8-2022

Assessment in Undergraduate Research

Jill Singer
Buffalo State University, singerjk@buffalostate.edu

Daniel Weiler

Bridget Zimmerman

Sean Fox

Elizabeth L. Ambos

Recommended Citation

Singer, J., Weiler, D., Zimmerman, B., Fox, S., & Ambos, E. (2022). Assessment in Undergraduate Research: The EvaluateUR Method. In H. Mieg, E. Ambos, A. Brew, D. Galli, & J. Lehmann (Eds.), The Cambridge Handbook of Undergraduate Research (Cambridge Handbooks in Education, pp. 158-171). Cambridge: Cambridge University Press. doi:10.1017/9781108869508.021

15

Assessment in Undergraduate Research

The EvaluateUR Method

Jill Singer, Daniel Weiler, Bridget Zimmerman,
Sean Fox, and Elizabeth L. Ambos

This chapter focuses on assessment as an integral element of the continued success and sustainability of undergraduate research (UR) and describes the EvaluateUR method, a proven approach to assessing the skills and competencies of UR students, which also contributes directly to student learning. The chapter concludes with brief remarks on the assessment of course-based UR and two projects now in progress that are adapting the EvaluateUR method to these assessments.

15.1 Background: Undergraduate Research Growth and Assessment Trends

Elizabeth L. Ambos

UR is increasingly viewed as a highly desirable practice worthy of substantial investment, as evidenced by its expansion from natural sciences to many other disciplines (e.g., Crawford et al., 2014), and the formation of international organizations or coalitions to promote and support UR. Prominent examples of the latter development include the Council on Undergraduate Research (BCUR: bcur.org.uk), the British Conference of Undergraduate Research (BCUR: www.bcur.org/), the Australasian Council for Undergraduate Research (ACUR: www.acur.org.au/), and the International Conference of Undergraduate Research (ICUR: www.icurportal.com). In no small part, UR's rise is due to the realization that high-impact learning practices that include UR increase student engagement and success (e.g., Kuh, 2008), and that undergraduate teaching and research can be better integrated through the inclusion of UR (e.g., Brew, 2006).

The rapid expansion of UR programs over the last half century is driving the emergence of a vigorous research area: assessing myriad aspects of the

https://doi.org/10.1017/9781108869508.021 Published online by Cambridge University Press

increasingly global UR movement. These assessments seek to determine why, how, and to what extent UR impacts student participants and others. Most assessment efforts address UR that follows a student–mentor format, meaning that students generally work one-on-one or (less often) in small research groups with mentors (usually faculty) who counsel them and oversee their work. Crowe and Brakke (2008, 2020) offer comprehensive summaries, and a Zotero resource of UR assessment references is provided by CUR on its website (www.cur.org/impact/assessment_strategies/).

The most fundamental line of research inquiry is the effort to understand *who* is participating in UR, *what* they are researching, and *when* and *where* they are undertaking this research. With the rise of UR and/or experiential learning offices on many campuses, records of UR participation are now often kept by academic institutions, and may be linked to students' academic records. While these records were initially limited to UR in the physical sciences, descriptive or assessment criteria particular to arts and humanities disciplines are now also commonly available (e.g., Gilliams et al., 2008; Johnson & Gould, 2009). In addition, most campuses now conduct surveys or interviews of UR participants in order to assess the quality of their experiences. The data gathered by this research provide important insights into the benefits that students perceive from their UR participation.

While collecting UR participation data can be challenging and time-consuming (Blockus, 2012), many institutions are gradually improving their data acquisition practices, particularly when UR offices are formally tasked with tracking UR participants. Thorough records include data on UR participants' demographic attributes, indicate whether UR work is paid (and if so from what source), show whether UR is formally course-based or extracurricular, and note whether UR students are at an introductory level or undertaking capstone projects that may have presentation and/or publication requirements. Institutional, regional, national, and now international conferences are common venues for UR presentations (Rivera et al., 2018), and specialized UR journals and library repositories of capstone UR projects provide highly visible evidence of UR achievements (Hensley & Johnson, 2019). Conference exit surveys or interviews are another assessment resource that have added to our understanding of UR's extent, practices, and impact (Crowe et al., 2010; Hill & Walkington, 2016; Spronken-Smith et al., 2013).

