

Joanna Gilmore  
Michelle Vieyra  
Briana Timmerman  
David Feldon  
JE Michelle Maher

---

## The Relationship between Undergraduate Research Participation and Subsequent Research Performance of Early Career STEM Graduate Students

*Undergraduate research experiences have been adopted across higher education institutions. However, most studies examining benefits derived from undergraduate research rely on self-report of skill development. This study used an empirical assessment of research skills to investigate associations between undergraduate research experiences and research skill performance in graduate school. Research experience characteristics including duration, autonomy, collaboration, and motivation were also examined. Undergraduate research experience was linked to heightened graduate school performance in all research skills assessed. While autonomy and collaboration were highlighted in student interviews, duration was most strongly correlated to significant increases in research skill performance. Based on these findings, we advocate for the inclusion of research experiences into the undergraduate science curriculum coupled with the creation of centralized offices of undergraduate research and faculty incentives for involving undergraduates in their research.*

*Keywords: undergraduate research, STEM*

Government agencies generally agree that the United States is experiencing a shortage of professionals trained in science, technology, mathematics, and engineering (STEM) (National Academy of Sciences

---

*Joanna Gilmore is a Research Associate with the Charleston County School District, Charleston, SC; Joanna.Gilmore1981@gmail.com. Michelle Vieyra is Associate Professor, Department of Biology at the University of South Carolina–Aiken; MichelleV@usca.edu. Briana Timmerman is a Research Associate Professor in the Department of Biological Sciences, University of South Carolina, and Director of Development for Advanced Career at the Southern Regional Education Board; Briana.Timmerman@gmail.com. David Feldon is Associate Professor, Department of Instructional Technology and Learning Sciences, and director of the USU STE<sup>2</sup>M Center at Utah State University; david.feldon@usu.edu. Michelle Maher is Professor of Higher Education Administration, University of Missouri-Kansas City, MO; mahermi@umkc.edu.*

*The Journal of Higher Education*, Vol. 86, No. 6 (November/December 2015)  
Copyright © 2015 by The Ohio State University

[NAS], 2007a; 2007b; President's Council of Advisors on Science and Technology, 2012; State Educational Technology Director's Association, 2008; For a discussion of the debate around the shortage of STEM professionals, see National Science Board, 2015). Retaining STEM undergraduates to baccalaureate graduation, and thus increasing the number of degree holders eligible to pursue STEM graduate degrees, appears critical in addressing these differentials. Discussions about how to increase STEM retention stress the importance of improving the quality of undergraduate teaching (e.g., Fairweather, 2008; Feldon, Timmerman, Stowe, & Showman; 2010; NAS, 2007a; National Research Council, 2000; Seymour & Hewitt, 1997; Stage & Kinzie, 2009). One proposed method to improve STEM instruction is the inclusion of research experiences in the undergraduate curriculum. The Council on Undergraduate Research noted that undergraduate research is the "pedagogy of the 21<sup>st</sup> century" (2005, n.p.) and the Association of American Colleges and Universities identified undergraduate research as one of ten "high impact education experiences" (2008, p. 1). Further, many colleges and universities have incorporated research experiences into their STEM undergraduate curricula (Blanton, 2008; Kinkead, 2003; Strasburger, 1995).

Extant research has linked undergraduate research participation with several positive educational outcomes, including increased STEM retention, rates of STEM graduate school matriculation, baccalaureate cumulative grade point average, and receipt of national awards (e.g., Bauer & Bennett, 2003; Craney, McKay, & Morris, 2011; Foertsch, Alexander, & Penberthy; 2000; Jones, Barlow, & Villarejo, 2010; Nagda, Gregerman, Jonides, von Hippel, & Lerner, 1998; Summers & Hrabowski, 2006). As well, undergraduate research participation has been associated with enhanced ability to "think like a scientist" and/or "do science," intellectual curiosity, oral communication, professional writing, research skills, and academic skills (Hunter, Laursen, & Seymour, 2007; Kardash, 2000; Russell, Hancock, & McCullough, 2007; Seymour, Hunter, Laursen, & DeAntoni, 2004). Commitment to STEM-related career and educational goals has also been linked to undergraduate research participation (Hunter, Laursen, & Seymour, 2007; Thiry, Laursen, & Hunter, 2011).

While valuable, these studies rely primarily on student self-report of skill development, which may be problematic, because students' estimations of their own abilities and skill improvement are typically inconsistent with performance-based evidence (Bowman, 2010; Dunning, Johnson, Ehrlinger, & Kruger, 2003; Feldon, Maher, Hurst, & Timmerman, 2015). For this reason, Kardash (2000) compared student self-ratings with faculty ratings of students' research skills to provide a more ob-

jective measure. Although this represents an improvement from studies that rely exclusively on student self-report of learning, even faculty perceptions of students' research skills can be suspect (Feldon et al., 2015).

Given the resources dedicated to implementing and sustaining undergraduate research participation across an increasing number of higher education institutions, the assessment of associated outcomes deserves closer attention (Blanton, 2008; Hunter et al., 2007). Undoubtedly, the impressive range of benefits perceived as deriving from undergraduate research participation is a testimony to its value. Curiously, however, in a time of decreased educational funding and increased scrutiny of student learning outcomes (Suskie, 2009), efforts to empirically link undergraduate research participation with a primary outcome, subsequent graduate research performance, are scarce to nonexistent. In response, this multi-institutional study empirically examines the relationship between undergraduate research participation and subsequent research performance of early career STEM graduate students.

### **Conceptual Framework**

This study was informed by examining research on skill acquisition, studies specific to undergraduate student research, and theories regarding the social nature of knowledge construction. Research on skill acquisition (Ericsson & Charness, 1994; Timmerman, Feldon, Maher, Strickland, & Gilmore, 2013) suggests that the development of any new skill requires deliberate practice over time, and the skill level achieved is directly linked to the duration of that experience. Studies on undergraduate research also highlight that duration of the research experience is important in understanding subsequent outcomes (Fechheimer, Weber, & Kleiber, 2010; Jones et al., 2010; Russell et al., 2007; Zydney, Bennett, Shahid, & Bauer, 2013).

In addition to undergraduate research experience duration, student motivation for conducting research may also help explain outcomes. Our interest in studying motivation emerged from Russell's (2008) assertion that forcing undergraduates to conduct research may be counterproductive. However, Vieyra, Gilmore, and Timmerman (2011) examined the self-reported benefits of participation in required undergraduate research in biology and found that all students perceived the experience as valuable, even those who indicated they would not have participated without the mandate. Within the current study, we recruited students who conducted undergraduate research as part of degree requirements and those who self-selected to conduct research. The current effort extends our earlier work by allowing a comparison between mandated

and self-selected undergraduate participation. The current multi-institutional study also includes a larger student sample and represents more disciplines.

Researchers frequently suggest the potency of undergraduate research participation springs from its theoretical underpinning in which knowledge creation is socially and culturally situated (e.g., Hunter, Laursen, & Seymour, 2007; Kardash, 2000; Vandermaas-Peeler, Nelson, Ferretti, & Finn, 2011). Thus, Brown and colleagues' (Brown, Collins, & Duguid, 1989; Collins, Brown, & Newman, 1989) cognitive apprenticeship model and Lave and Wegner's (1991) community of practice model, both of which theoretically situate knowledge creation within social and cultural realms, frame this study.

Within a cognitive apprenticeship model, apprentices (in this case, undergraduate researchers) gain disciplinary knowledge and skills through close interaction with recognized disciplinary experts (faculty mentors and relevant others, such as advanced peers) as they collaboratively address authentic disciplinary problems (Gafney, 2005; Lopatto, 2003; Pearson & Brew, 2002; Russell et al., 2007). The emphasis in this model is on "cognitive" because, as Walker, Golde, Jones, Conklin Bueschel and Hutchings (2008) note, the model "focuses on intellectual skills and practices, because it makes thought visible through formal representations (talking, writing, mathematical equations and the like), and because it expects teachers and learners to be thinking explicitly about what they're doing" (p. 109).

