

Virtually the Same? Evaluating the Effectiveness of Remote Undergraduate Research Experiences

Riley A. Hess,¹ Olivia A. Erickson,² Rebecca B. Cole,¹ Jared M. Isaacs,² Silvia Alvarez-Clare,³ Jonathan Arnold,⁴ Allison Augustus-Wallace,⁵ Joseph C. Ayoob,⁶ Alan Berkowitz,⁷ Janet Branchaw,⁸ Kevin R. Burgio,⁷ Charles H. Cannon,³ Ruben Michael Ceballos,⁹ C. Sarah Cohen,¹⁰ Hilary Collier,¹¹ Jane Disney,¹² Van A. Doze,¹³ Margaret J. Eggers,¹⁴ Edwin L. Ferguson,¹⁵ Jeffrey J. Gray,¹⁶ Jean T. Greenberg,¹⁵ Alexander Hoffmann,¹⁷ Danielle Jensen-Ryan,¹⁸ Robert M. Kao,¹⁹ Alex C. Keene,²⁰ Johanna E. Kowalko,²¹ Steven A. Lopez,²² Camille Mathis,²³ Mona Minkara,²⁴ Courtney J. Murren,²⁵ Mary Jo Ondrechen,²² Patricia Ordoñez,²⁶ Anne Osano,²⁷ Elizabeth Padilla-Crespo,²⁸ Soubantika Palchoudhury,²⁹ Hong Qin,³⁰ Juan Ramírez-Lugo,³¹ Jennifer Reithel,³² Colin A. Shaw,³³ Amber Smith,³⁴ Rosemary J. Smith,^{32,35} Fern Tsien,³⁶ and Erin L. Dolan^{2,*}

¹Department of Psychology, ²Department of Biochemistry and Molecular Biology, and ³Department of Genetics, University of Georgia, Athens, GA 30602; ⁴Center for Tree Science, Morton Arboretum, Lisle, IL 60532; ⁵Department of Medicine and Office of Diversity and Community Engagement and ⁶Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA 70112; ⁷Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260; ⁸Education Department, Cary Institute for Ecosystem Studies, Millbrook, NY 12545; ⁹WISCIENCE and Department of Kinesiology and ¹⁰Wisconsin Institute for Science Education and Community Engagement, University of Wisconsin–Madison, Madison, WI 53706; ¹¹Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701; ¹²Department of Biology, Estuary and Ocean Science Center, San Francisco State University, San Francisco, CA 94132; ¹³Department of Molecular, Cell and Developmental Biology and ¹⁴Department of Microbiology, Immunology, and Molecular Genetics and Institute for Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles, CA 90095; ¹⁵Community Environmental Health Laboratory, Mt. Desert Island Biological Laboratory, Salisbury Cove, ME 04672; ¹⁶Department of Biomedical Sciences, University of North Dakota, Grand Forks, ND 58202; ¹⁷Department of Microbiology and Cell Biology and; ¹⁸Undergraduate Scholars Program and Department of Earth Sciences, Montana State University, Bozeman, MT 59717; ¹⁹Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 606307; ²⁰Department of Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218; ²¹Department of Math and Sciences, Laramie County Community College, Cheyenne, WY 82007; ²²Science Department, College of Arts and Sciences, Heritage University, Toppenish, WA 98948; ²³Department of Biology, Texas A&M University, College Station, TX 77840; ²⁴Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015; ²⁵Department of Chemistry & Chemical Biology and ²⁶Department of Bioengineering, Northeastern University, Boston, MA 02115; ²⁷American Psychiatric Association, Washington, DC 20024; ²⁸Department of Biology, College of Charleston, Charleston, SC 29424; ²⁹Department of Computer Science and ³⁰Department of Biology, University of Puerto Rico–Rio Piedras, San Juan, PR 00925; ³¹Department of Natural Sciences, Bowie State University, Bowie, MD 20715; ³²Department of Science and Technology, Inter American University of Puerto Rico–Aguadilla, Aguadilla, PR 00605; ³³Chemical and Materials Engineering, University of Dayton, Dayton, OH 45469; ³⁴Department of Computer Science and Engineering and Department of Biology, Geology, and Environmental Science, University of Tennessee at Chattanooga, Chattanooga, TN 37403; ³⁵Rocky Mountain Biological Laboratory, PO Box 519, Crested Butte, CO 81224; ³⁶Department of Biological Sciences, Idaho State University, Pocatello, ID 83209