Starting in the 1990s, the availability of multi-year records of UR participation, together with student and faculty surveys, allowed researchers to better understand UR's positive impact on student success. For the first time, the question of why and how UR can benefit students could be understood in terms of the correlations between UR participation and increased student success as measured by grades, retention in degree program, time to graduation, and (from faculty testimony) better acculturation to professional practices (e.g., Lopatto, 2010). Studies of post-undergraduate

https://doi.org/10.1017/9781108869508.021 Published online by Cambridge University Press

career success (e.g., Bauer & Bennett, 2003; Schmitz & Havholm, 2015; Willison et al., 2020) have also suggested that there can be long-term benefits of UR participation.

One of the most exciting findings of studies conducted within the United States is that UR participation by individuals from groups that have been historically underrepresented in higher education often results in statistically significant gains in graduation rates and progress to postgraduate study. For example, landmark experimental studies conducted by Sandra Gregerman and colleagues at the University of Michigan, Ann Arbor in the 1990s to early 2000s (e.g., Hathaway et al., 2002; Nagda et al., 1998) demonstrated statistically significant differences in academic success between African-American students who participated in UR and those who did not. Over the past decade, careful design and implementation of quasi-experimental studies to replicate and extend knowledge of UR's benefits for historically underrepresented groups in the United States has led both to affirmation of earlier findings and a more detailed understanding of why and how UR has contributed to these successful outcomes (Eagan et al., 2013; Estrada et al., 2018; Finley & McNair, 2013; Hurtado et al., 2014).

UR's success is now well documented as being closely tied to the learning, skill-building, and psychosocial aspects of students' research experiences (Hunter et al., 2006; Laursen et al., 2010; Lopatto, 2010). Not surprisingly, effective mentoring is identified most frequently as being crucial to this success, and within the last decade research has proliferated on who undertakes this role, the best way to prepare to be a mentor, and what specific aspects of mentoring are most valuable for student success (e.g., Vandermaas-Peeler et al., 2018).

It is important to note that despite the recent interest in quasi-experimental studies of UR's impact, most studies of UR's success still rely on self-reported data. Unless substantial care is taken in the design, testing, and validation of self-reported data such studies may not fully meet the standard of proof recommended by educational researchers (Laursen, 2015). An innovative evaluation method that does provide valid and reliable evidence of UR student success is described below.

15.2 EvaluateUR: A Method for Assessing Undergraduate Research Student Outcomes and Contributing to Student Learning

Jill Singer, Daniel Weiler, Bridget Zimmerman, and Sean Fox

Buffalo State College (SUNY: State University of New York) undertook a multi-year effort aimed at developing and field-testing an evaluation methodology for measuring student learning and related outcomes in its summer UR program, in which students work one-on-one with faculty

https://doi.org/10.1017/9781108869508.021 Published online by Cambridge University Press

mentors. The goal was to extend the findings of the many valuable studies, described above, that had already examined the impacts of UR on participating students. Previous studies focused mainly on student and faculty perceptions of the value of UT. To move beyond these studies, a key activity at the beginning of this process included a multi-day retreat attended by faculty from the physical and social sciences, the arts, and humanities. This diverse group considered the student outcomes that should be measured by the evaluation and identified a range of outcomes critically important both for graduate studies and the workplace, such as communication skills, creativity, autonomy, critical thinking, and problem-solving ability. The retreat led to the identification of eleven outcome categories, each defined by several measurable components. The evaluation approach centered on having both faculty mentors and their student researchers assess student knowledge and skills on each of these outcome components three times (at the beginning, middle, and end) during the student's research project, followed each time by student–mentor conversations to compare and discuss the reasons for their respective assessments. These conversations aimed to provide students with new insights into their thinking processes and learning strategies. In this way, the evaluation approach sought to collect reliable data on specific student outcomes and also contribute directly to student learning. One of the novel features of this approach to evaluation is that it is embedded within the research and mentoring processes, while at the same time generating reliable data that can be used by directors of UR programs to document their programs' impacts. More details about the development of the method are provided in Singer and Weiler (2009). Singer and Zimmerman (2012) discuss the method at further length and provide data that demonstrates the method's robustness for students in all academic disciplines and for both new and experienced faculty mentors. With awards from the United States' National Science Foundation, Buffalo State's summer research evaluation method has since been scaled up to the national level and is now known as EvaluateUR.