As Lopatto (2003) indicates, effective mentoring of STEM undergraduate research experiences is critical because students "learn from the mentor how scientists think, how obstacles are tolerated and how a career path develops" (p. 141). The faculty mentor and advanced others make thought visible by modeling disciplinary-appropriate thinking patterns and behaviors. Further, they scaffold the undergraduate's efforts around authentic disciplinary tasks, successively developing the student's ability to complete tasks independently. This involves the student learning how to design aspects of his or her research studies, identify and solve problems, and take ownership of his or her work (Lopatto, 2003). The end goal of a cognitive apprenticeship is achieved when the student develops and demonstrates autonomy in their research endeavors. In support of this model, the literature on undergraduate research highlights the importance of developing student autonomy (Burke & Cummins, 2010; Hefferan, Heywood, & Ritter, 2002; Lopatto, 2003).

The practice of STEM research is increasingly collaborative and cross-disciplinary (Fox, & Mohapatra, 2007; B. F. Jones, Wuchty, & Brian, 2008), and learning how to effectively interact in a research

community is critical to researcher development. During undergraduate research experiences, the student often engages with others as a peripheral member (Lave & Wenger, 1991) of the small, specialized scholarly community of a research lab. In this capacity, the student/apprentice learns the operational procedures and disciplinary norms under the guidance of faculty, graduate students, and postdoctorates associated with the particular lab (Finlay & Faulkner, 2005; Waite & Davis, 2011). Often undergraduates may interact with a wide network of researchers, and this may have implications for their skill development. For example, undergraduates in Waite and Davis' (2011) study indicated that interacting with peers around their research exposed them to more diverse perspectives. This might help these students learn to situate and communicate their research.

As the size of an undergraduate's research network increases, Lave and Wenger's (1991) community of practice may be a better conceptualization of the undergraduate research experience. Lave and Wenger propose that learning occurs through active participation in social communities through which we reconstruct our identity. From this point of view, the most productive learning experiences are those that engage students in "actions, discussions, and reflections that make a difference to the communities they value" (Wenger, 1998, p. 10). Thus, both the cognitive apprenticeship and community of practice model emphasize the importance of collaboration and autonomy during the undergraduate research experience.

### **Study Purpose, Research Questions, and Hypotheses**

As noted above, extant research has linked undergraduate research participation with several positive educational outcomes. Prior research, however, has focused on retention outcomes (e.g., Nagda et al., 1998) or has relied exclusively on self-report of skill development (e.g., Bauer & Bennett, 2003) to evaluate gains from undergraduate research participation. Missing from this line of inquiry are studies linking undergraduate research participation with subsequent graduate student research skill performance using an empirical assessment of students' research skills. The current study examines the relationship between undergraduate research participation and early career graduate students' research skill performance. The research questions and hypotheses that guided this study include:

- Research Question 1: To what extent is participation in undergraduate research related to research skill performance at the beginning and end of a student's first year of graduate school?

- Hypothesis 1: We predict significantly better research skill performance among students with undergraduate research experiences than among students with no undergraduate research participation based on exploratory literature indicating that undergraduate research participation contributes to research skill development (Bauer & Bennett, 2003; Hunter, et al., 2007).
- Research Question 2: In terms of undergraduate research participation, to what extent are duration of undergraduate research, degree of autonomy, collaboration/research network size, and motivation related to research skill performance?
  - Hypothesis 2a: We predict that duration of undergraduate research participation will be positively associated with research performance. This prediction is informed by Fechheimer, Weber, and Kleiber (2010) who found that duration of undergraduate research was positively related to student grade point averages.
  - Hypothesis 2b: We predict that students with higher levels of autonomy during undergraduate research participation will demonstrate higher research skill performance, as prior researchers have suggested that autonomy is important in learning from undergraduate research participation (Burke & Cummins, 2010; Hefferan, Heywood, & Ritter, 2002; Lopatto, 2003).
  - Investigations of the relationships between graduate research skill performance and undergraduate research participation collaboration/research network size and motivation for undergraduate research participation are exploratory in nature and thus not informed by prior studies. As such, we do not offer specific hypotheses about the outcomes of these investigations.

## **Method**

### *Context*

All data were gathered as part of a larger NSF-funded project examining the development of graduate students' teaching and research skills. Researchers recruited students from STEM programs at three universities. Two universities are located in the southeastern United States, including a research-extensive university (77.6% of the study sample) as designated by the Carnegie Foundation for the Advancement of Teaching (2014) and a primarily baccalaureate college of arts and sciences with a master's degree program (10.3%). The third, a larger master's-granting university (12.1%), is located in the northeastern United States. The institutions selected represent three different contexts: a flagship doctoral institution, a small masters-granting institution

in the Southeast, and a large masters granting institution in the Northeast. All participants came from research-intensive programs regardless of their institution; students were involved in research from the start of their graduate program and were expected to complete empirical theses/dissertations. Homogeneity of variance within the sample indicates that properties observed reflect effects of undergraduate research and not institutional factors.

### *Participants*

The NSF Primary Investigators (who include the third, fourth, and fifth authors of this study) either worked in STEM fields at one of the universities or established relationships with STEM departmental faculty. The Primary Investigators shared information about the opportunity to participate in the study with these contacts who, in turn, shared the information with their graduate students who represent the study participants.

All 58 participants were first-year graduate students. All were seeking their master's degree ( $n = 25$ , 43.1%) or Ph.D. ( $n = 33$ , 56.9%). All but two (96.6%) were Graduate Research Assistants at the time of the study. Forty students (69.0%) pursued a science-related degree (e.g., biology, chemistry), 5 (8.6%) pursued a degree in technology (e.g., biotechnology, computer science), 11 (19.0%) sought an engineering degree (e.g., mechanical or chemical engineering), and 2 (3.4%) pursued degrees in mathematics or statistics. Twenty-four students (41.4%) were female. Thirty students (51.7%) self-identified as nonnative English speakers. These percentages are generally representative of the overall graduate student population at the respective institutions. Each student received \$500 as compensation for participation.

### *Data Collection*

Data collection was undertaken to assess students' research skills at the beginning and end of their first academic year of graduate school and to gather descriptive information about students' undergraduate research experiences. Research skills were assessed by applying a rubric to student research proposals. Upon recruitment at the beginning of the academic year, students were directed to write research proposals in their field of study. They were provided with descriptions of proposal evaluation criteria, writing prompts, and desired section headings (e.g., *Introduction*, *Methods*), and were instructed to use resources and citation styles typical for their field. The proposal was framed as an effort that could be directly applicable to NSF Graduate Fellowship applications or students' eventual thesis or dissertation proposals. Students sub-

mitted their research proposal within three weeks after starting their first semester of graduate study (preproposal); they revised their proposals at the end of their first year of graduate study (postproposal).

Proposals were evaluated using a previously validated rubric for assessing scientific reasoning skills through writing (Timmerman, Strickland, Johnson, & Payne, 2011). This rubric represents a consensus among STEM experts on acceptable research proposals. In the current study, we use the rubric to measure our dependent variable of research skill performance. The rubric has been validated for this purpose, however, we do not advocate for using the rubric to gauge performance of an individual researcher. Two trained raters with expertise in STEM fields independently rated each proposal using the rubric. Intraclass correlation coefficients were computed to assess inter-rater reliability and were high (0.6–0.9) (Feldon et al., 2011; also see supplemental online materials accompanying Feldon et al., 2011). Discrepancies were resolved by discussion until consensus was reached (Johnson, Penny, & Gordon, 2000; Johnson, Penny, Gordon, Schumate, & Fisher, 2005).

The rubric includes 12 components, 10 of which were examined in this study.<sup>1</sup> For example, the rubric plank delineating criteria characterizing quality of data presentation (which is one of the 12 components) stated:

Expected or preliminary data are summarized and presented in a logical format. A rough picture of the anticipated results should be provided that include, as appropriate, tables or graphs to show anticipated trends. Quantitative data should be presented using appropriate unit labels. If graphs are used, axes are appropriately labeled and scaled, and captions are informative and complete.

Rubric components can be aggregated into four subscales. For example, the *Context Subscale* includes three components: *Introduction and Context*, *Primary Literature*, and *Testable Hypothesis*. Total scores were computed by summing all component scores. To ensure that proposals accurately reflected each student's own academic work, they were submitted to *SafeAssign*© plagiarism detection software. Papers with notable plagiarism issues were returned to students for revision and resubmission (see Gilmore, Strickland, Timmerman, Maher, & Feldon, 2010 for details).

