Assessment of Undergraduate Research Learning Outcomes: Poster Presentations as Artifacts

Rachel Hayes-Harb, University of Utah
Mark St. Andre, Megan Shannahan, University of Utah

Abstract
The authors have developed a set of undergraduate research learning outcomes that address the traditions of research and mentoring across campus. Achievement of these outcomes is assessed at annual, institution-wide, undergraduate research events by employing a poster presentation evaluation rubric and deploying graduate students, postdoctoral scholars, and faculty as ad hoc raters. Between April 2018 and April 2019, 2,721 rubrics evaluating 803 undergraduate research posters by 436 raters were collected. Students participating in the one-semester funded and mentored undergraduate research program performed significantly higher on all four quantified learning outcomes than did nonparticipants. It was further found that disciplines exhibited different profiles of relative strength and weakness with respect to the various learning outcomes. Together, these findings inform future programmatic decision-making at the institution.

Keywords: assessment, learning outcomes, poster presentations, rubrics, undergraduate research

doi: 10.18833/spur/3/4/10

Undergraduate research (UR) offices and administrators on higher education campuses have become increasingly accountable for assessing the impact of their programs on various measures of student success (see, e.g., Banta 2007; Crowe and Brakke 2008, 2019; Rogers and McDowell 2015). A number of studies have revealed that UR experiences may positively impact student retention, graduation, graduate school attendance, and career pathways, among other metrics of student achievement (e.g., Bauer and Bennett 2003; Cibelli 2015; Eagan et al. 2013; Follmer et al. 2017; Gregerman et al. 1998; Hathaway, Nagda, and Gregerman 2002; Jones, Barlow, and Villarejo 2010; Lopatto 2004; Parker 2018; Vincent-Ruz, Grabowski, and Schunn 2018), although many studies lack the controls necessary to provide evidence of a causal relationship (notable exceptions include Gregerman et al. 1998 and Sell, Naginey, and Stanton 2018). Other efforts have focused on assessing achievement of learning outcomes associated with UR (Crowe and Brakke 2008; Laursen 2015), although these efforts are often discipline specific, dependent on indirect methods of assessment (e.g., post-experience self-assessment by students) and small in scale (e.g., limited to a single academic program or discipline). This article describes an ongoing, campus-wide, UR learning outcomes assessment project at the University of Utah that was designed to address these challenges by (a) employing a set of UR learning outcomes that can be applied regardless of discipline; (b) directly assessing student poster presentations as UR artifacts via rubrics; and (c) deploying volunteer poster assessors at campus-wide undergraduate research events.

Undergraduate Research Learning Outcomes
The Office of Undergraduate Research at the University of Utah serves students working with faculty research mentors in all of the institution’s 18 colleges and schools, ranging from fine arts and humanities to engineering, medicine, and pharmacy. Identifying UR learning outcomes that can be applied regardless of discipline poses a challenge given that “students and their research advisors experience UR differently, depending on their discipline and its intellectual and pragmatic ways of working” (Laursen 2015, 10; see also Kardash 2000 and Stokking et al. 2004). Indeed,
although the literature provides examples of STEM-oriented UR learning outcomes assessment (Hopper, Schumacher, and Stachnik 2013; Kardash 2000; Seymore et al. 2004; Weston and Laursen 2015), there are relatively few examples of learning outcomes that can be applied regardless of discipline (exceptions include Russell, Hancock, and McCullough 2007 and Singer and Zimmerman 2012). Despite varied traditions of research and mentoring across the wide range of disciplines, this article describes an attempt to capture essential learning outcomes associated with the range of UR experiences at the university. This effort involved a review of UR learning outcomes published by various institutions and was particularly influenced by the work of Singer and Zimmerman (2012), Singer and Weiler (2009), and the UR learning objectives provided by Loyola University Chicago (n.d.). The current set of UR learning outcomes are as follows:

1. Identify and utilize relevant previous work that supports the research.
2. Articulate a timely and important research question or creative objective.
3. Identify and utilize appropriate methodologies to address the research question or creative objective.
4. Present the research effectively in a conference setting and a written publication.
5. Meet the relevant field’s standards for the responsible conduct of research and effectively navigate challenges that arise in the research process.
6. Work collaboratively with other researchers, demonstrating effective communication and problem-solving skills.
7. Reflect constructively on the research experience, identifying what was learned, personal strengths and opportunities for growth, and how the experience informs future educational and career goals.