15.2.1 Student Outcome Categories

As the Buffalo State evaluation design was refined, each of the eleven outcome categories was further defined by specifying several measurable student behaviors, resulting in a total of thirty-five outcome components (see Table 15.1).

The outcome categories shown in Table 15.1 are also closely aligned with the wide range of essential workplace competencies identified by the Office of Career, Technical, and Adult Education, United States Department of Education (www.cte.ed.gov/employabilityskills) and National Association of Colleges and Employers (www.naceweb.org/career-readiness/competencies/career-readiness-defined).

Table 15.1. *EvaluateUR: Outcome categories and outcome components*

Outcome categories	Outcome components
Communication	Understands and uses discipline-specific language Expresses ideas orally in an organized, clear, and concise manner Writes clearly and concisely using correct grammar, spelling, syntax, and sentence structure
Creativity	Displays insight about the topic being investigated Shows ability to approach problems from different perspectives Uses information in ways that demonstrate intellectual resourcefulness Effectively connects multiple ideas/approaches
Autonomy	Demonstrates an ability to work independently and identify when guidance is needed Accepts constructive criticism and uses feedback effectively Uses time well to ensure work gets accomplished Sets and meets project deadlines
Ability to deal with obstacles	Is not discouraged by unforeseen problems and perseveres when encountering challenges or setbacks Shows flexibility and a willingness to take risks and try again Trouble-shoots problems and searches for ways to do things more effectively
Intellectual development	Recognizes that problems are often more complicated than they first appear Approaches problems with an understanding that there can be more than one right explanation or even none at all Displays accurate insight into the limits of his/her own knowledge and an appreciation for what isn't known
Critical thinking and problem solving	Challenges established thinking when appropriate Looks for the root causes of problems and develops or recognizes the most appropriate corrective actions Recognizes flaws, assumptions and missing elements in arguments
Practice and process of inquiry	Demonstrates ability to formulate questions and hypotheses within the discipline Demonstrates ability to properly identify and/or generate reliable data Shows understanding of how knowledge is generated, validated, and communicated within the discipline
Nature of disciplinary knowledge	Shows understanding of the way practitioners think within the discipline and view the world around them Shows understanding of the criteria for determining what is valued as a contribution in the discipline Shows awareness of important contributions in the discipline and who was responsible for those contributions Reads and applies information obtained from professional journals and other sources Is aware of professional societies in the discipline
Content knowledge and methods	Displays knowledge of key facts and concepts Displays a grasp of relevant research methods and is clear about how these methods apply to the research project being undertaken Demonstrates an appropriate mastery of skills needed to conduct the project
Ethical conduct	Recognizes that it is unethical to create, modify, misrepresent, omit, eliminate, or misreport data or findings, or to misrepresent authorship Behaves with a high level of collegiality and treats others with respect
Career goals	Is clear about academic and/or professional/work plans Is aware of how research skills relate to academic and/or professional/work plans

https://doi.org/10.1017/9781108869508.021 Published online by Cambridge University Press

15.2.2 EvaluateUR Assessments

The EvaluateUR instrument reflects the student outcomes shown in Table 15.1. It uses a five point scale ranging from "always" to "not yet" to indicate how often a student displays the behavior described by the relevant outcome component. This instrument is first provided to each student–mentor pair at an orientation session prior to beginning student research activities, so that both students and mentors can familiarize themselves with the outcome categories and components, with the expectations of the evaluation, and with the methodology that emphasizes parallel student and mentor assessments and the importance of student–mentor conversations. The orientation session also provides an opportunity for students and mentors to discuss how the EvaluateUR outcome categories and outcome components are related to the forthcoming student research.