Kyle Frantz, *Monitoring Editor*

Submitted Jan 4, 2022; Revised Feb 27, 2023; Accepted Mar 17, 2023

CBE Life Sci Educ June 1, 2023 22:ar25

DOI:10.1187/cbe.22-01-0001

*Address correspondence to: Erin L. Dolan (eldolan@uga.edu).

© 2023 R. A. Hess et al. CBE—Life Sciences Education © 2023 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 4.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/4.0>).

"ASCB®" and "The American Society for Cell Biology®" are registered trademarks of The American Society for Cell Biology.

ABSTRACT

In-person undergraduate research experiences (UREs) promote students' integration into careers in life science research. In 2020, the COVID-19 pandemic prompted institutions hosting summer URE programs to offer them remotely, raising questions about whether undergraduates who participate in remote research can experience scientific integration and whether they might perceive doing research less favorably (i.e., not beneficial or too costly). To address these questions, we examined indicators of scientific integration and perceptions of the benefits and costs of doing research among students who participated in remote life science URE programs in Summer 2020. We found that students experienced gains in scientific self-efficacy pre- to post-URE, similar to results reported for in-person UREs. We also found that students experienced gains in scientific identity, graduate and career intentions, and perceptions of the benefits of doing research only if they started their remote UREs at lower levels on these variables. Collectively, students did not change in their perceptions of the costs of doing research despite the challenges of working remotely. Yet students who started with low cost perceptions increased in these perceptions. These findings indicate that remote UREs can support students' self-efficacy development, but may otherwise be limited in their potential to promote scientific integration.

INTRODUCTION

Undergraduate research experiences (UREs) are critical for shaping students' decisions regarding whether to pursue graduate education and research careers in the life sciences (Laursen *et al.*, 2010; Lopatto and Tobias, 2010; Gentile *et al.*, 2017). Although UREs vary widely in duration and structure, they share some common characteristics (Lopatto, 2003; Gentile *et al.*, 2017). Typically, undergraduate researchers join faculty members' research groups to collaborate in or carry out some aspect of their research. Undergraduates are guided in their research by a more experienced researcher, such as a graduate student, postdoctoral associate, or faculty member, who is typically called their "research mentor" (Thiry and Laursen, 2011; Aikens *et al.*, 2016; Joshi *et al.*, 2019). During UREs, students are expected to engage in the practices of the discipline, including collecting and analyzing data, interpreting results, troubleshooting and problem solving, collaborating with other researchers, and communicating findings both orally and in writing (Gentile *et al.*, 2017). Often, undergraduate researchers assume increasing ownership of their research over time, taking on greater responsibility and autonomy in their work as they gain experience and expertise (Hanauer *et al.*, 2012).

In 2020, the COVID-19 pandemic caused massive disruptions of research, slowing or stopping research altogether at colleges and universities across the country (Korbel and Stegle, 2020; Redden, 2020). Summer URE programming was not spared these effects. In 2019, there were 125 National Science Foundation (NSF)-funded URE Sites in the biological sciences; in Summer 2020, 80% of Sites were cancelled (S. O'Conner, NSF program manager for BIO REU Sites, personal communication). Remarkably, about 20% of the Sites opted to proceed with their Summer 2020 programs. The programs that opted to proceed were modified to operate on an entirely remote basis. Research projects had to be modified, or changed entirely, to accommodate a remote format (Erickson *et al.*, 2022). These modifications typically included a shift from experimental, laboratory, and field-based research and techniques to research questions or problems that could be addressed using computational and analytical approaches. Additionally, program leaders and research mentors were tasked with adapting their typical

program timelines, meeting schedules, communication platforms, and curricula (e.g., seminars, workshops) to an online format.