Researchers gathered information about students' undergraduate research experiences through semistructured interviews conducted at the beginning of the graduate students' first semester of study. Interviews were conducted with all 58 study participants and generally lasted about

one hour in length. Interviews addressed students' prior and anticipated teaching and research experiences. Later in the NSF study, when our interest in studying undergraduate research emerged, those who had participated in undergraduate research were asked to describe their participation in more depth. For example, they were asked: "Tell me what your undergraduate research experience was like," "How long did you participate in undergraduate research," "Was your undergraduate research collaborative," "To what extent did you have control over your undergraduate research," and "How did doing undergraduate research impact you?" Because our interest in undergraduate research emerged after a year of data had already been collected for the larger NSF project, some participants in this study were not directly asked these questions.

### *Data Analysis*

This study used an explanatory sequential mixed methods design (Creswell & Plano Clark, 2007) in which quantitative data (proposal scores) were first analyzed and qualitative data (student descriptions of undergraduate research experiences and the coding of those data) were then used to explore mechanisms underlying quantitative results. Specifically, quantitative analyses were first conducted to identify differences in research skill performance between students who had and who had not participated in undergraduate research. Qualitative analyses were then conducted to identify and explore emergent patterns in characteristics of undergraduate research participation (e.g., level of autonomy) associated with observed differences in research skill performance between the groups. Qualitative data were quantitized to examine relationships between undergraduate research characteristics and research skill performance. Because interviews were conducted as part of the larger study which did not focus exclusively on undergraduate research experiences, transcript data were not rich enough to perform in-depth qualitative analysis.

Quantitization of interview data began when the first and second author independently reviewed relevant interview transcript portions to determine if a student had conducted undergraduate research. Exact inter-rater agreement was computed to be 100.0%. Mann-Whitney tests then were used to identify significant differences in research skill performance between students with and without undergraduate research experience. The Mann-Whitney test is appropriate as it does not assume a normal distribution and is suitable for ordinal data analysis. A one-tailed *p*-value was used to examine differences in research skill performance between students with and without undergraduate research experiences because exploratory literature supports the supposition that undergradu-

ates with research experience demonstrate stronger research skill performance (Bauer & Bennett, 2003; Hunter, et al., 2007).

### *Definitions and Coding of Influencing Factors*

We also sought to investigate the influence of key characteristics of undergraduate research participation on subsequent graduate research skill performance. The first and second authors reviewed students' descriptions of their undergraduate research experience to identify information related to four factors (duration, autonomy, collaboration, and motivation). As previously mentioned, in-depth qualitative analyses of these characteristics was not possible as our data set was limited because we did not anticipate conducting this study when data were originally collected. Although in-depth qualitative analyses were not viable, the categories we used to quantitize interview data were developed inductively through multiple reviews of the data.

We defined duration as the number of semesters that students reported conducting undergraduate research, with one semester coded as "1," two semesters coded as "2," and so forth. When students recorded a summer research experience, we coded the description as one semester. However, we recognize that this is a rough measure of duration as some summer research experiences, such as NSF's Research Experiences for Undergraduates, involve intensive, full-time work (e.g., SJB Research Consulting Inc., 2006). This also means that students could have completed three semesters of research experience within one calendar year. Student descriptions that reported no undergraduate research experience were coded "0."

Autonomy concerns whether a student reported some control in research problem selection or methodological decision-making during undergraduate research. If a student's description indicted some control over these it was coded as having some autonomy. If a student's description indicated no decision-making authority it was coded as not having autonomy. If not enough data were available to determine if the student did or did not experience autonomy during their undergraduate research, we coded the description as missing data (this was also the case for the collaboration and motivation variables that will be described next).

Collaboration was operationalized as research network size, although we realize that this represents only one component of collaboration. If a student described collaborating solely with a mentor, the description was coded as such. If a student indicated interacting with individuals other than their mentor around his or her undergraduate research, the description was coded as experiencing a wider collaborator network.

We defined motivation as whether students reported that they voluntarily conducted undergraduate research or were required to do so by their undergraduate degree programs. If they reported they voluntarily chose to conduct undergraduate research, their descriptions were coded as self-selecting to do research. If they reported that they engaged in undergraduate research to fulfill degree requirements, such as an honors thesis, their descriptions were coded as being required to conduct undergraduate research (although the argument could be made that these students are also self-selecting because they chose the institution and degree they pursued). For each of these variables, the coding scheme only provides a rough measure of the nature of participants' undergraduate research. In reality, each of these variables could be better measured along a spectrum of experience. However, the granularity of our dataset only allowed us to reliably code participants' data along less-refined categories.

We independently coded all transcripts with respect to these four factors. Inter-rater agreement was generally high: 82.5% for degree of autonomy, 75.4% for size of their research network, and 75.4% for motivation for conducting undergraduate research. A Pearson's correlation of  $r = 0.936$  was computed to assess inter-rater reliability for duration of undergraduate research experience because it is a continuous rather than a dichotomous variable. All discrepancies were resolved by discussion until consensus was reached (Johnson et al., 2000; Johnson et al., 2005). In addition to coding students' transcripts with respect to these four variables (duration, autonomy, collaboration, and motivation), we examined transcripts for evidence that students perceived these factors as facilitating and/or hindering their development as researchers.

Upon coding completion, we computed Pearson's correlation coefficients to examine the relationship between duration of students' undergraduate research experience and graduate research skill performance. A one-tailed  $p$ -value was used for this analysis as, based on results from the work of Fechheimer, Webber, & Kleiber (2010), we predicted that longer durations of undergraduate research experience would be related to stronger graduate research skill performance. To examine the relationship between the three other variables (autonomy, collaboration, and motivation) and graduate research skill performance, we divided students' data into quartiles based on students' total preproposal scores. We then reviewed interview data associated with these three variables to identify emergent patterns. For example, we sought to determine if students in the highest performing quartile more commonly reported having autonomy during their undergraduate research, and we sought to determine if autonomy was characterized differently for students in the various performance quartiles.

## Results

### *Participation in Undergraduate Research and Relationship to Graduate Research Skill Performance*

Forty-seven of 58 (81.0%) students reported participating in undergraduate research. Tables 1 and 2 present students' scores on the preproposal and postproposal by presence or absence of undergraduate research involvement. On their preproposal, students with undergraduate research experience outscored those without this experience on all rubric components except *Testable Hypothesis*. These differences were statistically significant for *Data Presentation* ( $p = .011$ ), *Results Subscale* ( $p = .016$ ), and *Total Score* ( $p = .026$ ). On their postproposal, with the exception of *Introduction and Context*, students with undergraduate research experience outscored those without this experience on all rubric components as well as on total scores. However, only the *Discussion Subscale* ( $p = .028$ ) met the  $p < .05$  threshold for significance for the postproposal (Table 2). To contextualize these data, we examined participation in undergraduate research by participants' status as a native/nonnative English speaker as this represented a substantial portion of our sample (51.7%). Twenty-one of 30 (70.0%) English language learners conducted undergraduate research compared with 26 of 28 (92.9%) native English speakers.

### *Undergraduate Research Characteristics and Their Relationship to Research Skills*

Informed by the literature, we examined the influence of four undergraduate research characteristics' including (1) duration of undergraduate research experience, (2) degree of students' autonomy in conducting undergraduate research, (3) size of students' research network, and (4) students' motivation for conducting undergraduate research. Factors were examined with respect to the quartiles (which divide the student data according to students' preproposal total scores) to identify any relationships between each of these characteristics of undergraduate research and research skill outcomes. Analytic results concerning each factor are discussed in turn.