Beginning with a "baseline assessment" before research begins, and followed by two additional assessments (at the mid- and end-points of the student's research project), students score themselves on each outcome component, and their research mentors, using the same instrument, also independently score their students. Repeating the assessments provides the equivalent of pre-, mid-, and post-scores. On each of these occasions, students and mentors compare their assessments and discuss the reasons for the scores they have each assigned. By using identical outcome categories and their respective explanatory components, as well as an identical scale and scale rubric for all assessments, EvaluateUR conducts assessments according to explicit and uniform standards across varied disciplines and across different student–faculty pairs. These and related features of the EvaluateUR design are intended to overcome the usual objections to assessing student outcomes based on faculty judgments, which typically rely on disparate standards without common assessment parameters across disciplines or among different faculty. At the same time, EvaluateUR emphasizes that the assessment scores are less important than the conversation that follows the assessments, at which time the student and mentor share their rationales for assigning particular scores and discuss the reasons for differences, if any, in their perceptions. These conversations are supported by score reports, sent to each student–mentor pair as soon as they submit their respective scores to the EvaluateUR website, that show how they scored each outcome component. The reports highlight all scores that are two or more scale points apart, so that students and their mentors can focus their conversations on those outcome components where their assessments were farthest apart.

A feature of EvaluateUR is automated statistical reporting called EZStats. For each outcome component it generates the assessment scores entered by students, showing the pre- to mid- mean score gain (or decline), the mid- to post- mean gain (or decline), and the difference between pre- and post-means. Similar statistics are provided for assessment scores entered by mentors. These are composite scores for students and mentors, in a format that makes them readily usable for publication.

15.2.3 Outcomes from an Independent Evaluation

In 2019, EvaluateUR was implemented by UR directors at thirty-seven institutions with 799 student–mentor research pairs completing all of the EvaluateUR steps. After analyses of assessment scores for this group and feedback from user surveys, an independent evaluation of EvaluateUR arrived at a number of conclusions about the method:

- EvaluateUR succeeded in introducing student participants to a comprehensive list of competencies and skills they would need in order to go on to graduate work and/or succeed in the workplace.
- All EvaluateUR assessment components saw statistically significant positive student gains and the EvaluateUR evaluation model successfully measured objective student growth.
- Post-assessment student–mentor conversations contributed to the development and enhancement of student metacognitive skills, characterized by learners becoming aware of what learning strategies they are pursuing and why, and then using that awareness to make intentional adjustments to those strategies in order to learn more effectively. These conversations also helped most students confirm their plans for graduate school or employment in their major field.
- Research mentors found it easier to identify the academic strengths and weaknesses of the students they mentored, enabling them to focus their mentoring efforts more productively.
- The mentors could also more easily see areas where students might be over- or underestimating their competencies and could help students gain new insights into their academic strengths and weaknesses and the relative efficacy of their learning strategies.
- UR program directors were provided with tools to help them manage EvaluateUR on their campuses, and with reliable, readily understandable evidence of the potential benefits of their UR programs.

In sum, the independent evaluation found that EvaluateUR tested an innovative method for evaluating UR in a way that could reliably measure specific knowledge and skill outcomes while also contributing directly to student learning. The evaluation concluded that EvaluateUR succeeded in meeting all its objectives and thereby demonstrated the feasibility of its method in a wide variety of college and university settings.

15.2.4 Benefits of EvaluateUR

EvaluateUR provides a range of benefits to participating students and mentors, as well as to UR directors. These benefits are described below and summarized in Table 15.2.

> Students: EvaluateUR helps students to become aware of the wide range of competencies and skills they should strive to master in order to succeed in graduate work and/or the workplace.

https://doi.org/10.1017/9781108869508.021 Published online by Cambridge University Press

Post-assessment student–mentor conversations contribute to the development and enhancement of student metacognitive skills, characterized by learners becoming aware of what learning strategies they are pursuing and why, and then using that awareness to improve their learning strategies. Metacognitive learning strategies have been shown to improve problem solving (Schraw et al., 2006; Scharff et al., 2017), critical thinking (Hogan et al., 2015), and the transfer of key skills from one learning context to another (Billing, 2007). Panadero and colleagues (2017) point to a positive connection between improved self-assessment and improved self-efficacy. Wolters and Hussain (2015) note a connection between metacognition and perseverance. Metacognition abilities are also critically important in the workplace, because they make it possible for employees to make accurate appraisals of their work-related strengths and weaknesses and make needed adjustments.

