Challenges and opportunities for measuring student outcomes of undergraduate research

Sandra L. Laursen

Ethnography & Evaluation Research, University of Colorado Boulder 580 UCB, Boulder, CO 80309-0580 USA

Abstract. Inherent in the practice of apprentice-model undergraduate research (UR) is a fundamental tension between the educational goals of UR and its basis in faculty scholarship. This tension leads to challenges for faculty in guiding student researchers in their daily work and in positioning their own UR work within institutionally bifurcated domains of teaching and research. It also generates a disconnect when it comes to measuring the outcomes of UR. Traditional outcome measures emphasize students' career outcomes and research productivity, while education research has documented students' personal and professional learning from UR, including new skills and understandings of disciplinary inquiry, growth in confidence and responsibility, and scientific identity development. Thus far, self-report measures including surveys and interviews have dominated this young body of research. I discuss why assessing the outcomes of apprentice-model undergraduate research is inherently difficult, outline some strengths and limitations of the approaches tried to date, and suggest areas for future research, including the design and measurement challenges that arise in attempting to incorporate undergraduate research into courses.

PACS: 01.40.Di, 01.40.Fk, 01.50.Qb

I. THE FUNDAMENTAL TENSION OF UR

In traditional apprentice-model undergraduate research (UR), students work with experienced scientists in the context of an active research lab, investigating a question that advances the group's scholarly agenda and (ultimately) contributes new scientific knowledge of interest to the field. By working with real scientists using real disciplinary tools and confronting everyday problems as they arise in the research setting, students develop conceptual knowledge, practical and problem-solving skills, build habits of critical thinking and analysis, and come to understand how scientific knowledge is built.[1] They grow in confidence, take on increasing responsibility for their decisions, and become socialized into the work, behaviors, and values of scientists, thus enabling them to make an informed personal choice about joining the profession or taking their very transferable skills to other lines of work.

Our large-scale interview studies of apprentice-model UR make clear that the strong personal and professional benefits of participating in undergraduate research arise directly from students' engagement in faculty scholarship, participating in "real science," as they carry out research activities and try on the role of scientist themselves.[2] The good learning outcomes for students are intrinsically tied to the scholarly goals, methods and context of the project.

However, also inherent in the practice of apprenticemodel undergraduate research is a fundamental tension: accomplishing the educational benefits for students may be at odds with the scholarly agenda of the project. Students learn powerfully from trying something, failing, analyzing the failure and trying again. As scientists know, this is an authentic experience of science—but failures can be costly too, with wasted time and materials, and broken equipment. Skilled UR advisors navigate this tension daily in their work with students, judging when to let students run with an idea and when to rein them in or offer guidance, when to speak up or stay quiet. They acknowledge this as a creative tension that attends any work with student researchers—not a problem that can be solved once and for all.[3]

This creative tension between the educational value of UR and its scholarly goals also surfaces in considering how to measure the outcomes of UR for students. Traditional measures have valued scholarly products, counting studentcoauthored publications and presentations, and alumni who go on to earn advanced degrees. But multiple studies agree

²⁰¹⁵ PERC Proceedings, edited by Churukian, Jones, and Ding; Plenary, doi:10.1119/perc.2015.plenary.002 Published by the American Association of Physics Teachers under a Creative Commons Attribution 3.0 license. Further distribution must maintain attribution to the article's authors, title, proceedings citation, and DOI.

that students do make significant educational gains and concur on the general nature of these gains—which are valuable to both future scientists and to those who enter teaching, medicine, and other fields. With institutions and funders eager to demonstrate the added value of their UR programs, there is understandable interest in and need for better measures of these outcomes.

Despite this broad consensus, the outcomes of UR as an educational intervention are inherently tricky to document and attribute in detail. Students' experience of UR varies by discipline, by research group, and even by the nature and stage of the project, as well as in the motivation and background they (and their advisors) bring to it. Outcomes in the domain of "thinking like a scientist"—including skill in carrying out scientific practices such as designing experiments and analyzing data, and broader skills in thinking critically and solving problems in applied contexts—are especially highly valued but are challenging to assess in a way that works well across UR's varied contexts. It is thus hard to draw causal conclusions about what components of UR account for these good outcomes.

II. MEASURING UR: THE EXAMPLE OF URSSA

URSSA, the Undergraduate Research Student Self-Assessment, offers a good example of these measurement challenges and opportunities. Recognizing the need for simple, reliable and low-cost measures for evaluating UR programs, our group developed URSSA as an assessment tool for UR programs.[4] This free survey instrument can be used to compare similar programs or to track a program's evolution over time. A core set of multiplechoice and open-ended items asks students to self-rate their gains in four broad domains: skills, thinking and working like a scientist, personal gains related to research (such as confidence and collaboration skills), and attitudes and behaviors that indicate adoption of the identity and status of a scientist. Other core items probe UR activities known to influence these outcomes, such as having responsibility for a project and presenting findings to others, while optional items allow program directors to gather feedback on specific design elements of their UR program.