**UR Learning Outcomes Assessment**

The UR learning outcomes assessment efforts began in spring 2016, with the start of self-assessment data collection from students upon completion of the one-semester funded research program (the Undergraduate Research Opportunities Program, or UROP) and the small research grants and research travel grants programs. Each learning outcome was presented following the carrier phrase “I can . . . .” and students were asked to select from *strongly agree, agree, disagree,* and *strongly disagree* in response to each statement. Response rates have been quite high (between 76 and 93 percent, by semester and program), due to an office policy requiring the completion of final reports as a prerequisite for future funding program participation. In addition, faculty mentors were invited to rate students’ achievements relative to each learning outcome using the same scale. Those response rates have been somewhat lower (23 to 47 percent), as this task was optional for mentors. Given that ease of data collection using Likert scales facilitates multipoint assessment (Singer and Zimmerman 2012), since fall 2018 student self-assessments and mentor assessments also have been collected prior to UROP participation. Although self-assessment with Likert scales is widely used, convenient, and low cost and lends itself well to repeated assessments (Singer and Zimmerman 2012), its limitations also are well known (Pike 2011). Indeed, even mentor assessments of students raise validity concerns: as the mentor-student relationship evolves, mentors become more likely to exhibit bias in favor of their students (Johnson 2008). An alternative to asking students to self-assess and asking mentors to assess their own students is to involve independent raters to score assignments, learning portfolios, or other artifacts using rubrics against learning outcomes, because they are direct evidence of student work (Diller and Phelps 2008; Montgomery 2002).

In summer 2017, a rubric was developed to use in evaluating student achievement of the seven UR learning outcomes (Office of Undergraduate Research n.d.). The rubric was designed to be used quickly and easily by untrained raters. It includes four scored elements, relating to the first four learning outcomes listed above. Each learning outcome is associated with “indicators of excellence.” For example, for the learning outcome “Articulate a timely and important research question or creative objective,” the indicators are: (a) explicitly states the research question or creative objective, (b) explains why the research question is important and timely, and (c) the research question or creative objective follows logically from the previous work cited. Each of the four outcomes can earn up to 25 points. A 25-point scale was selected to encourage raters to make finer distinctions among student presentations than would be possible on a smaller scale. Raters were instructed that a score ranging from 0 to 10 points indicates the student is in the earliest stages of development with respect to the learning outcome. A score ranging from 11 to 20 points indicates that the student is in the process of achieving the learning outcome. Finally, a score ranging from 21 to 25 points indicates the student has mastered the learning outcome. The final three learning outcomes—relating to the responsible conduct of research; collaboration, communication, and problem solving; and reflection—are less likely to be directly reflected in a poster presentation. Therefore, raters were asked to describe evidence of the student’s achievement of these outcomes if observed. The rubric was piloted by colleagues in the National Science Foundation–funded Research Experience for Undergraduates in physics and astronomy at the first annual Summer Research Symposium in 2017 (45 evaluations of 14 posters by 8 raters) and at a small, informal poster session hosted in December 2017 (27 evaluations of 6 posters by 7 raters). Rater feedback and the scores they provided indicated that the rubric was easy to use and that it elicited a wide range of scores.
Assessment Protocol Implementation

In spring 2018, an institution-wide undergraduate research assessment project was launched, taking advantage of the institution’s annual undergraduate research events (the Undergraduate Research Symposium, held each April, and the Summer Symposium, held each August); their hundreds of student poster presentations; and their visibility to graduate student, postdoctoral, and faculty researchers throughout the campus. All undergraduate students involved in mentored research are eligible to present at these events. In addition, presentation at the Undergraduate Research Symposium is required for students funded by the Office of Undergraduate Research, for those earning the honors degree, and for some capstone courses. Note that the statistics presented in this article illustrate the institution-specific implementation of a quality improvement assessment project and may not generalize to other institutions. As such, the project does not meet the University of Utah’s definition of research and is not subject to Institutional Review Board oversight.

Figure 1 (top) shows that, although students in STEM disciplines were well represented at these events, students from a wide range of disciplines participated. While the proportion of raters associated with the health sciences was higher than the proportion of students from the health sciences presenting, the opposite was true for social and behavioral sciences. Relative patterns of participation by students and mentors is a topic worthy of further study.

Existing undergraduate research presentation events were leveraged to engage expert researchers in directly assessing the poster presentation artifacts. The institutional response was enthusiastic: at the 2018 Undergraduate Research Symposium, 154 raters completed 1,006 rubrics for 243 student posters. At the much smaller (and newer) 2018 Summer Research Symposium, the ratio of raters to student presenters was even higher: 97 raters completed 544 rubrics for 102 undergraduate student posters. The number of raters grew to 185 at the 2019 Undergraduate Research Symposium, although the ratio of raters to posters declined due to a very large increase in the number of student posters (to 458; the result of a campaign by the Office of Undergraduate Research to increase student involvement in the event). Table 1 presents the assessment project statistics for these undergraduate research events, showing the numbers of student posters and raters in addition to numbers of rubrics completed per rater and per poster.