> Mentors: EvaluateUR enables faculty mentors to observe changes in student competencies and skills using an identical assessment scale and rubric at several different stages of student research. They can therefore more easily identify areas where students may be over- (or under-) estimating their competencies and/or may need additional faculty guidance, and can help students gain new insights into areas where they should consider revising their learning strategies.
>
> UR Directors: EvaluateUR provides UR directors (or other administrators with core responsibilities for experiential learning) with reliable evidence on the potential benefits of UR in helping students to achieve specific knowledge, skills, and competencies that they will need for graduate work and/or the workplace. This evidence is provided in the form of statistical reports that can be used directly by research directors in their own reports to other interested parties.

The unique features of EvaluateUR are a departure from the *practice* but not the *purpose* of most evaluations. Whether evaluation is focused mainly on compliance issues or is concerned with a program's impact on students, the ultimate purpose of virtually all education program evaluation is the enhancement of student learning through program improvement. Evaluations ordinarily do not get directly involved in the teaching/learning process, out of concern that doing so might compromise objectivity in their measures of student progress. In that tradition, evaluation is seen not as a direct, real-time contribution to student learning but as a potential corrective to education programs that should be either improved or abandoned. EvaluateUR shares the evaluation goal of enhancing student learning but believes that, for UR, that goal can best be achieved by involving students and faculty together in the assessment task in order to develop and enhance student metacognitive skills, as described above. In this way, EvaluateUR both collects reliable data on specific student outcomes and uses this procedure to contribute directly to student learning. The EvaluateUR methodology is designed to accomplish this without sacrificing accuracy or

Table 15.2. *EvaluateUR benefits*

Students

1. Are introduced to a comprehensive list of competencies and skills that include but go beyond subject-area knowledge that they will require to pursue graduate work and/or succeed in the workplace
2. Are provided with regular feedback about their progress through repeated assessments and follow-up conversations with mentors
3. Obtain a realistic picture of their strengths and weaknesses across all competencies and skills they should strive to achieve
4. Develop or enhance their metacognitive skills
5. Gain greater self-awareness and confidence as they track their academic growth
6. Strengthen their applications to graduate programs or résumés for entering the workplace

Mentors

1. Are able to observe their research student over an extended period of time and have multiple opportunities to familiarize themselves with student work
2. Are able to make more consistent and reliable assessments of their students' academic strengths and weaknesses
3. Are able to focus mentoring efforts on specific areas where students may need extra guidance, thereby making the research more productive

Undergraduate Research Directors

1. Obtain support for campus assessment efforts
2. Are provided with statistical analyses of assessment score data that constitute a highly reliable and explicit portrait of student growth in knowledge and skills across a wide range of outcomes. The data demonstrate the impact of undergraduate research to campus administrators and/or external funding source(s)
3. Are provided with evidence that can be used to present their program impacts/outcomes at professional meetings and can be published in journals

objectivity and without posing an undue burden for participating faculty and students. It can therefore meet the need for reliable student outcome data and also make a more powerful and immediate contribution to student learning than typical independent program evaluation can ordinarily accomplish. Based on evaluation findings of EvaluateUR in 2019, these benefits are confirmed by students, mentors, and UR directors.

15.3 Looking Forward: Course-Based Undergraduate Research Assessment

In both the United States and Europe, the one-on-one student–mentor UR model is being expanded to embrace UR conducted within the context of courses attended by many students (Healey & Jenkins, 2009; Mieg, 2019). These activities have come to be known as course-based UR experiences (CURE) (Karukstis & Elgren, 2007; Shanahan, 2012). In reality there is often no clear distinction between the one-on-one student–mentor UR model and the CURE model, because many student–mentor UR programs are also

https://doi.org/10.1017/9781108869508.021 Published online by Cambridge University Press

linked to courses and/or degree programs both in the United States and other countries.

The CURE expansion in the United States is designed to make the benefits of UR available to more students and to embed UR in courses earlier (i.e., before thesis or capstone studies) in students' undergraduate degree programs (Rodenbusch et al., 2016; Schinske et al., 2017; Stanford et al., 2017). Faculty in biological sciences disciplines have been most prolific in designing research-centric curricula for introductory life science courses (Dolan, 2016; Jordan et al., 2014; Shaffer et al., 2010). Also, Willison (2009) and colleagues at the University of Adelaide in Australia have created the Research Skill Development (RSD) framework, which identifies sequential skill levels developed through UR and provides templates and self-report strategies to help assess these skills for a variety of disciplines.