This unprecedented and massive shift raises the question of whether undergraduates who participate in remote research programs realize the same outcomes as undergraduates who have participated in in-person URE programs. This question is important to address for several reasons. First, graduate programs and employers can benefit from knowing about the experiences and outcomes of applicants whose research experience occurred remotely during Summer 2020. Second, if remote URE programs are beneficial to students, they have the potential to expand access to research experiences, especially for students who would otherwise be excluded from in-person UREs due to geographic constraints. Third, remote URE programs may reduce some of the cost associated with in-person programming (e.g., housing), allowing reallocation of these funds to pay additional undergraduate researchers. Finally, remote UREs may allow both students and their mentors greater flexibility in balancing work-life demands, including eliminating the hassle of relocating for a temporary summer research position. The present study aims to provide insight about whether remote UREs benefit students and thus should be considered an option for URE programming in the future.

THEORETICAL FRAMEWORK

For the most part, UREs have been designed to allow students to explore research as a path for further education and careers (Seymour *et al.*, 2004; Hunter *et al.*, 2007; Laursen *et al.*, 2010; Lopatto and Tobias, 2010; Thiry *et al.*, 2011; Gentile *et al.*, 2017). Multiple theories related to career development and decision making have been used to explore and explain the outcomes students realize from participating in research. For example, Estrada, Hernandez, and colleagues carried out a series of studies framed by the tripartite integration model of social influence (TIMSI), arguing that three social factors influence students' integration into the scientific community (Estrada *et al.*, 2011; Hernandez *et al.*, 2018). Specifically, students' scientific self-efficacy, scientific identity, and perceptions of the alignment between their personal values and the values of the

scientific community (i.e., values alignment) predict whether students engage in research experiences and continue in a science research–related path (Estrada *et al.*, 2011, 2018). Furthermore, students' engagement in research increases their scientific self-efficacy, which in turn positively influences their scientific identity (Adedokun *et al.*, 2013; Robnett *et al.*, 2015; Frantz *et al.*, 2017). Thus, from an empirical perspective, research experiences can stimulate a recursive process through which students develop their research skills, feel more capable of performing research, identify and share values with the research community, and choose to continue in research (Hernandez *et al.*, 2020). Theoretically, the TIMSI illustrates how research experiences embed students in the social environment of a research group, thereby promoting their integration into the scientific community (Hernandez *et al.*, 2020).

It is unclear whether remote research affords the same social environment for students to carry out research as does an in-person experience. For example, the types of research activities that can be done at a distance are more limited, which may limit students' development of research skills and, in turn, their scientific self-efficacy. The extent to which research mentors can provide in-the-moment guidance to help students overcome challenges is also likely to be limited, because they are not working side by side. This may affect the extent to which students are successful in their research tasks, which could stymie their scientific self-efficacy development. Furthermore, students may feel less engaged in the social environment of their research group, because their interactions are more time and space limited. This may in turn limit their feelings of being part of the research community, thereby limiting their scientific identity development. Thus, it is reasonable to question whether remote UREs would foster the same level of scientific integration as in-person UREs.

Prior research has also used expectancy-value-cost theory (EVT; Eccles and Wigfield, 2002; Barron and Hulleman, 2015) as a framework for examining students' value of UREs as a predictor of their motivation to continue in research (Ceyhan and Tillotson, 2020). EVT posits that individuals' expectations about the degree to which they will be successful in a task (i.e., their self-efficacy) and their perceptions of the value and costs associated with a task or pursuit influence their motivation to engage in the task or pursuit in the future (Eccles and Wigfield, 2002; Barron and Hulleman, 2015). From this theoretical perspective, one would expect undergraduates to decide whether to pursue graduate education or research careers based on whether they perceived they were sufficiently competent and whether doing research would provide sufficient value over costs. Value can take the form of being personally interesting (intrinsic value), being useful (utility value), and providing prestige or respect (attainment value; Eccles and Wigfield, 2002). Cost can be experienced in terms of effort spent, emotional or psychological tolls, or missed opportunities (Ceyhan and Tillotson, 2020).

