduate students and postdocs on mentoring their undergraduates, suggesting that there is a need for mentor and teacher training in order to offer undergraduate researchers more consistent and targeted support.

ACKNOWLEDGEMENTS

This work was funded by the National Science Foundation *Improving Undergraduate STEM Education* initiative (NSF IUSE 1712001), and the Alfred P. Sloan Foundation. Additional support comes from the Berkeley Graduate School of Education Barbara Y. White Fund. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 1752814. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Adedokun, O. A., Parker, L. C., Childress, A., Burgess, W., Adams, R., Agnew, C. R., ... & Teegarden, D. (2014). Effect of time on perceived gains from an undergraduate research program. *CBE—Life Sciences Education*, *13*(1), 139-148.

Brewer, C. A., & Smith, D. (2011). Vision and change in undergraduate biology education: a call to action. *American Association for the Advancement of Science, Washington, DC*.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32-42.

Khoukhi, A. (2013). A structured approach to honours undergraduate research course, evaluation rubrics and assessment. *Journal of Science Education and Technology*, *22*(5), 630-650.

Laursen, S., Hunter, A.-B., Seymour, E., Thiry, H., & Melton, G. (2010). *Undergraduate research in the sciences: Engaging students in real science.* San Francisco, CA: Jossey-Bass.

Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: The scaffolded knowledge integration framework. *Journal of Science Education and Technology*, *4*(2), 103-126.

Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. Routledge.

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., & Stone, E. (2015). Undergraduate research experiences: Impacts and opportunities. *Science*, *347*(6222), 1261757.

Lopatto, D. (2004). Survey of undergraduate research experiences (SURE): First findings. *Cell biology education*, *3*(4), 270-277.

National Academies of Sciences, Engineering, and Medicine. (2017). *Undergraduate research experiences for STEM students: Successes, challenges, and opportunities*. National Academies Press.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By State*s. Washington, DC: The National Academies Press.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press, 2102 Constitutions Avenue, NW, Lockbox 285, Washington, DC 20055.

White, R., & Gunstone, R. (2014). *Probing understanding*. Routledge.

Wilson, A., Howitt, S., Roberts, P., Åkerlind, G., & Wilson, K. (2013). Connecting expectations and experiences of students in a research-immersive degree. *Studies in Higher Education*, *38*(10), 1562-1576.

FIGURES

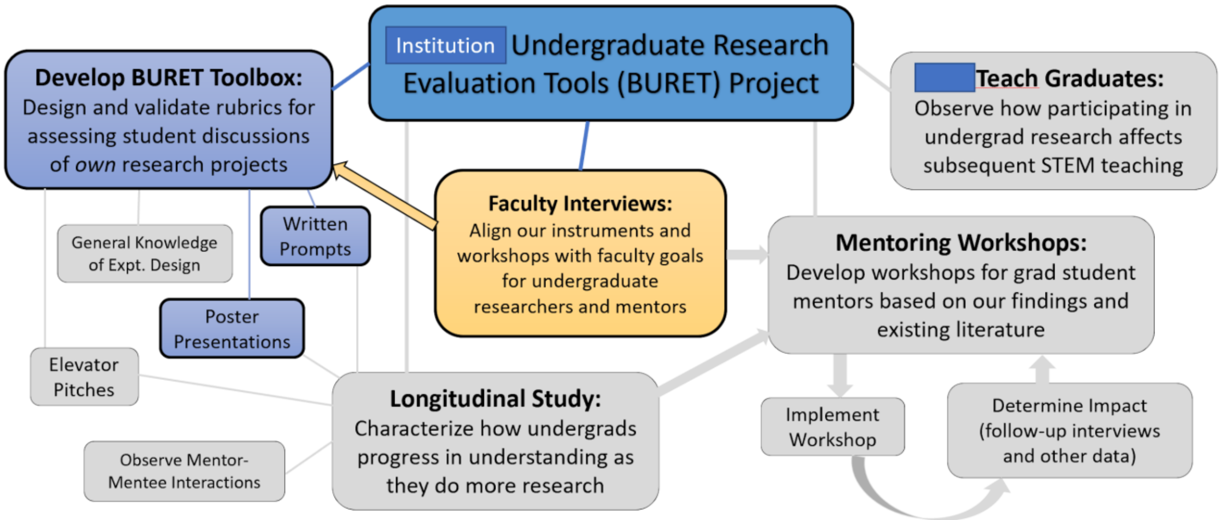

Figure 1. Project overview; portions discussed in this submission are highlighted.