As the CURE movement gains momentum, it is generating new assessment approaches designed to understand modifications to mentorship processes dictated by the transition from the one-on-one student–mentor UR model to UR within classrooms. In particular, assessment designs have had to be modified in order to measure ways in which UR within a classroom context has been able to maintain a focus on the student skills and competencies that the student–mentor model has excelled in nurturing (Nadelson et al., 2010; Wuetherick et al., 2018). One line of current inquiry is to undertake surveys or collect case studies of CURE practices across various disciplines (Mieg, 2019; Zimbardi & Myatt, 2014) in order to determine what patterns of CURE deployment exist. A particular challenge for assessment strategies has been the need to assess students' cognitive abilities and skill acquisition (a relatively straightforward task for mentors overseeing individual UR students) when UR is conducted within a classroom context (Auchincloss et al., 2014; Shortlidge & Brownell, 2016).

The success of EvaluateUR has also provided insights into ways in which this method can be adapted to the evaluation of CUREs, which would potentially greatly enlarge the number of students and faculty members who could benefit from the use of the EvaluateUR method. A project designed to test these ideas, known as EvaluateUR-CURE, is currently underway with funding from the United States National Science Foundation's ATE (Advanced Technological Education) program. In addition, the ATE program is supporting the adaptation of the core ideas of EvaluateUR to the evaluation of students participating in regional and international underwater ROV (remotely operated vehicle) competitions. This effort is a collaboration between EvaluateUR and the Marine Advanced Technology Education (MATE) Center in Monterey, California, and is known as EvaluateUR-Compete. Additional information about EvaluateUR, EvaluateUR-CURE, and EvaluateUR-Compete can be found on the EvaluateUR website (https://serc.carleton.edu/evaluateur).

References

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., Lawrie, G., McLinn, C. M., et al. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE–Life Sciences Education*, *13*(1), 29–40. https://doi.org/10.1187/cbe.14-01-0004

Bauer, K. W., & Bennett, J. S. (2003), Alumni perceptions used to assess undergraduate research experience. *Journal of Higher Education*, *74*(2), 210–230.

Billing, D. (2007). Teaching for transfer of core/key skills in higher education: Cognitive skills. *Higher Education*, *53*, 483–516.

Blockus, L. (2012). The challenge of the count. *Council on Undergraduate Research Quarterly*, *32*(3), 4–8.

Brew, A. (2006). *Research and teaching: Beyond the divide*. Palgrave Macmillan.

Crawford, I., Orel, S. E., & Shanahan, J. O. (Eds.). (2014). *How to get started in arts and humanities research with undergraduates*. Council on Undergraduate Research.

Crowe, M., & Brakke, D. (2008). Assessing the impact of undergraduate research experiences on students: An overview of current literature. *Council on Undergraduate Research Quarterly*, *28*(1), 43–50.

Crowe, M., & Brakke, D. (2020). Assessing undergraduate research experiences: An annotative bibliography. *Scholarship and Practice of Undergraduate Research*, *3*(2), 21–30. https://doi.org/10.18833/spur/3/2/3

Crowe, M., Stanford, D., & Shattell, M. (2010). Student and faculty perceptions of the National Conference on Undergraduate Research (NCUR). *Council on Undergraduate Research Quarterly*, *31*(1), 14–17.

Dolan, E. L. (2016). Course-based undergraduate research experiences: Current knowledge and future directions. Committee on Strengthening Research Experiences for Undergraduate STEM Students / National Science Foundation. Retrieved from https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_177288.pdf

Eagan, M. K. Jr., Hurtado, S., Chang, M., Garcia, G. A., Herrera, F. A., & Garibay, J. C. (2013). Making a difference in science education: The impact of undergraduate research programs. *American Educational Research Journal*, *50*(4), 463–713.

Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE–Life Sciences Education*, *17*(1), 1–13. https://doi.org/10.1187/cbe.17-04-0066

Finley, A., & McNair, T. (2013). *Assessing underserved students' engagement in high-impact practices*. Association of American Colleges and Universities.