Work from Ceyhan and Tillotson (2020) indicates that undergraduates express intrinsic and utility value as well as opportunity costs of in-person research. However, students may experience remote research differently, ascribing different values and costs to research and differing in their motivation to continue research in the future. For example, students carrying out research remotely may not be responsible for the hands-on

collection of their data, which may limit their interest in the work (i.e., less intrinsic value). In contrast, students may perceive greater utility value, because they learn computational skills that are useful in a variety of career paths and in high demand among employers. In addition, students may perceive less opportunity cost of doing remote research because of its inherent flexibility (e.g., no need to physically relocate, options to schedule research tasks around other personal demands).

In summary, prior research using TIMSI and EVT shows that UREs influence students' scientific self-efficacy, scientific identity, and perceptions of the value and costs of doing research, which can in turn influence their intentions to pursue a graduate degree and/or a research career as well as their actual pursuit of these paths. Here, we used these frameworks to study of the influence of remote UREs on student outcomes. Specifically, we sought to address the following research questions:

1. To what extent do undergraduates who engage in remote research programs experience scientific integration in terms of gains in their scientific self-efficacy, scientific identity, values alignment, and intentions to pursue graduate education and science research-related careers?
2. To what extent do undergraduates who engage in remote research programs shift their perceptions of the values and costs of doing research?

Due to COVID-19, it was not possible to include a comparison group of in-person undergraduate researchers. Thus, we report our results here and interpret them with respect to published results of in-person UREs, which include students in URE Sites and other URE formats (e.g., Robnett *et al.*, 2015; Frantz *et al.*, 2017; Ceyhan and Tillotson, 2020; Hernandez *et al.*, 2020).

METHODS

Here we describe the results of a single-arm, comparative study. We collected data using established survey measures of the constructs of interest, which we administered before and after students participated in a remote research program. We evaluated the measurement models and then we addressed our research questions by fitting a series of latent growth models within a structural equation model framework. The results reported here are part of a larger study of remote UREs that was reviewed and determined to be exempt by the University of Georgia Institutional Review Board (STUDY00005841, MOD00008085).

Context and Participants

We contacted the 25 institutions that planned to host remote research programs during Summer 2020 (S. O'Connor, personal communication) to invite them to collaborate in this study. A total of 23 programs hosted by 24 research institutions in 18 states and one U.S. territory agreed to participate by distributing study information to their Summer 2020 cohorts of undergraduate researchers. The sample included five non-degree granting research institutes as well as three master's universities, one doctoral university, two high research activity universities, and 11 very high research activity universities according to the Carnegie Classification of Institutes of Higher Education. Three universities were classified as Hispanic-serving institutions. At the time of enrollment, undergraduate researchers did not yet know that their summer programs would take place

TABLE 1. Duration of URE programs: Remote URE programs in this study varied in duration, with most being about 10 weeks long

Duration in weeks	Number of programs
5	1
8	3
9	4 ^a
10	12
11	2

^aOne program had staggered end dates with most students engaging in research for 9 weeks.

remotely. One institution did not have the capacity to host its complete program remotely, so they partnered with another institution to host a joint program. Additionally, one of the 24 institutions offered two distinct programs funded from different sources. We treated these as a single program, because the participating students, their research projects, and the program activities were quite similar (Erickson *et al.*, 2022). In total, 307 students received the recruitment email and study information. This number includes students ($n = 27$) who participated primarily in person who were later excluded from the analysis. A total of 227 remote students in 22 programs (average group size = ~12) completed both the pre and post surveys. The average program duration was ~9 weeks; detailed duration data can be found in Table 1. Of the 227 students who responded to both the pre and post surveys, 153 identified as women, 69 identified as men, and 4 identified as non-binary. There were 45 students who indicated they were transfer students and 54 who indicated that they were first generation college students (i.e., no parent or guardian completed a bachelor's degree). Program details are described elsewhere (Erickson *et al.*, 2022).