Evaluation studies show that URSSA items can discriminate the level of gains made by novice and more experienced student researchers,[5] identify amplified gains among students from groups underrepresented in science,[6] and show relationships of gains to students'

experience of mentoring.[7] URSSA is designed as a postonly instrument because we find that students' understanding of some of the items, especially those in the important domain of "thinking like a scientist," shift as a result of UR experience—making a pre/post comparison unreliable as scale endpoints shift in meaning.

URSSA is useful because it is free, reliable, and comprehensive in examining the range of previously known gains from UR. It was carefully developed, empirically grounded in prior research and constructed using best practices for survey development.[4] Confirmatory factor analysis shows that the four-factor structure designed into URSSA is the best fit to the survey structure.[8] However, URSSA and like instruments[e.g. 9] based on student selfreport have limitations.[8,10-12] Some of these arise from students' lack of familiarity or good feedback on the skills and knowledge represented, and others from the measurement conditions under which surveys are typically applied, e.g. when student samples are small, highly selected, and non-anonymous, and when high individual and programmatic stakes may attach to students' responses.

Why haven't educational studies of UR included controlled studies that use validated assessments to measure specific UR outcomes? Many practical challenges loom: studies of UR must nearly always be conducted in partnership with institutional UR programs, as UR is too labor-intensive an intervention to operate solely to support educational research. Yet the breadth and interlinked nature of UR outcomes means that programs are understandably reluctant to focus evaluation measurements only on a narrow subset of outcomes that a researcher may wish to probe deeply; the stakes are too high for them to agree to random assignment to treatment and control groups. At the same time, there is a limit to the number of detailed assessments that can be given to any one set of UR students. These factors make it difficult to gather detailed and generalizable data about specific UR outcomes.

III. THE APPEAL OF COURSE-BASED RESEARCH EXPERIENCES

While better tools for measuring UR outcomes and processes remain a goal, sufficient evidence is available to argue that the benefits of UR are both significant and distinctive, and that they emerge from a type of apprenticeship in which the organizing principle is scientific authenticity of the research problem, methods and standards to which work is held. Thus institutions and funders seek to expand the availability of UR and reduce its cost per student. Earlier entry to UR is also desired, based on evidence that students benefit from the more extensive experience made possible by an earlier start in research[5] and that early research encourages students from firstgeneration college families and underrepresented minority groups to pursue science degrees and research careers.[1]

Thus there is keen interest in developing alternatives to apprentice-model UR that can accommodate more students earlier and at lower cost—especially curricular forms that may be called "research-based courses," "course-embedded research," or "course-based undergraduate research experiences." Variations include lab courses based on systems well suited to student experimentation,[13] analysis of existing citizen science data,[14] or very large-scale distributed projects that use students to gather many variants of a particular observation.[15,16] Efforts to define whether and how such courses constitute "research" have focused on identifying their key elements, for example, those proposed by Corwin Auchincloss et al.:[17]

- 1) The project includes an element of *scientific discovery*; its outcomes are not predetermined.
- 2) Students learn and use scientific practices.
- 3) *Iteration* is built in as a source of learning, through trying, failing, critiquing, and retrying.
- 4) *Collaboration* is built in as a source of learning, and deepens skills, understanding, and metacognition.
- 5) The topic is *broadly relevant* beyond the class; it offers opportunity for impact or action.

Taking the view that the desired outcomes should determine the design of a research-based course, Wilson, Howitt and Higgins[18] usefully contrast the value-added learning that emerges from doing research with outcomes that are extensions of conventional courses, and define four main categories of this value-added learning. We used their categories to map outcomes of a pilot research-based course for beginning undergraduates.[19] Students made good gains in what Wilson et al. called the "nitty-gritty" of the project: project-specific technical skills, general skills such as keeping a lab notebook, and understanding that research takes time and requires care. Students also developed a sense of themselves as scientists, gaining confidence in their ability to do research, feeling ownership of the project, and making progress in their ability to tolerate repetition and failure. This occurred even when students' interest in pursuing a research career declined: they valued the selfknowledge gained from trying out research as a possible path. But students did not make strong gains in two other value-added domains, the more general capacities to carry out scientific practices and to think critically. Moreover, some of students' more pronounced gains, such as content knowledge, differed little from what might have been acquired in a standard lecture/lab course. This researchbased course was well designed and executed in many respects—yet it alerts us that simply engaging students in "real research" does not automatically involve a level of inquiry needed for students to develop their abilities to think and work like scientists.

These quite preliminary but interesting results raise many questions about the nature, degree and ordering of outcomes from research-based courses: What outcomes are indeed possible from a one-term research experience? How do these depend on the design of the course, and on student maturity? Are certain skills or outcomes typically more readily developed before others? Given that course design choices must be made, which outcomes do we value most? There is much yet to learn as educators across disciplines consider whether and how to design and teach such courses.