Gilliams, T., Morris, L., Woodward, K., Rice, K., & Osgood, D. (2008). Models and assessment of collaborative research in the arts and humanities. *Council on Undergraduate Research Quarterly*, *29*(3), 34–37.

Hathaway, R. S., Nagda, B., & Gregerman, S. (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: An empirical study. *Journal of College Student Development*, *43*(5), 614–631.

Healey, M., & Jenkins, A. (2009). *Developing undergraduate research and inquiry*. Higher Education Academy. www.researchgate.net/publication/256208546_Developing_Undergraduate_Research_and_Inquiry

Hensley, M. K., & Johnson, H. R. (2019). The library as collaborator in student publishing: An index and review of undergraduate research journals. *Scholarship and Practice of Undergraduate Research*, *2*(4), 58–67.

Hill, J., & Walkington, H. (2016). Developing graduate attributes through participation in undergraduate research conferences. *Journal of Geography in Higher Education*, *40*(2), 222–237. https://doi.org/10.1080/03098265.2016.1140128

Hogan, M., Dwyer, C., Harney, O., Noone, C., & Conway, R. (2015). Metacognitive skill development and applied systems science: A framework of metacognitive skills, self-regulatory functions and real-world applications. In A. Peña-Ayala (Ed.), *Metacognition: Fundaments, applications, and trends* (pp. 75–106). Springer International Publishing.

Hunter, A. B., Laursen, S. L., & Seymour, E. (2006). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education*, *91*(1), 36–74.

Hurtado, S., Eagen, K., Figueroa, T., & Hughes, B. (2014). Reversing underrepresentation: The impact of undergraduate research programs on enrollment in STEM graduate programs. Retrieved from www.heri.ucla.edu/nih/downloads/AERA-2014-Undergraduate-Research-And-STEM-Grad-Enrollment.pdf

Johnson, D. C., & Gould, C. (2009). Special challenges of assessing undergraduate research in the arts and humanities. *Council on Undergraduate Research Quarterly*, *29*(3), 33–38.

Jordan, T. C., Burnett, S. H., Carson, S., Caruso, S. M., Clase, K., DeJong, R. J., Dennehy, J. J., Denver, D. R., et al. (2014). A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *mBio*, *5*(1), e01051–13. https://doi.org/10.1128/mbio.01051-13

Karukstis, K. K., & Elgren, T. E. (Eds.). (2007). *Designing and sustaining a research-supportive curriculum: A compendium of successful practices*. Council on Undergraduate Research.

Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges and Universities.

Laursen, S. L. (2015). Assessing undergraduate research in the sciences: The next generation. *Council on Undergraduate Research Quarterly*, *35*(3), 9–14.

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

Lopatto, D. (2010). *Science in solution: The impact of undergraduate research on student learning*. Research Corporation for Science Advancement.
Mieg, H. A. (Ed.). (2019). *Inquiry-based learning – Undergraduate research: The German multidisciplinary experience*. Springer (open access). https://doi.org/10.1007/978-3-030-14223-0
Nadelson, L., Walters, L., & Waterman, J. (2010). Course-integrated undergraduate research experiences structured at different levels of inquiry. *Journal of STEM Education: Innovations and Research*, *11*(1–2), 27–44.
Nagda, B. A., Gregerman, S. R., Jonides, J., von Hippel, W., & Lerner, J. S. (1998). Undergraduate student–faculty research partnerships affect student retention. *Review of Higher Education*, *22*(1), 55–72. https://doi.org/10.1353/rhe.1998.0016
Panadero, E., Jönsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98.
Rivera, J., Khelifa, M., Hamdah, B. A., Al-Hamadi, A. M., & Zdgiebloski, E. S. (2018). A global conversation: Reflections from the first World Congress on Undergraduate Research. *Scholarship and Practice of Undergraduate Research*, *2*(1), 55–59. https://doi.org/10.18833/spur/2/1/4
Rodenbusch, S., Hernandez, P. R., Simmons, S. L., & Dolan, E. L. (2016). Early engagement in course-based research increases graduation rates and completion of science, engineering and mathematics degrees. *CBE–Life Sciences Education*, *15*(2), article 20. https://doi.org/10.1187/cbe.16-03-0117
Scharff, L., Draeger, J., Verpoorten, D., Devlin, M., Dvorakova, L., Lodge, J., & Smith, S. (2017). Exploring metacognition as support for learning transfer. *Teaching & Learning Inquiry*, *5*(1), 1–14.
Schinske, J., Balke, V. L., Bangera, G., Bonney, K. M., Brownell, S. E., Carter, R. S., Curran-Everett, D., Dolan, E. L., et al. (2017). Broadening participation in biology education research (BER): Engaging community college students and faculty. *CBE–Life Sciences Education*, *16*(2), 1–11. www.lifescied.org/doi/10.1187/cbe.16-10-0289
Schmitz, H. J., & Havholm, K. (2015). Undergraduate research and alumni: Perspectives on learning gains and post-graduation benefits. *Council on Undergraduate Research Quarterly*, *35*(3), 15–22.
Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, *36*, 111–139.
Shaffer, C. D., Alvarez, C., Bailey, C., Barnard, D., Bhalla, S., Chandrasekaran, C., Chandrasekaran, V., Chung, H.-M., et al. (2010). The Genomics Education Partnership: Successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE–Life Sciences Education*, *9*(1), 55–69. https://doi.org/10.1187/09-11-0087
Shanahan, J. (2012). Building undergraduate research into the curriculum. In N. Hensel & E. Paul (Eds.), *Faculty support and undergraduate research:*