The programs in this study were funded by the NSF or the U.S. Department of Agriculture. The NSF supports UREs through two funding mechanisms: Research Experience for Undergraduate (REU) Sites, which host cohorts of students each year, or REU Supplements, which typically support one or two undergraduate researchers associated with a funded research project (National Science Foundation, n.d.). Here, we focus on URE Sites, which typically offer some combination of networking with faculty and professional development to complement the mentored research experience (National Science Foundation, n.d.). In the past, URE participants have typically been junior- or senior-level undergraduate students who have committed to a science, technology, engineering, and mathematics (STEM) major, but programs are increasingly involving students at earlier points in their undergraduate careers in order to attract students to a STEM career who were not already on this path (National Science Foundation, n.d.).

Data Collection

We surveyed students twice using the secure survey service Qualtrics: 1) at the beginning of the program (pre survey or time 1) and 2) after all program activities had been completed (post survey or time 2). Students participating in programs that offered pre-program workshops were asked to complete the initial survey before engaging in these workshops. Students were sent emails with the final survey within a week of finishing their URE programs with up to two

reminders. Monetary incentives were not offered. Only students who completed both surveys were included in the sample (Table 2). The survey measures are described briefly here and included in their entirety in the Supplemental Material.

Scientific Self-Efficacy. Scientific self-efficacy is the extent to which students are confident in their ability to carry out various science research practices, such as developing a hypothesis to test. We used a nine-item scientific self-efficacy measure that was a combination of seven published items (Chemers *et al.*, 2011; Estrada *et al.*, 2011) and two items (“Use computational skills” and “Troubleshoot an investigation or experiment”) that we authored based on input from the directors of the URE programs in this study. These items were intended to more fully capture the forms of scientific self-efficacy students could develop by engaging in remote research. Response options ranged from 1 (“not confident”) to 6 (“extremely confident”).

Scientific Identity. Scientific identity is the extent to which students see themselves as scientists and as members of the scientific community. We used a seven-item scientific identity measure using seven published items (Chemers *et al.*, 2011; Estrada *et al.*, 2011). An example item is “I have a strong sense of belonging to the community of scientists.” Response options ranged from 1 (“strongly disagree”) to 6 (“strongly agree”).

Values Alignment. Values alignment is the extent to which students see their personal values as aligning with values of the scientific community. We used a published four-item values alignment measure (Estrada *et al.*, 2011), the structure of which was based upon the Portrait Value Questionnaire (Schwartz *et al.*, 2001). Response options ranged from 1 (“not like me”) to 6 (“extremely like me”). An example item is “A person who thinks it is valuable to conduct research that builds the world’s scientific knowledge.”

Intrinsic Value. Intrinsic value refers to how much students find research personally interesting and enjoyable. We adapted a published six-item intrinsic value measure (Gaspard *et al.*, 2015b). Response options ranged from 1 (“strongly disagree”) to 6 (“strongly agree”). An example item is “Research is fun to me.”

Personal Importance. Personal importance (also known as attainment value) refers to the importance that students place on doing well in research, including how relevant doing well in research is for their identity. We adapted a three-item personal importance measure (Gaspard *et al.*, 2015b). Response options ranged from 1 (“strongly disagree”) to 6 (“strongly agree”). An example item is “Research is very important to me personally.”

Utility Value. Although EVT conceptualizes utility value as a single construct, work from Gaspard and others has shown that students perceive different forms of utility from their educational experiences, such as utility for their future careers or for helping their communities (Thoman *et al.*, 2014; Gaspard *et al.*, 2015a, b). Thus, we chose to measure three forms of utility value (i.e., job, life, and social utility) by adapting existing scales (Gaspard *et al.*, 2015b). Job utility refers to students’

TABLE 2. Demographics of study participants^a

Race/ethnicity	Previous research experience					Total
	None	1 Term	2 Terms	3 Terms	>3 Terms	
African American or Black	7	6	7	2	9	31
Central and East Asian	6	5	8	7	4	30
Latinx	10	13	16	11	10	60
Middle Eastern	—	—	1	—	1	2
Native American or Native Hawaiian	2	2	2	—	1	7
South Asian	—	3	1	—	4	8
White	18	30	34	13	21	116

^aNote that students were able to indicate multiple races or ethnicities, so race/ethnicity counts do not sum to the total sample size.

perceptions of how useful the ability to do research would be in the context of a workplace. We adapted three job utility items, such as “The skills I develop in research will help me be successful in my career.” Life utility refers to students’ perceptions of how useful the ability to do research would be for their everyday lives. We adapted three life utility items, such as “Research comes in handy in everyday life.” Social utility refers to students’ perceptions of how useful the ability to do research would be for their communities. We adapted three social utility items, such as “Being well versed in research will prepare me to help my community.” For all utility items, the response options ranged from 1 (“strongly disagree”) to 6 (“strongly agree”).