**Duration.** Duration of undergraduate research experience refers to the length of time student engaged in undergraduate research. Often, students reported duration in number of semesters. At times, however, students described their undergraduate research in terms of the number of years that they were involved. The researchers assumed that a year referred to an academic year (equivalent to two semesters). Data regarding duration were reported by 36 of the 47 (76.6%) students who participated in undergraduate research. The median number of semesters

TABLE 1

## Preproposal Mean Component, Subscale, and Total Scores by Involvement in Undergraduate Research

|                            | Did Not Engage in Undergraduate Research |        |       | Engaged in Undergraduate Research |        |       | Z      | p      | 1 - $\beta$ |
|----------------------------|--|--------|-------|-----------------------------------|--------|-------|--------|--------|-------------|
|                            | N  | Mean   | SD    | N                                 | Mean   | SD    |        |        |             |
| Introduction and Context   | 11                                       | 1.455  | 0.660 | 47                                | 1.596  | 0.671 | 0.573  | 0.283  | 0.156       |
| Primary Literature         | 11                                       | 1.045  | 0.757 | 47                                | 1.457  | 0.740 | 1.347  | 0.089  | 0.495       |
| Testable Hypothesis        | 11                                       | 1.250  | 0.447 | 47                                | 1.229  | 0.541 | 0.021  | 0.492  | 0.065       |
| <i>Context Subscale</i>    | 11                                       | 3.750  | 1.369 | 47                                | 4.282  | 1.657 | 0.805  | 0.211  | 0.297       |
| Validity and Reliability   | 10                                       | 1.100  | 0.543 | 46                                | 1.223  | 0.655 | 0.420  | 0.338  | 0.154       |
| Experimental Design        | 11                                       | 1.523  | 0.541 | 47                                | 1.548  | 0.636 | -0.111 | 0.456  | 0.065       |
| Data Selection             | 11                                       | 1.205  | 0.590 | 47                                | 1.484  | 0.734 | 1.412  | 0.079  | 0.382       |
| <i>Methods Subscale</i>    | 10                                       | 3.725  | 1.244 | 46                                | 4.234  | 1.642 | 0.718  | 0.237  | 0.294       |
| Data Presentation          | 10                                       | 0.550  | 0.587 | 46                                | 1.288  | 0.960 | 2.304  | 0.011* | 0.935       |
| Data Analysis              | 10                                       | 0.775  | 0.381 | 46                                | 1.016  | 0.680 | 0.964  | 0.168  | 0.457       |
| <i>Results Subscale</i>    | 10                                       | 1.325  | 0.808 | 46                                | 2.304  | 1.439 | 2.147  | 0.016* | 0.904       |
| Conclusions Based on Data  | 11                                       | 0.795  | 0.430 | 47                                | 1.149  | 0.787 | 1.538  | 0.062  | 0.655       |
| Discussion of Limitations  | 11                                       | 0.523  | 0.607 | 47                                | 0.798  | 0.889 | 0.791  | 0.215  | 0.338       |
| <i>Discussion Subscale</i> | 11                                       | 1.318  | 0.822 | 47                                | 1.947  | 1.433 | 1.214  | 0.113  | 0.616       |
| Total Score                | 10                                       | 10.100 | 2.193 | 46                                | 12.685 | 4.938 | 1.948  | 0.026* | 0.823       |

*Note.* All components were measured on a scale of 0 to 3. Zero indicates an absence of the targeted skill within the proposal and 3 indicates that the skill was proficiently demonstrated. Equal variances not assumed in calculation of test statistic and  $p$ -value. Research proposals in the area of mathematics are not evaluated for validity and reliability, data presentation, and data analysis as these criteria were determined to be inappropriate for math. This resulted in a lower potential total scores for students in math, thus they have been excluded from analyses conducted with total research proposal scores.  $Z$  represents the test statistic resulting from the Mann-Whitney test and  $p$  represents the associated one-tailed probability that there are no group differences.  $1 - \beta$  represents the statistical power of the test to correctly reject the null hypothesis. 0.80 is generally considered to be adequate statistical power to detect group differences. Sample sizes differ from pre to postproposal because of study attrition.

\* significant at  $p \leq .05$ .

TABLE 2  
 Postproposal Mean Component, Subscale, and Total Scores by Involvement in Undergraduate Research

|                            | Did Not Engage in Undergraduate Research |        |       | Engaged in Undergraduate Research |        |       | Z      | p      | 1 - $\beta$ |
|----------------------------|--|--------|-------|-----------------------------------|--------|-------|--------|--------|-------------|
|                            | N  | Mean   | SD    | N                                 | Mean   | SD    |        |        |             |
| Introduction and Context   | 9  | 1.917  | 0.781 | 43                                | 1.785  | 0.609 | -0.674 | 0.250  | 0.122       |
| Primary Literature         | 9  | 1.194  | 1.059 | 43                                | 1.692  | 0.742 | 1.317  | 0.094  | 0.382       |
| Testable Hypothesis        | 9  | 1.278  | 0.404 | 43                                | 1.326  | 0.599 | 0.197  | 0.422  | 0.089       |
| <i>Context Subscale</i>    | 9  | 4.389  | 2.077 | 43                                | 4.802  | 1.526 | 0.521  | 0.301  | 0.140       |
| Validity and Reliability   | 9  | 1.333  | 0.857 | 42                                | 1.512  | 0.767 | 0.674  | 0.250  | 0.143       |
| Experimental Design        | 9  | 1.611  | 0.674 | 43                                | 1.657  | 0.559 | 0.381  | 0.352  | 0.073       |
| Data Selection             | 9  | 1.389  | 0.601 | 43                                | 1.593  | 0.707 | 0.870  | 0.192  | 0.227       |
| <i>Methods Subscale</i>    | 9  | 4.333  | 1.980 | 42                                | 4.768  | 1.756 | 0.891  | 0.187  | 0.150       |
| Data Presentation          | 9  | 1.111  | 0.849 | 43                                | 1.395  | 0.934 | 0.970  | 0.166  | 0.227       |
| Data Analysis              | 9  | 0.972  | 0.655 | 42                                | 1.244  | 0.693 | 1.470  | 0.071  | 0.299       |
| <i>Results Subscale</i>    | 9  | 2.083  | 1.132 | 42                                | 2.625  | 1.459 | 1.190  | 0.117  | 0.340       |
| Conclusions Based on Data  | 9  | 1.000  | 0.729 | 43                                | 1.227  | 0.692 | 0.772  | 0.220  | 0.215       |
| Discussion of Limitations  | 9  | 0.806  | 0.855 | 43                                | 0.983  | 1.738 | 0.088  | 0.465  | 0.117       |
| <i>Discussion Subscale</i> | 9  | 1.806  | 1.402 | 43                                | 2.209  | 1.949 | 1.916  | 0.028* | 0.180       |
| Total Score                | 9  | 12.611 | 5.195 | 42                                | 14.393 | 4.756 | 1.149  | 0.125  | 0.243       |

Note. All components were measured on a scale of 0 to 3. Zero indicates an absence of the targeted skill within the proposal and 3 indicates that the skill was proficiently demonstrated. Equal variances not assumed in calculation of test statistic and p-value. Research proposals in the area of mathematics are not evaluated for validity and reliability, data presentation, and data analysis as these criteria were determined to be inappropriate for math. This resulted in a lower potential total scores for students in math, thus they have been excluded from analyses conducted with total research proposal scores. Z represents the test statistic resulting from the Mann-Whitney test and p represents the associated one-tailed probability that there are no group differences. 1 -  $\beta$  represents the statistical power of the test to correctly reject the null hypothesis. 0.80 is generally considered to be adequate statistical power to detect group differences. Sample sizes differ from pre to postproposal because of study attrition.

\* significant at  $p \leq .05$ .

of undergraduate research reported was two (mean = 2.81,  $SD = 2.00$ ). Figure 1 identifies the length of students' undergraduate research experiences in semesters.

Although length of undergraduate research participation was frequently discussed when students described their undergraduate research experience, no student explicitly connected duration to his or her research skill development. However, when duration of undergraduate research participation and proposal scores were correlated, several significant coefficients were found on both pre and postproposals rubric criteria, including pre and postproposal *Total Scores* ( $p = .024$  and  $.037$ , respectively).

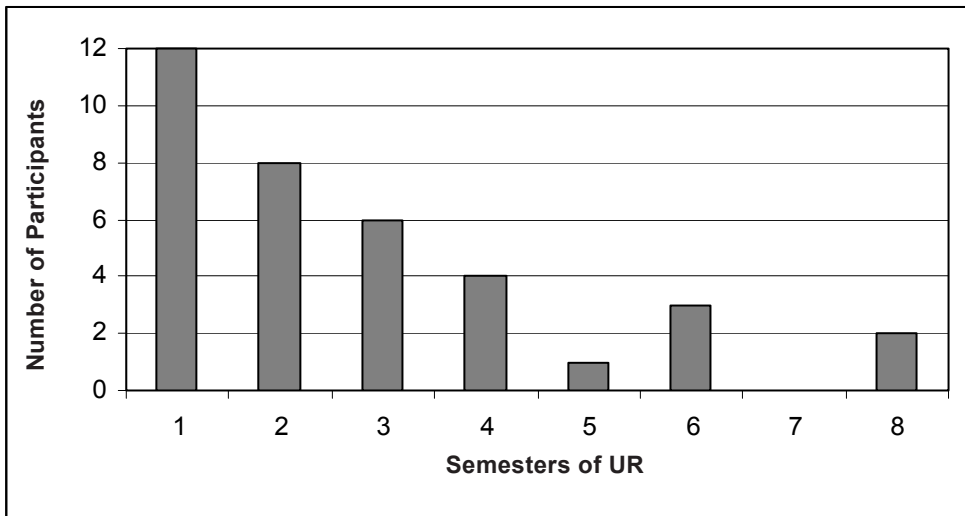


FIGURE 1  
Duration of Students' Undergraduate Research (UR) Experiences

TABLE 3

Correlation between Duration of Undergraduate Research and Component, Subscale, and Total Preproposal and Postproposal Scores