IV. KEEPING THE TENSION CREATIVE

Our early work sought to establish the educational benefits of apprentice-model UR[20] and showed how these benefits emerged in a holistic way from the scientific authenticity of the research project.[1] When viewed as a balancing act between the educational and scientific goals, the main risk of tipping too far toward educational benefits lies not reaching the scholarly goals. This may risk the UR advisor's scientific productivity, reputation and career prospects, or her motivation to continue to involve undergraduates in the lab. In contrast, too little emphasis on the educational benefits turns students into technicians: "A well-designed robot" could do my job, said one student who was not fully involved in his project's intellectual work.[21]

The metaphor of creative tension applies equally to course-based research, but is exacerbated because courses face significant constraints that traditional apprentice-model UR does not: students' laboratory hours are few and must be artificially fit into a class schedule, and costs must be minimized. Institutions' long-term investment in sustaining such a course may be uncertain. If instructors of researchbased courses have little prior experience with inquiry teaching, some iteration may be needed to learn to teach effectively in this setting—yet there is immediate pressure to gather data to "prove" that this approach works. For course-based research, it is hard to identify any serious risks of too much focus on student outcomes: it is a course, after all, and students are supposed to learn. The risks instead emerge from over-weighting course design toward its scientific goals. Emphasizing "real research" is tempting, as institutions and funders today find prestige in boasting of the undergraduate teaching commitments of their top scientists. The creative tensions then become:

- Imposing the structure needed to support scientific goals *versus* providing opportunities for students to mess up, analyze their failure, revise their plan, and start again.
- Mitigating the open-endedness and high stakes of "real" research, which can be daunting for new researchers *versus* offering students sufficient independence that they can overcome legitimate challenges and experience the intellectual excitement of making a discovery.
- Providing enough structure for students to master certain skills and knowledge *versus* leaving room for curiosity and exploration.

How ironic it would be if, in an effort to ensure that students obtain sufficient scientific results to feel their

- S. Laursen *et al.*, Undergraduate Research in the Sciences: Engaging Students in Real Science (Jossey-Bass, San Francisco, 2010).
- [2] S.A. Barab & K.E. Hay, J. Res. Sci. Teach. 38(1), 70 (2001).
- [3] S. Laursen, E. Seymour, & A.-B. Hunter, *Change* 44(2), 30 (2012).
- [4] A.-B. Hunter, T.J. Weston, S.L. Laursen, & H. Thiry, Council Undergrad. Res. Quarterly 29(3), 15 (2009).
- [5] H. Thiry, T.J. Weston, S.L. Laursen, & A.-B. Hunter, *CBE Life Sci. Educ.* **11**(Fall), 260 (2012).
- [6] T. Weston, *BIO REU third year report*. ATLAS Assessment and Research Center, 2013.
- [7] C. Hayward, H. Thiry, & S. Laursen, Evaluation of the undergraduate research programs of the Biological Science Initiative: Advisors' perspectives. Ethnography & Evaluation Research, 2013.
- [8] T.J. Weston & S.L. Laursen, *CBE-Life Sci. Educ.* 14(3), ar33 (2015).
- [9] D. Lopatto, Cell Biol. Educ. 3(4), 270 (2004).

contributions are "real," they develop an unreal belief that research marches along so predictably that it is not really very interesting... and thus reject a future research career. Rather than emphasize whether the research is "real" in its scholarly goals, the focus should be to design, implement and institutionalize good, extended inquiry experiences that are "real" in how they involve students in the scientific process. This may or may not require a connection to instructors' scientific pursuits, but does require a question that will engage and motivate students, clear and selective learning goals, and good partnership between educational designers, instructors, researchers and institutional leaders.

ACKNOWLEDGEMENTS

These ideas benefited from conversations with Tim Weston, Chuck Hayward, Heather Thiry, Sarah Wise, Erin Dolan, Sara Brownell, Lisa Corwin Auchincloss, Joanne Stewart and Tom Higgins. Prior work cited here was funded by the U.S. National Science Foundation and Howard Hughes Medical Institute.

- [10] S.L. Laursen, Council Undergrad. Res. Quarterly 35(3), 9 (2015).
- [11] M.K. Eagan et al., Amer. Educ. Res. J. 50(4), 683 (2013).
- [12] M.C. Linn et al., Science 347, 1261757 (2015).
- [13] S.E. Brownell et al., CBE-Life Sci. Educ. 14(2), ar21 (2015).
- [14] I. Betancourt & C.M. McLinn, J. Microbiol. Biol. Educ. 13(1), 86 (2012).
- [15] T.C. Jordan et al., MBio 5(1), e01051-13 (2014).
- [16] J. T. Wang et al., J. Microbiol. Biol. Educ. 16(1), 50 (2015).
- [17] L. Corwin Auchineloss *et al.*, *CBE-Life Sci. Educ*.**13**(1), 29 (2014).
- [18] A. Wilson, S. Howitt & D. Higgins, Assess. Eval. Higher Educ., 1 (2015).
- [19] S. Laursen & S. Wise, unpublished results (2015).
- [20] E. Seymour, A.-B. Hunter, S.L. Laursen, & T. DeAntoni, *Sci. Educ.* 88(4), 493 (2004).
- [21] H. Thiry, S.L. Laursen, & A.-B. Hunter, J. Higher Educ. 82(4), 357 (2011).