Cost. Cost is the extent to which students perceive research as requiring them to make sacrifices. We adapted the three-item cost scale (Gaspard *et al.*, 2015b). Response options ranged from 1 (“strongly disagree”) to 6 (“strongly agree”). An example item is “I have to give up a lot to do well in research.”

Graduate and Career Intentions. Graduate and career intentions refer the extent to which students intend to pursue a graduate degree or science- or research-related career. The career-related item was used from Estrada *et al.* (2011), and the graduate degree-related item was similarly worded, with “career” replaced with “graduate degree.” Response options ranged from 1 (“I DEFINITELY WILL NOT pursue a graduate degree in science/a science research-related career”) and 5 (“I DEFINITELY WILL pursue a graduate degree in science/a science research-related career”).

Previous Research Experience. To better characterize the study sample and explore possible differential effects of remote research experiences for students with different levels of research experience, we asked students how much research experience they had before they participated in the study. Response options included: none, one semester or summer, two semesters or summers, three semesters or summers, and more than three semesters or summers.

Missing Data

Data were evaluated for missingness. Most variables in the pre and post surveys were not missing any observations. Out of all 22,635 data points, only 65 were missing. Only participants with post observations were included in the data set. To check for attrition biases, we compared pre-survey item means of

participants who did or did not complete the post survey ($n = 37$) using Welch’s two-sample t tests. We observed a significant difference only for one values alignment item, which asks participants to rate the extent to which they agree that “I can do better in the world based on my ability to do research.” Students who completed the post survey had higher pre-survey scores ($M = 5.22$, $SD = 0.92$) than those who did not ($M = 4.80$, $SD = 0.76$). This difference may be a false positive, given that we ran 25 t tests to test for item-level differences related to missingness. Based on the very limited number of missing values and the absence of meaningful group differences in missingness, we assume that data were missing at random and thus not likely to impact our results.

Data Analysis

Following the Anderson and Gerbing (1988) two-step approach, we first tested confirmatory measurement models for all measures before fitting our structural models. To attain optimum model fit for our measurement model, we followed an iterative process of model specification using confirmatory factor analysis (CFA) with robust maximum likelihood estimation. We also evaluated the internal consistency and invariance of the measures. Then, we used latent growth modeling within a structural equation model framework to address our research questions. All analyses were conducted in R v. 4.0.1 and RStudio using the R package lavaan (Rosseel, 2012; Bates *et al.*, 2014). We provide an overview of our analyses in the following sections and include details in the Supplemental Material.

Assessment of Measurement Model Fit. We used several fit indices to assess how adequately our CFA models reproduced their variance-covariance matrices. We provide a detailed description of our approaches and the resulting model fit statistics in the Supplemental Material, with a brief summary here. First, we assessed measurement model fit by conducting a chi-square test (χ^2) for each model (Kline, 2015). Then we assessed goodness of fit using equivalence testing (Yuan *et al.*, 2016; Marcoulides and Yuan, 2017; Peugh and Feldon, 2020). We supplemented evaluation of our measurement models by interpreting factor loadings to estimate the extent to which each survey item reflects its respective latent variable and coefficient omega (Ω) values as a measure of internal consistency, or the degree of item correlation within the factor (Dunn *et al.*, 2014). Ultimately, we balanced evidence from fit indices, factor loadings, and omega values to determine our final measurement models. Finally, we evaluated each measure for invariance over time points.