|                            | Preproposal |          |          |              | Postproposal |          |          |              |
|----------------------------|-------------|----------|----------|--------------|--------------|----------|----------|--------------|
|                            | <i>n</i>    | <i>r</i> | <i>p</i> | <i>1 - β</i> | <i>n</i>     | <i>r</i> | <i>p</i> | <i>1 - β</i> |
| Introduction and Context   | 47          | .286     | 0.026*   | 0.627        | 43           | .223     | 0.075    | 0.424        |
| Primary Literature         | 47          | .272     | 0.033*   | 0.598        | 43           | .324     | 0.017*   | 0.703        |
| Testable Hypothesis        | 47          | .153     | 0.153    | 0.268        | 43           | .068     | 0.334    | 0.117        |
| <i>Context Subscale</i>    | 47          | .929     | 0.023*   | 1.000        | 43           | .290     | 0.030*   | 0.621        |
| Validity and Reliability   | 46          | .122     | 0.210    | 0.201        | 43           | .207     | 0.092    | 0.398        |
| Experimental Design        | 47          | .183     | 0.110    | 0.341        | 43           | .172     | 0.135    | 0.297        |
| Data Selection             | 47          | .180     | 0.113    | 0.341        | 43           | .269     | 0.041*   | 0.565        |
| <i>Methods Subscale</i>    | 46          | .222     | 0.069    | 0.445        | 43           | .254     | 0.051    | 0.508        |
| Data Presentation          | 46          | .384     | 0.004*   | 0.864        | 43           | .382     | 0.006*   | 0.842        |
| Data Analysis              | 46          | .045     | 0.383    | 0.096        | 43           | .542     | 0.048*   | 0.994        |
| <i>Results Subscale</i>    | 46          | .309     | 0.022*   | 0.676        | 43           | .277     | 0.031*   | 0.593        |
| Conclusions Based on Data  | 47          | .212     | 0.077    | 0.423        | 43           | .249     | 0.054    | 0.508        |
| Discussion of Limitations  | 47          | -.006    | 0.484    | 0.057        | 43           | -.189    | 0.113    | 0.346        |
| <i>Discussion Subscale</i> | 47          | .461     | 0.055    | 0.968        | 43           | -.074    | 0.319    | 0.117        |
| Total Score                | 46          | .294     | 0.024*   | 0.647        | 43           | .275     | 0.037*   | 0.593        |

Note. The *p*-value presented represents a one-tailed *p*-value.

\* significant at  $p \leq .05$ .

**Degree of Autonomy.** Degree of autonomy concerns the extent to which students reported having control in research problem selection or methodological decision-making during their undergraduate research experience. Examples include autonomously identifying research questions to pursue or selecting data collection or analysis methods to implement. Among the 47 students who reported undergraduate research experience, 27 (57.4%) provided information about their level of autonomy during this experience. Of those, 13 (48.1%) reported having little or no autonomy. These students described a research experience limited to conducting literature reviews or collecting data using a prescribed protocol. For example, one student shared, “I just followed what [my advisor] did. She knew better than I did.”

In contrast, 14 of 27 students (51.9%) who provided information about their level of autonomy reported meaningfully contributing to the study topic, research design, and/or data analysis. For example, one student noted, “[The study] was prescribed in the beginning, but over time I was at the point where I could, of course, I’d have to speak with my advisor, but I was given responsibility to design experiments and things like that.” Students reporting substantial autonomy during their undergraduate research experience tended to be in the lowest and highest preproposal performance quartiles (accounts for 71.4% of students who were coded as having autonomy). Examination of relevant interview data revealed that students in the upper quartile more commonly described having control over their research project from the beginning through identifying a topic and research question. For example, one student in the upper quartile noted,

It was very much on me to pursue wherever it [the research study] went. She [my advisor] gave me some papers and said, “Hey, these are kind of interesting, you should read them” and then expected me to figure out what to do with them. It was a really great experience to form questions and try to see just what and how previous literature affected my current research.

In contrast, students in the lowest quartile more commonly described having autonomy over smaller subsections of the research project, such as choosing the data analysis methods. For example, one student shared “I started working on the project by going into the field and collecting the data. And after that I was able to analyze the data, look at it using different methods.”

Several students who discussed the nature of their autonomy during their undergraduate research experience also described the benefit of this intellectual contribution. One student who spoke broadly about the value of autonomy shared, “I did appreciate the fact that I could control

my own learning because all my classes were what the professor wanted to teach and this was the investigation that I wanted to conduct.” Another student found the intellectual rigor she sought when she became involved in a research project in which she had some control over the research design:

I think it [undergraduate research] made college more challenging. To be quite honest, I was not challenged by the coursework and I was like “This is boring.” It’s high school all over again.” When I started doing the undergraduate research it was like, “This is why I’m here. This is what I want to do for the rest of my life is research. I’m finally doing it.” I felt like I had grown up a little bit.

Other students articulated more specifically how their research skills developed through having some control over their undergraduate research, including a perceived increase in disciplinary knowledge, self-efficacy for doing research, ability to read and apply primary literature, and problem-solving skills:

It was completely an independent study and we got to pick what direction we wanted to go in. So I picked coral reefs because I love scuba diving. And I learned so much about ecotourism problems, global warming, how it affects the ocean, and ocean acidification. I got back to school and I switched my major the next day.

He [my advisor] said “I sort of put this [research project] on the backburner, now see what you can do with it.” And I made some progress with it. It was a scary thing to begin with. When you hear the word, ‘research,’ you say, “I don’t even know the basic material, how can I do research in this area?” But when you really see, you sit down with the problem and you work at it and you see this is not so bad . . . I mean you just keep plugging away at it. So it was much more approachable after that . . .

It was a really great experience to form questions and try to see just what and how previous literature affected my current research and then when I had failures, how to learn from those failures, of which there were many, and then see how your research questions change over time and how sometimes, the things that you asked at the beginning are completely meaningless, but they at least formed the foundation for something that is important or further avenues of explanation.

Some students who reported limited autonomy during their undergraduate research saw this restriction as a challenge to their intellectual

growth. For example, one student shared, “Sometimes you are treated as the undergraduate and you work with the graduates and so you do the undergraduate things and graduates do the graduate things. It can be frustrating. You can get left in the dark.”

**Research Network Size.** Research network size refers to the extent to which students described engaging with others within the context of the research project. Based on interview responses, we divided students into two groups: those who reported being supervised by a single mentoring faculty member, and those who reported collaborating with multiple individuals. Sixty-six percent (31 of 47) of students provided information about whether they collaborated with individuals other than their faculty mentor during undergraduate research participation. Of these, 9 (29%) reported only interacting with their advisor. For example, one participant shared, “It was just myself and my advisor.”

In contrast, 22 (71.0%) reported collaborating with a network of individuals. These networks included students’ advisors, other faculty members both within and outside of the student’s department, professional research staff, peers, graduate students, and postdoctoral fellows. For example, one student explained:

Much of [our work] in our lab, we work together as far as, “Well, how do you think? Here’s my problem. How would you attack this?” Things like that, and it is all the same subject area, but we all have our individual projects. It is definitely a collaborative environment. We had a lab meeting every week and we talked about all of our individual projects, and we did minipresentations.

Another student described her experience working with a large research group and local/state agencies during her undergraduate research:

The whole organization participated in different projects, but they all focused on manatees, and so they cooperated with a lot of state agencies and local agencies as well, because some of the studies were ongoing, like 20-year manatee tracking, and population and survey studies, and so a bunch of different groups participated over the years, like Florida Fish and Wildlife, US Fish and Wildlife definitely used the data, and I helped all the research scientists there at some point.