Innovation in faculty role definition, workload, and reward (pp. 68–76). Council on Undergraduate Research.

Shortlidge, E. E., & Brownell, S. E. (2016). How to assess your CURE: A practical guide for instructors of course-based undergraduate research experiences. *Journal of Microbiology and Biology Education*, *17*(3), 399–408. www.asmscience.org/content/journal/jmbe/10.1128/jmbe.v17i3.1103

Singer, J., & Weiler, D. (2009). A longitudinal student outcomes evaluation of the Buffalo State College summer undergraduate research program. *Council on Undergraduate Research Quarterly*, *29*(3), 20–25.

Singer, J., & Zimmerman, B. (2012). Evaluating a summer undergraduate research program: Measuring student outcomes and program impact. *Council on Undergraduate Research Quarterly*, *32*(3), 40–47.

Spronken-Smith, R., Brodeur, J., Kajaks, T., Luck, M., Myatt, P., Verburgh, A., Walkington, H., & Wuetherick, B. (2013). Completing the research cycle: A framework for promoting dissemination of undergraduate research and inquiry. *Teaching & Learning Inquiry*, *1*(2), 105–118.

Stanford, J. S., Rocheleau, S. E., Smith, K. P. W., & Mohan, J. (2017). Early undergraduate research experiences lead to similar learning gains for STEM and non-STEM undergraduates. *Studies in Higher Education*, *42*(1), 115–129. https://doi.org/10.1080/03075079.2015.1035248

Vandermaas-Peeler, M., Miller, P., & Moore, J. (Eds.). (2018). *Excellence in mentoring undergraduate research*. Council on Undergraduate Research.

Willison, J. W. (2009). Multiple contexts, multiple outcomes, one conceptual framework for research skill development in the undergraduate curriculum. *Council on Undergraduate Research Quarterly*, *29*(3), 10–14.

Willison, J., Zhu, X., Xie, B., Yu, X., Chen, J., Zhang, D., Shahshoug, I., & Sabir, F. (2020). Graduates' affective transfer of research skills and evidence-based practice from university to employment in clinics. *BMC Medical Education*, *20*, article 89. https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909–020-1988-x

Wolters, C., & Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning*, *10*(3), 293–311.

Wuetherick, B., Willison, J. W., & Shanahan, J. (2018). Mentored undergraduate research at scale: Undergraduate research in the curriculum and as pedagogy. In M. Vandermaas-Peeler, P. Miller, & J. Moore (Eds.), *Excellence in mentoring undergraduate research* (pp. 181–202). Council on Undergraduate Research.

Zimbardi, K., & Myatt, P. (2014). Embedding undergraduate research experiences within the curriculum: A cross-disciplinary study of the key characteristics guiding implementation. *Studies in Higher Education*, *39*(2), 233–250. www.tandfonline.com/doi/abs/10.1080/03075079.2011.651448