Rater Recruitment

Beginning several weeks in advance of the event, graduate students, postdoctoral researchers, and faculty on campus were invited to serve as raters. An invitation and flyer were sent to the office’s faculty and staff listserv, directors of graduate studies, graduate program professional advisers, postdoctoral researchers, the mentors of the student presenters, and raters who participated in previous events. All were asked to share the information with their colleagues, and the information was also shared via the office’s Twitter and Instagram accounts. The invitation included a link to sign up via an online form, enabling the office to track interest, provide instructions, and remind raters before the event. To encourage rater participation, there were drawings for the winners’ choices of an Apple Watch or Amazon Echo Show. The invitation and flyer also appealed to the interest of the campus community in supporting undergraduate researchers and presented the experience as an opportunity for professional development. Across all three
events under consideration, 26 percent of raters were faculty, 20 percent were postdoctoral scholars, 6 percent were master’s students, and 48 percent were doctoral students. Figure 1 (bottom) presents the distribution of these raters by discipline, which roughly matches the distribution of disciplines represented by the undergraduate student presenters (Figure 1; top).

Data Quality
To maximize the ability to normalize the data, rubrics were distributed to raters in packets of five with a randomly pre-assigned poster number at the top of each rubric. Because a goal of these events was to help students develop their ability to communicate their research to broad audiences, the rater’s scholarly discipline was not considered in poster assignment; raters were thus experienced researchers who did not necessarily have specialist knowledge in the topics of the posters they evaluated. As an incentive to complete the evaluation of at least five posters, for each complete packet returned the rater received a single entry for the prize drawing. Raters were further provided with instructions to help them avoid conflicts of interest: they were told that they were considered to be a conflicted rater if they had a personal or professional relationship with the poster presenter and to randomly select an alternative poster to judge in its place. They also were instructed to select an alternative poster if the presenter was not at the poster after multiple visit attempts.

Data Analysis and Reporting
Rubrics were checked for completeness and scanned for archiving purposes, and data was input by undergraduate student interns; posters and raters were identified by numbers to preserve student and rater privacy. Raters were found to use the full range of scores (0–25 points). Scores could be presented in their absolute forms as means and ranges for a given poster. In a modest attempt at rater normalization, ratings could be converted to z-scores to allow for comparisons across groups of students. Although a discussion of the results of this assessment project is beyond the scope of the present article, it is worth noting that the rater-normalized rubric data collected revealed that students who participated in the semester-long funded research program (UROP) earned significantly higher scores on all four quantified learning outcomes than did students who had not (see Table 2).

These differential scores confirmed that the rubric as deployed by the raters was sensitive to differences in the profiles and research experience levels of students.

Raters also provided written comments supporting their quantitative evaluations for the first four learning outcomes and were invited to provide additional written comments relating to the final three outcomes. These written comments could be shared with students and mentors and coded for additional quantitative analysis. Raters were told that their ratings and comments might be shared anonymously with students and mentors for their benefit, and upon request aggregated data might be presented to units for individual academic program assessment purposes.

The data collected through this project may further be useful in determining discipline-specific UR programming. Table 3 presents the highest- and lowest-rated learning outcomes for each of the six disciplinary categories, suggesting that students in the various disciplines may benefit from programming that responds to their particular strengths and weaknesses.

Discussion
The UR learning outcomes assessment data collected can be leveraged to further a number of institutional goals. In addition to contributing to what Banta (2007) refers to as
Finally, this project demonstrates a commitment to the transparent and collaborative assessment of UR learning outcomes at the University of Utah and beyond. Banta (2007) recommends sharing rubrics across campuses to assess “senior projects, capstone papers, and products of undergraduate research” (11). In this spirit, open access to the rubric is provided on the office’s website (Office of Undergraduate Research n.d.), and colleagues are invited to share feedback for further refinement of learning outcomes assessment efforts.