To investigate the relationship between undergraduate research network size and graduate students’ research skill performance, we examined research network size by preproposal total score quartiles. We found no patterns in the data. Further, only one student indicated that being part of a larger research team helped him develop basic research

skills and problem-solving that later helped him with his dissertation work:

Three people were working on three different genes. They have impacted me a lot. They provided me the basics how to do research in a lab, how to write notes after every day. That's very important because when in the last month when I was writing my six month dissertation, it helped me a lot and two of the PhDs down there, they helped me. They helped us in all the steps, especially solving problems.

No students described working only with their advisor as beneficial or harmful to their research skill development.

**Motivation.** Motivation for conducting research concerns whether students reported voluntarily conducting undergraduate research or reported being required to do so by their undergraduate degree programs (mandatory). For students who engaged in multiple projects, if they reported undertaking any project to fulfill a requirement, their response was coded as 'mandatory.'

Twenty-six of the 47 (55.3%) students who reported conducting undergraduate research provided interview data concerning voluntary or mandatory undergraduate research participation. Twenty (76.9%) reported voluntary engagement, noting that they conducted undergraduate research to explore their interests, explore career options, obtain research experience necessary for entry into graduate/medical school, assist a colleague or professor who needed help, have more responsibility, be cognitively challenged, or develop a closer relationship with a faculty member. For example, one participant who was coded as voluntarily participating in undergraduate research shared, "I just wanted to have a little lab experience before I applied to grad school and before I got into grad school." In contrast, 6 students (23.1%) reported engaging in mandatory research, such as an undergraduate thesis or capstone project. For example, one participant shared, "In Ghana every science student is supposed to do a thesis as an undergraduate. I did a thesis."

To examine the relationship between motivation for undergraduate research and graduate research skill performance, we examined motivation by preproposal total score quartiles. We found no consistent patterns across performance quartiles in terms of whether students were mandated or volunteered to conduct undergraduate research. Further, no students reported connections between their motivation for doing research and their research skill development.

## **Discussion and Conclusions**

Findings suggest that participation in undergraduate research may be linked to heightened research skills in graduate school which supports hypothesis 1 of this study. Students who conducted undergraduate research had higher scores on almost all components and subscales for both pre and postproposals. Although a pronounced trend was observed, differences between students who had conducted undergraduate research and those who had not were not statistically significant in many cases, likely due to the limited study sample size and resulting low statistical power (or the ability to detect group differences). Thus, we view these nonsignificant results as inconclusive.

When duration of undergraduate research experience was considered, pronounced and striking effects were seen wherein as semesters of research experience accumulate, proposal scores increase, supporting hypothesis 2a. This conclusion is further supported by the greater number of significant correlations found between duration and postproposal scores. This is consistent with previous research indicating that duration of undergraduate research mediates outcomes (Fechheimer et al., 2010; Russell et al., 2007; Zydney et al., 2013), and with research underlining the importance of accumulated practice in developing expertise (Ericsson & Charness, 1994). This finding highlights the benefit of undergraduate research experiences that span several semesters. Although we gathered information about undergraduate research duration, we did not ask how long ago research experiences occurred. This information will be useful in reducing ‘noise’ in future studies. However, we note that undergraduate research duration was a robust predictor of research skill performance despite our inability to control for length of time between undergraduate research and participants’ entry into current degree programs.

Students perceived having control over their undergraduate research activities was beneficial to research skill development, as indicated by our qualitative findings. Numerous students explicitly connected the opportunity to select their own research problem or make methodological decisions as valuable in their development as researchers (providing partial support for hypothesis 2b). Specifically, they discussed the benefits of autonomy in terms of increasing intellectual rigor in their studies and strengthening their research interests, disciplinary knowledge, self-efficacy for doing research, ability to read and apply primary literature, and problem-solving skills. This finding is consistent with studies identifying the importance of autonomy in student research (Burke & Cummins, 2010; Hefferan, Heywood, & Ritter, 2002; Lopatto, 2003).

Examination of qualitative data characterizing students' experience with autonomy during undergraduate research indicated that students with the strongest research skills also had the most substantial autonomy, often designing their research projects from the outset beginning with devising a research question. Prior work supports the importance of early involvement in study design as autonomy is related to "significant improvement in the remaining phases of inquiry" (as cited in Kuhn & Pease, 2008, p. 516).

These triangulated findings suggest the criticality of autonomy in learning through undergraduate research. However, an alternative explanation may be that these students were afforded autonomy because they had already developed substantial research skills. This alternative interpretation would align with research suggesting that learners benefit from more self-directed learning when they have substantial prior knowledge (Kirschner, Sweller, & Clark, 2006). Future studies should examine when in an undergraduate's development it is most helpful to give them control over their research.

Researchers may also find it valuable to parse out the aspects of research over which students have control (e.g., research problem, data analysis, writing) to better understand how autonomy contributes to research skill development. Our qualitative finding comparing students with autonomy in the lowest quartile vs. the highest quartile suggests that autonomy in determining the research focus and questions matters more than autonomy in selecting data analysis tools. As our sample size was limited, we view this result as exploratory, although research by Maher, Gilmore, Feldon, and Davis (2013) substantiates this finding. Like the undergraduates in this study, novice graduate research assistants studied by Maher and colleagues frequently did not participate in the early phases of designing research. The researchers noted that this is likely because students employed in research-intensive settings are typically connected to grant funding that was initiated prior to the graduate student joining the lab. Thus, the student is brought into the work once the research has already been conceptualized. Perhaps as a result of this arrangement, the supervisors perceived that the student's primary role in the lab was "to complete the experimental study at hand" rather than to develop as a researcher (Maher et al., 2013, p. 18). This calls to question whether student research experiences should be characterized using a cognitive apprenticeship model. Thus, we suggest that future studies investigate supervisors' motives for employing student researchers, the goals they hold for their research assistants, their beliefs about how to achieve those goals, and their mentoring practices.

In addition to further exploring how and when autonomy benefits researcher development as well as if, how, and when supervisors scaffold student autonomy, future studies may also tease apart the duration of research and level of student autonomy. Although we were unable to ensure that these variables were independent in the current study, it does not change the significant correlations we observed for duration or the emerging qualitative findings for autonomy. However, an appropriate place to focus future research would be differentiating autonomy and time spent doing research to see if they each independently or jointly contribute to learning outcomes.

Work conducted in the area of collaboration during undergraduate research indicates that undergraduates perceive the opportunity to collaborate with individuals other than their faculty research advisor as beneficial, particularly in understanding alternative perspectives, increasing motivation, and minimizing stress (Waite & Davis, 2011). However, students in this study did not generally connect being part of a larger research team with their research skill development. Qualitative analyses in which we examined students' undergraduate research across levels of research skill performance did not reveal differences between students who worked only with a research advisor and those who worked with a larger research team. The fact that we did not find a benefit to being part of a larger research team may reflect that even among STEM research teams, students often pursue individual projects and as a result cooperation with other team members is limited (Maher et al., 2013). Although this study cannot confirm Waite and Davis' findings, this is an area ripe for further study as STEM research is becoming increasingly collaborative (Fox & Mohapatra, 2007; Jones et al., 2008). We particularly encourage researchers to explore the extent to which research labs are appropriately characterized as a community of practice. Part of this work should involve examining the interactions between the undergraduate researcher and their mentors and other members of the research team.

Students' motivation for engaging in undergraduate research was also examined in connection to research skill development. No students described how their motivation for engaging in undergraduate research impacted their skill development, and no differences in motivation were found when examined with respect to research proposal scores. These preliminary findings are consistent with those of Vieyra et al. (2011) in that they suggest that students who are required to conduct undergraduate research may benefit from the experience as much as undergraduates who self-select to conduct research. However, the impact of requiring undergraduate research on research skill development should be further

investigated given calls advocating for mandatory undergraduate research engagement (Healey & Jenkins, 2009; Houlden, Raja, Collier, Clark, & Waugh, 2004). Further, Russell's (Russell et al., 2007; Russell, 2008) concern was that mandatory research experiences may result in a decreased likelihood of pursuing a doctorate. Given that all students in this study were selected because they were enrolled in graduate school, we were unable to refute this hypothesis. This represents another direction for future research.

### *Study Limitations*

While intriguing, our findings are based on a relatively small sample. Further, statistical analyses utilized nonparametric tests that make fewer assumptions about one's data but require larger samples to detect group differences. If sample size was increased and/or if assumptions of parametric tests were met, statistical power would also increase.