TABLE 3. Descriptive statistics for outcome variables

	N		Mean $\pm$ SD	
	Pre	Post	Pre	Post
Scientific Self-efficacy	259	221	3.65 $\pm$ 0.91	4.27 $\pm$ 0.88
Scientific identity	257	226	4.64 $\pm$ 0.92	4.92 $\pm$ 0.97
Values alignment	262	225	5.29 $\pm$ 0.07	5.34 $\pm$ 0.77
Graduate school intentions	262	227	4.36 $\pm$ 0.78	4.38 $\pm$ 0.81
Career intentions	262	227	4.21 $\pm$ 0.84	4.31 $\pm$ 0.79
Enjoyment	257	227	5.17 $\pm$ 0.88	5.19 $\pm$ 0.99
Intrinsic value	261	227	5.43 $\pm$ 0.69	5.35 $\pm$ 0.97
Personal importance	262	226	5.31 $\pm$ 0.71	5.28 $\pm$ 0.85
Job utility	262	227	5.54 $\pm$ 0.67	5.49 $\pm$ 0.77
Life utility	260	226	5.12 $\pm$ 0.78	5.04 $\pm$ 0.96
Social utility	262	227	5.28 $\pm$ 0.72	5.18 $\pm$ 0.93
Cost	257	226	3.47 $\pm$ 1.28	3.44 $\pm$ 1.49

Substantive Analyses. We calculated intraclass correlations (ICC) using the R package psychometric v. 2.3 (Fletcher, 2010). Specifically, we calculated ICC1, which estimates the influence of the group on scores (Bliese, 2000). We fit our data in a structural equation model framework using latent growth models (LGMs) with robust maximum likelihood estimation using the lavaan R package. The models for our one-item measures of graduate intentions and career intentions would not converge, so these two outcomes were analyzed using a latent growth model with both items included in the same equation (i.e., one slope and one intercept).

We fit five LGMs altogether for 11 total variables. We fit four LGMs related to the TIMSI, one for each of four variables (i.e., scientific self-efficacy, scientific identity, values alignment, and graduate school and career intentions). We fit one LGM to estimate changes in the seven benefit and cost variables. For each model, we report seven parameters: 1) where students are at the start of the remote URE (i.e., intercept of the fixed effect, κ_1); 2) any observed growth pre- to post-URE (i.e., slope of the fixed effect, κ_2); 3) any influence of prior research experience on students' starting values (prior research intercept; β_1) and 4) growth (prior research slope; β_2); 5) any influence of students' program on their starting values (program intercept; β_1) and 6) growth (program slope; β_2); and 7) the correlation of the random intercept and slope (Φ_{21}). We interpret a positive correlation as indicating that students starting at a higher value (e.g., greater incoming self-efficacy) grew *more* pre- to post-URE, whereas a negative correlation indicates that students starting at a higher value grew *less* from pre- to post-URE. This coefficient helps determine whether students with higher or lower scores at the start of the URE changed the most from pre- to post-URE. All reported scores are unstandardized. Means and standard deviations for each measure at both time points are reported in Table 3.

Because we conducted 77 statistical tests altogether (seven parameters for 11 variables), we used the Benjamini-Hochberg procedure for controlling the false discovery rate (Benjamini and Hochberg, 1995). This procedure calculates a critical value for each p value using the formula $(i/m)*Q$, where i is the rank of the p value from lowest to highest, m is the total number of tests run, and Q is our chosen false discovery rate. With a total of 77 tests and a false discovery rate of 5%, we determined that

all tests with a value of $p < 0.021$ would be considered significant.

RESULTS

Here we report the significant results of our LGM analyses. Given that students were grouped by program, we first calculated ICCs with program as the grouping variable to estimate the similarity in scores between students of the same program. Across both time points, intraclass correlations were small, with the highest being ICC = 0.08 for scientific self-efficacy at time 2. These results suggest that students in the same programs did not score more similarly one another than to students in other programs.

Indicators of Scientific Integration

In alignment with the TIMSI, students who participated in remote UREs grew in their scientific self-efficacy. Collectively, students did not grow in their scientific identity, values alignment, or graduate and career intentions pre- to post-URE. However, many students began their remote UREs with up to three terms of prior research experience and started their remote UREs at high levels on these variables. When we analyzed growth related to students' pre-URE levels, we found that students with lower starting scientific identity, values alignment, and graduate and career intentions grew, while those with higher starting levels did not. We report the specific results for each outcome in the following sections and in Tables 3 and 4.