a)

Research Experience	n	1	2	3	4	5	6	7	Mean
0-1 Semesters	34	2.9	2	2.1	2.4	0.9	2.6	0.3	1.9
2-3 Semesters	16	3.3	2.7	2.4	2.9	1.8	2.9	0.8	2.4
4+ Semesters	12	4.2	2.9	2.8	3.3	2.2	3.7	1.4	2.9

b)

Research Experience	n	1	2	3	4	5	6	7	8	9	10	11	Mean
0-1 Semesters	24	2.3	3.6	3.5	2.8	2.7	2.5	3.1	2.3	3.1	1.9	1.5	2.7
2-3 Semesters	15	3.3	3.5	3.7	3	3.1	3.3	3.3	2.5	3.1	2.8	2.3	3.1
4+ Semesters	22	3.8	3.7	4.2	3.7	3.8	3.6	3.9	2.7	3.5	3.4	3	3.6

Figure 2. Mean scores on prompt response and poster presentation items for students with different levels of prior research experience.

a) Mean scores on prompt responses for the following items: 1. Connecting their work to a broader context, 2. Providing rationale for their experimental design choices, 3. Identifying the relevant variables and how they connect, 4. Describing the manipulation of their raw data to yield useful information, 5. Interpreting their data, 6. Proposing next steps for the project, and 7. Integrating additional content knowledge into their presentation.

b) Mean scores on poster presentations for the following items: 1. Connecting their work to a broader scientific context, 2. Connecting their work to a broader societal context, 3. Providing rationale for their experimental design choices, 4. Assessing limitations of their experimental design choices, 5. Comparing their experimental design choices to potential alternatives, 6.

Number of design choices for which some rationale was given (max. 5), 7. Interpreting their data, 8. Analyzing sources of error and uncertainty, 9. Proposing next steps for the project, 10. Incorporating references to previous work, and 11. Integrating additional content knowledge into their presentation.

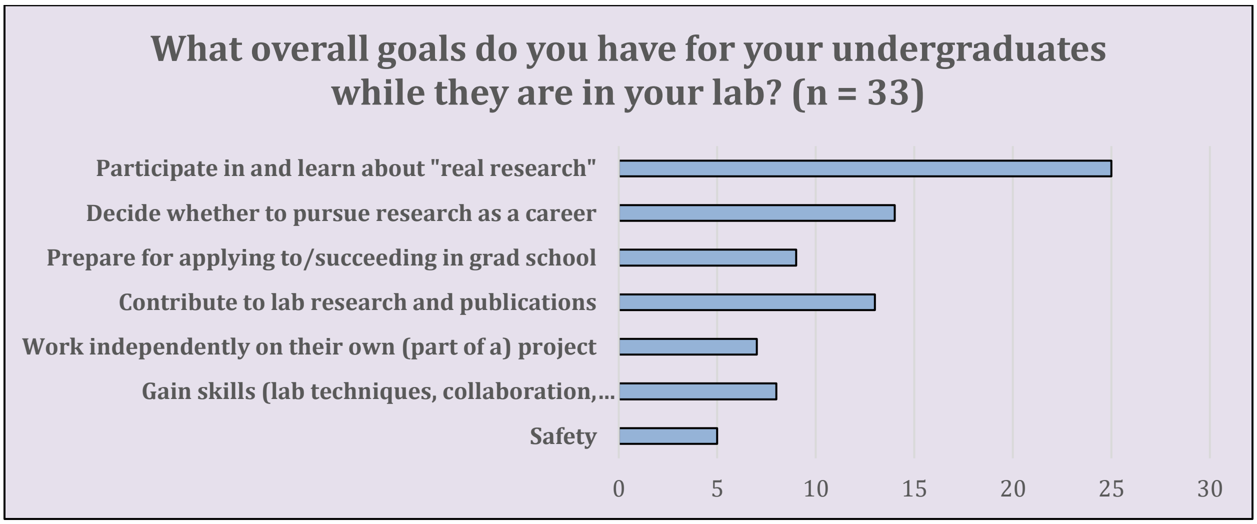

Figure 3. Most common faculty responses to open-ended question about goals for undergraduate researchers.