Although this study employed a direct assessment of research skill performance, we relied on self-report to examine undergraduate research characteristics (duration, autonomy, research network size, and motivation). Participants could have inaccurately remembered or misreported their experiences. However, relevant research suggests that factual recall is accurate (e.g., "I did these things"). Where accuracy fails is when inference is required (e.g., "This is the impact these events had," "This is the mechanism I used to find a solution to a problem;" Feldon, 2007). We used the most valid kind of self-report data in this study. However, future studies could capture information about the nature of undergraduate research experiences through more trustworthy methods such as observation or concurrent reporting.

Lack of precision in coding of the four undergraduate research characteristics (duration, autonomy, research network size, and motivation) also limits conclusions that can be drawn from this analysis. Each variable could be better measured along a spectrum of experience. However, the granularity of our dataset only allowed us to reliably code participants' data along less-refined categories. Efforts to replicate this study would benefit from a deeper pool of qualitative data around the nature of participants' undergraduate research experiences. We note, however, that while the roughness of our measures introduces more error, it does not bias the results in one particular direction (i.e., with respect to duration of undergraduate research, we are as likely to have overestimated duration of experience as we are to have underestimated it). As the standard error increases, it makes statistical significance less likely to attain, increasing the likelihood of a Type I error (incorrectly failing to reject

the null hypothesis). Our ability to reject the null indicates that the impact of duration may be greater than what we found due to the limitations of our measure.

Finally, the correlational nature of this study also limits the interpretation of its findings, in part due to self-selection effects. Future studies using quasi-experimental or experimental designs will enable differentiation between direct influences of undergraduate research experiences on research skill development. They can also determine if students with preexisting differences in research skills or capacity for building research skills are more likely to participate in undergraduate research.

### *Implications for Policy and Practice*

Although this study is exploratory and correlational, our most pronounced finding is that undergraduate research is linked to heightened performance in vital STEM research skills. We detected this trend despite wide variation in implementation and multiple institutional contexts. Thus, coupled with accumulating evidence about the benefits of undergraduate research, we advocate for the inclusion of undergraduate research in the STEM curriculum (Bauer & Bennett, 2003; Craney, McKay, & Morris, 2011; Foertsch, Alexander & Penberthy, 2000; Hunter, Laursen, & Seymour, 2007; Jones, Barlow, & Villarejo, 2010; Kardash, 2000; Nagda et al., 1998; Russell, Hancock, & McCullough, 2007; Seymour, Hunter, Laursen, & DeAntoni, 2004; Summers & Hrabowski, 2006; Thiry, Laursen, & Hunter).

Centralized offices for undergraduate research represent one method for supporting undergraduate research at the institutional level. Such offices may be particularly worthwhile in light of the goal of increasing the number of STEM graduates. Creating policies at the graduate level that promote undergraduate research may also be helpful. For example, graduate admissions processes might prioritize selection of applicants with undergraduate research experiences as these experiences are associated with increased graduate school performance.

Our study, coupled with that of Vieyra et al., (2011), indicates that requiring research experience is *not* associated with negative outcomes and is a viable option in the STEM disciplines. Mandatory engagement may also reduce faculty ' concerns about undergraduate student attrition prior to research project completion (Gates, Teller, Bernat, & Delgado, 1999). Additionally, undergraduates often express concern about the time required to complete a voluntary research project (Gates et al., 1999). Mandatory engagement may alleviate this concern, as it is more likely to be associated with course credit and, in turn, degree completion.

Our findings, in concert with those of others (Fechheimer et al., 2010; Russell et al., 2007; Zydney et al., 2013), indicate that undergraduate research engagement spanning multiple semesters is most beneficial to student development. Therefore, we encourage institutions to create opportunities, guidelines, and incentives for undergraduates to conduct research across multiple semesters. Increased length of research engagement may also address faculty concerns regarding time needed to train undergraduate researchers (Gates et al., 1999).

Our study, coupled with prior work (Burke & Cummins, 2010; Heferan, Heywood, & Ritter, 2002; Lopatto, 2003), highlights the value of students experiencing control over their research. Helping students become self-directed and conduct original, independent research is the primary goal of graduate education (Gardner, 2008). If graduate study is designed to train autonomous researchers, we see undergraduate research as an effective mechanism for developing self-directed learning, particularly when the research experience spans multiple semesters. When undergraduates participate in research across semesters, institutions should encourage and support faculty in designing the experience so that students develop new research skills (e.g., designing the study from the outset) or refine their research skills over time.

### *Significance*

This study is the first to connect undergraduate research participation with an empirical assessment of research skills in graduate school. By using a valid, direct assessment of learning outcomes, in this case, research skills, this study significantly extends the existing literature that relies upon self-report of learning outcomes. This study is also important because it provides preliminary insights into *why*, or under what conditions, undergraduate research experience is associated with increased research skill acquisition. This study also includes participants from diverse backgrounds, suggesting that our findings are applicable across institution types, STEM disciplines, student demographic groups (e.g., those who speak English as a Second Language), and despite how the experience is structured (e.g., across single- vs. multiple semesters, across levels of autonomy afforded, as an independent vs. collaborative research project, and whether it is required/not required for degree completion). Finally, our study offers well-defined avenues for further inquiry into undergraduate research.

Ultimately, we see undergraduate research as a way to improve undergraduate education, retain STEM majors, and promote success and retention of STEM graduate students. These outcomes will provide the qualified candidates needed for the STEM workforce.

## Notes

Acknowledgements: The work reported in this article was supported by a grant from the National Science Foundation, NSF-0723686. The views in this paper are those of the authors and do not necessarily represent the views of the supporting funding agency.

<sup>1</sup> The rubric components *Broader Impacts* and *Writing Quality* were excluded from analysis in this study because they showed low reliability.

## References

- Association of American Colleges and Universities (2008). High impact educational practices: What they are, who has access to them, and why they matter. Washington, DC. Retrieved from [http://www.neasc.org/downloads/aacu\\_high\\_impact\\_2008\\_final.pdf](http://www.neasc.org/downloads/aacu_high_impact_2008_final.pdf)
- Bauer, K. W., & Bennett, J. S. (2003). Alumni perceptions used to assess undergraduate research experience. *The Journal of Higher Education*, 74(2), 210–230.
- Blanton, R. L. (2008). A brief history of undergraduate research, with consideration of its alternative futures. In R. Taraban & R. Blanton (Eds.), *Creating effective undergraduate research programs in science: The transformation from student to scientist* (pp. 233–246). New York: Teachers College Press.
- Bowman, N. A. (2010). Can 1st-year college students accurately report their learning and development? *American Educational Research Journal*, 47(2), 466–496.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 35–42.
- Burke, L. A., & Cummins, M. K. (2010). Using undergraduate student-faculty collaborative research projects to personalize teaching. *College Teaching*, 50(4), 129–133.
- Carnegie Foundation. (2014). *Carnegie classifications institutional profiles*. Retrieved from [http://classifications.carnegiefoundation.org/lookup\\_listings](http://classifications.carnegiefoundation.org/lookup_listings)
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.
- Council on Undergraduate Research. (2005). *Joint statement on principles of support of undergraduate research, scholarship and creative activities*. Retrieved from <http://www.cur.org/SummitPosition.html>
- Craney, C., McKay, T., & Morris, J. (2011). Cross-discipline perceptions of the undergraduate research experience. *The Journal of Higher Education*, 82(1), 92–113.
- Creswell, J. W., & Plano Clark, V. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3), 83–87.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, 49(8), 725–747.
- Fairweather, J. (2008). Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) education: A status report. Commissioned paper