This assessment project also has a number of weaknesses that will be addressed in future work. For example, further rich insights may be revealed by qualitative and/or mixed methods of assessment (see, e.g., Childress 2015; Laursen, Seymour, and Hunter 2012; Seymour et al. 2004). This has not been undertaken due to the limited capacity of the office at present to expertly analyze the thousands of sets of rater comments. In addition, a systematic procedure for sharing rater feedback with students and their mentors has not yet been identified, despite interest among students and faculty to review this feedback. Finally, the number of volunteer raters has not increased at the same rate as the number of presenters — rater recruitment efforts must be increased to keep pace with increases in undergraduate research participation on campus. The authors continue to

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>UROP participants</th>
<th>Non-UROP students</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Identify and utilize relevant previous work that supports the research</td>
<td>20.2</td>
<td>19.2</td>
<td>t(632) = 2.713</td>
<td>0.007</td>
</tr>
<tr>
<td>2. Articulate a timely and important research question or creative objective</td>
<td>21.0</td>
<td>20.0</td>
<td>t(631) = 3.331</td>
<td>0.001</td>
</tr>
<tr>
<td>3. Identify and utilize appropriate methodologies to address the research question or creative objective</td>
<td>20.7</td>
<td>19.3</td>
<td>t(632) = 5.317</td>
<td>0.0005</td>
</tr>
<tr>
<td>4. Present the research effectively in a conference setting and a written publication</td>
<td>20.9</td>
<td>20.2</td>
<td>t(632) = 3.933</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

**TABLE 2. Assessment Project Outcomes Analysis 2018–2019**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Lowest-rated</th>
<th>Highest-rated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Architecture, humanities, and arts</td>
<td>Methods</td>
<td>Previous work</td>
</tr>
<tr>
<td>Business and law</td>
<td>Research question/creative objective</td>
<td>Presentation</td>
</tr>
<tr>
<td>Health sciences</td>
<td>Presentation</td>
<td>Methods</td>
</tr>
<tr>
<td>Social and behavioral sciences</td>
<td>Methods</td>
<td>Research question/creative objective</td>
</tr>
<tr>
<td>Science and engineering</td>
<td>Previous work</td>
<td>Methods</td>
</tr>
<tr>
<td>Social work and education</td>
<td>Presentation</td>
<td>Previous work</td>
</tr>
</tbody>
</table>

**TABLE 3. Lowest- and Highest-Rated Learning Outcomes by Discipline**

“assessment for accountability,” or the assessment that is increasingly required of higher education institutions by external sources, this assessment project was designed to serve several additional purposes.

Over time, this data will allow observation and evaluation of temporal relationships between institutional or programmatic developments and changes in student learning outcomes. As data is accumulated for increasing numbers of students, the quantitative thresholds necessary for the responsible use of propensity matching techniques will be reached and exceeded, allowing for inferences regarding causal relationships between participation in UR programming and student learning outcomes.

In addition, inviting faculty, postdoctoral researchers, and graduate students to serve as raters communicates to the campus community that the research expertise and commitment to mentoring provided by these colleagues are valued. The rubric further serves as a guide to students and their research mentors as they develop their projects and their presentations, reinforcing student and mentor awareness of the office’s values and goals. The data collected can be reported to academic units on campus in aggregated form to support their own internal assessment and program improvement efforts.
work to address these and other limitations of the assessment project.

In closing, we return to the ideals driving undergraduate research learning outcomes assessment. As Childress (2015) so eloquently puts it, “[w]e don’t want undergraduate research for its own sake . . . the primary focus of our assessment is not the work that we do, but the work that we enable; not the programs we run, but the lives we change” (7–8). It is for these reasons that the University of Utah and the Office of Undergraduate Research have made this investment in carefully assessing students’ achievement of undergraduate research learning outcomes.

Acknowledgments
The authors are grateful to colleagues in Undergraduate Studies at the University of Utah for their myriad contributions to the development and implementation of this project and thank the members of the University of Utah community who have contributed their time and expertise as raters. This manuscript has substantially benefited from feedback from colleagues at the 2018 CUR Biennial Conference, from Melinda Licht and Robert Kirby, and from the editors and reviewers of SPUR. Most of all, the authors are grateful to the University of Utah’s undergraduate students and their faculty mentors for their scholarly and mentoring contributions.

References


Rachel Hayes-Harb
University of Utah, r.hayes-harb@utah.edu

Rachel Hayes-Harb is an associate dean in Undergraduate Studies, director of the Office of Undergraduate Research and Capstone Programs, and a professor of linguistics at the University of Utah.

Mark St. Andre is an associate dean in Undergraduate Studies with responsibilities in the Office of General Education and Learning Outcomes Assessment and an adjunct assistant professor in the Department of Educational Psychology at the University of Utah.

Megan Shannahan is an academic program manager in the Office of Undergraduate Research and Capstone Programs at the University of Utah.

---

**Want to Review Manuscripts for SPUR?**

Send an email to SPUR@cur.org indicating areas of interest/specialization (e.g., assessment). Experience with qualitative/quantitative methods is highly desired.