- presented at NRC workshop on Evidence on Selected Promising Practices in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education; Washington, DC. Retrieved from [http://www7.nationalacademies.org/bose/Fairweather\\_CommissionedPaper.pdf](http://www7.nationalacademies.org/bose/Fairweather_CommissionedPaper.pdf)
- Fechheimer, M., Webber, K., & Kleiber, P.B. (2010). How well do undergraduate research programs promote engagement and success of students? *CBE Life Science Education*, 10(2), 156–163.
- Feldon, D. F. (2007). The implications of research on expertise for curriculum and pedagogy. *Educational Psychology Review*, 19, 91–110.
- Feldon, D. F., Maher, M. A., Hurst, M., & Timmerman, B. E. (2015). Faculty mentors, graduate students, and performance-based assessments of students' research skill development. *American Educational Research Journal*, 52(2), 334–370.
- Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D., Gilmore, J., Stiegelmeier, C. (2011). Graduate students' teaching experiences improve their methodological research skills. *Science*, 333(6045), 1037–1039.
- Feldon, D. F., Timmerman, B.E., Stowe, K. A., & Showman, R. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. *Journal of Research in Science Teaching*, 47(10), 1165–1185.
- Finlay, S., & Faulkner, G. (2005). Tête à tête: Reading groups and peer learning. *Active Learning in Higher Education*, 6(1), 32–45.
- Foertsch, J., Alexander, B., & Penberthy, D. (2000). Summer research opportunity programs (SROPs) for minority undergraduates: A longitudinal study of program outcomes 1986–1996. *Council of Undergraduate Research Quarterly*, 20(3), 114–119.
- Fox, M. F., & Mohapatra, S. (2007). Social-Organizational characteristics of work and publication productivity among academic scientists in doctoral-granting departments. *The Journal of Higher Education*, 78(5), 542–571.
- Gafney, L. (2005). The role of the research mentor/teacher: Student and faculty views. *Journal of College Science Teaching*, 34(4), 52–56.
- Gardner, S. K. (2008). “What’s too much and what’s too little?”: The process of becoming an independent researcher in doctoral education. *The Journal of Higher Education*, 79, 326–350.
- Gates, A. Q., Teller, J. P., Bernat, A., & Delgado, N. (1999). Meeting the challenge of expanding participation in the undergraduate research experience. *Journal of Engineering Education*, 88(4), 409–414.
- Gilmore, J., Strickland, D., Timmerman, B. E., Maher, M. & Feldon, D. (2010). Weeds in the flower garden: An exploration of plagiarism in graduate students' research proposals and its connection to enculturation, ESL, and contextual factors. *International Journal for Educational Integrity*, 6(1), 13–28.
- Healey, M., & Jenkins, A. (2009). Developing undergraduate research and inquiry. York, UK: Higher Education Academy. Retrieved from [http://business.heacademy.ac.uk/assets/documents/resources/publications/DevelopingUndergraduate\\_Final.pdf](http://business.heacademy.ac.uk/assets/documents/resources/publications/DevelopingUndergraduate_Final.pdf)
- Hefferan, K. P., Heywood, N. C., & Ritter, M. E. (2002). Integrating into field trips and classroom learning a capstone undergraduate research experience. *Journal of Geography*, 101, 183–190.

- Houlden, R. L., Raja, J. B., Collier, C. P., Clark, A. F., & Waugh, J. M. (2004). Medical students' perceptions of an undergraduate research elective. *Medical Teacher, 26*, 659–661.
- Hunter, A. B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education, 91*(1), 36–74.
- Johnson, R. L., Penny, J., and Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education, 13*(2), 121–138.
- Johnson, R. L., Penny, J., Gordon, B., Shumate, S. R., and Fisher, S. P. (2005). Resolving score differences in the rating of writing samples: Does discussion improve the accuracy of scores? *Language Assessment Quarterly: An International Journal, 2*(2), 117–146.
- Jones, B. F., Wuchty, S., & Brian U. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science, 322*, 1259–1262.
- Jones, M. T., Barlow, A. E. L., & Villarejo, M. (2010). Importance of undergraduate research for minority persistence and achievement in biology. *The Journal of Higher Education, 81*(1), 82–115.
- Kardash, C. M. (2000). Evaluation of undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology, 92*(1), 191–201.
- Kinkead, J. (2003). Learning through inquiry: An overview of undergraduate research. *New Directions for Teaching and Learning, 93*, 5–17.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.
- Kuhn, D., & Pease, M. (2008). What needs to develop in the development of inquiry skills? *Cognition and Instruction, 26*, 512–559.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Lopatto, D. (2003). The essential features of undergraduate research. *Council of Undergraduate Research Quarterly, 24*, 139–142.
- Maher, M., Gilmore, J., Feldon, D., & Davis, T. (2013). Cognitive apprenticeship and the supervision of science and engineering research assistants. *The Journal of Research Practice, 9*(2), 1–22.
- 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*, 55–72.
- National Academy of Sciences. (2007a). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academies Press. Retrieved from <http://www.utsystem.edu/competitive/files/RAGS-fullreport.pdf>
- National Academy of Sciences. (2007b). *Is America falling off the flat earth?* Washington, DC: National Academies Press. Retrieved from [http://www.sandia.gov/NINE/documents/Augustine\\_FallingOfftheFlatEarth.pdf](http://www.sandia.gov/NINE/documents/Augustine_FallingOfftheFlatEarth.pdf)
- National Research Council. (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.

- National Science Board. (2015). Revisiting the STEM Workforce, A Comparison to Science and Engineering Indicators 2014, Arlington, VA: National Science Foundation (NSB-2015-10). Retrieved from <http://www.nsf.gov/pubs/2015/nsb201510/nsb201510.pdf>
- Pearson, M., & Brew, A. (2002). Research training and supervision development. *Studies in Higher Education*, 27(2), 135–150.
- President's Council of Advisors on Science and Technology (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Retrieved from [http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-executive-report-final\\_2-13-12.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-executive-report-final_2-13-12.pdf)
- Russell, S. H. (2008). Undergraduate research opportunities: Facilitating and encouraging the transition from student to scientist. In R. Taraban & R. L. Blanton (Eds.), *Creative effective undergraduate research programs in science* (pp. 53–80). New York: Teachers College Press.
- Russell, S. H., Hancock, M. P., & McCullough, J. (2007). The Pipeline: Benefits of undergraduate research experiences. *Science*, 316(5824), 548–549.
- Seymour, E., & Hewitt, N. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview.
- Seymour, E., Hunter, A. B., Laursen, S., L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education*, 88(4), 493–534.
- SJB Research Consulting Inc. (2006). The Summer Undergraduate Program in Engineering Research at the University of California, Berkeley (SUPERB-IT Program): Final Evaluation Report. New York, NY.
- Stage, F. K., & Kinzie, J. (2009). Reform in undergraduate science, technology, engineering, and mathematics: The classroom context. *The Journal of General Education*, 58(2), 85–105.
- State Educational Technology Director's Association. (2008). Science, technology, engineering, and math. Retrieved from <http://www.learning.com/press/pdf/Science-Technology-Engineering-Mathematics-STEM-Report.pdf>
- Strassburger, J. (1995, May). Embracing undergraduate research. *American Association of Higher Education (AAHE) Bulletin*, 47, 3–5.
- Summers, M., & Hrabowski, F. (2006). Preparing minority scientists and engineers. *Science*, 311(5769), 1870–1871.
- Suskie, L. A. (2009). *Assessing student learning: A common sense guide*. San Francisco: Jossey-Bass.
- Thiry, H., Laursen, S. L., Hunter, A. (2011). What experiences help students become scientists? A comparative study of research and other sources of personal professional gains for STEM undergraduates. *The Journal of Higher Education*, 82(4), 357–388.
- Timmerman, B. E., Feldon, D., Maher, M., Strickland, D., & Gilmore, J. (2013). Performance-based assessment of graduate student research skills: Timing, trajectory, and potential thresholds. *Studies in Higher Education*, 38(5), 693–710.
- Timmerman, B. E., Strickland, D., Johnson, R. L., & Payne, J. (2011). Development of a 'universal' rubric for assessing students' scientific reasoning skills using scientific writing. *Assessment and Evaluation in Higher Education*, 36(5), 509–547.

- Vandermaas-Peeler, M., Nelson, J., Ferretti, L., & Finn, L. (2011). Developing expertise: An apprenticeship model of mentoring undergraduate research across cohorts. *Perspectives on Undergraduate Research Mentoring*, 1(1), 1–10.
- Vieyra, M., Gilmore, J., & Timmerman, B. E. (2011). Requiring research may improve retention in STEM fields for underrepresented women. *Council of Undergraduate Research Quarterly*, 32(1), 13–19.
- Waite, S., & Davis, B. (2011). Developing undergraduate research skills in a faculty of education: Motivation through collaboration. *Higher Education Research and Development*, 25(4), 403–419.
- Walker, G. E., Golde, C., Jones, L., Conklin Bueschel, A., & Hutchings, P. (2008). *The formation of scholars: Rethinking doctoral education for the twenty-first century*. San Francisco: Jossey-Bass.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge, MA: Cambridge University Press.
- Zydney, A. L., Bennett, J. S., Shahid, A., & Bauer, K. W. (2013). Impact of undergraduate research experience in engineering. *The Journal for Engineering Education*, 91(2), 151–157.

Copyright of Journal of Higher Education is the property of Ohio State University Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.