Students Grew in Their Scientific Self-Efficacy Regardless of Their Starting Levels.

Students began their UREs reporting moderate levels of scientific self-efficacy ($M = 3.65$, $SD = 0.091$, $\kappa_1 = 2.91$). On average, students increased in their scientific self-efficacy by a value of 0.87 on a 1 to 6 scale from pre- to post-URE ($\kappa_2 = 0.87$, $SE = 0.25$, $p = 0.0002$). In addition, students who started their UREs at a lower level of scientific self-efficacy experienced greater growth than those with higher starting values ($\Phi_{21} = -0.44$, $SE = 0.07$, $p < 0.0001$). Students' prior research experience significantly predicted their scientific self-efficacy at the start of their UREs ($\beta_1 = 0.23$, $SE = 0.04$, $p < 0.0001$), but did not significantly predict their self-efficacy growth from pre- to post-URE. Students did not differ in their starting self-efficacy ($p = 0.733$) or their self-efficacy growth based on their programs ($p = 0.859$).

TABLE 4. Students in remote UREs differ in their scientific integration based on their starting levels

Outcome	Parameter ^a	β	SE	z	p
Scientific self-efficacy	Starting level	2.91	0.25	11.64	0.000
	Growth	0.87	0.29	3.03	0.002
	Starting level by program	0.00	0.01	0.34	0.733
	Growth by program	0.00	0.01	0.18	0.859
	Starting level based on prior experience	0.23	0.04	5.37	0.000
	Growth based on prior experience	-0.09	0.05	-1.75	0.081
	Growth based on starting level	-0.44	0.07	-6.12	0.000
Scientific identity	Starting level	3.87	0.24	16.03	0.000
	Growth	0.43	0.24	1.79	0.074
	Starting level by program	0.02	0.01	2.37	0.018
	Growth by program	0.00	0.01	0.05	0.957
	Starting level based on prior experience	0.15	0.04	3.58	0.000
	Growth based on prior experience	-0.08	0.04	-1.81	0.070
	Growth based on starting level	-0.21	0.06	-3.62	0.000
Values alignment	Starting level	5.14	0.21	24.71	0.000
	Growth	-0.26	0.24	-1.13	0.261
	Starting level by program	0.00	0.01	-0.31	0.759
	Growth by program	0.01	0.01	1.38	0.166
	Starting level based on prior experience	0.08	0.03	2.61	0.009
	Growth based on prior experience	0.01	0.03	0.39	0.697
	Growth based on starting level	-0.14	0.04	-3.51	0.000
Graduate school and career intentions	Starting level	3.98	0.19	21.35	0.000
	Growth	0.02	0.17	0.14	0.886
	Starting level by program	0.00	0.01	0.47	0.635
	Growth by program	0.00	0.01	0.05	0.957
	Starting level based on prior experience	0.08	0.03	2.10	0.017
	Growth based on prior experience	0.07	0.03	0.24	0.815
	Growth based on starting level	-0.09	0.03	-2.75	0.006

^aWe interpret the intercept fixed effect (κ_i) as the level at which students started their UREs (starting level); the slope fixed effect (κ_j) as students' growth from pre- to post-URE (growth); intercept of program and prior research experience variables (β_i) as starting level by program and starting level based on prior experience, respectively; the slope of program and prior research experience variables (β_j) as students' growth by program and growth based on prior experience, respectively; and the correlation of the random intercept and slope (Φ_{21}) as an indicator of whether students experienced different growth based on starting level on a variable. A positive correlation indicates that students starting at a higher level grew more pre- to post-URE, and a negative correlation indicates that students starting at a higher value grew less. Significant results are bolded.

In analyzing the scientific self-efficacy data, we observed that the mean score for item 2 ("Use computational skills [software, algorithms, and/or quantitative technologies]") was lower than for the other items in the scale: $M = 3.08$ pre-URE (vs. $M = 3.42$ – 4.10 for other items) and $M = 4.00$ post-URE (vs. $M = 3.85$ – 4.74 for other items). This suggests that, even though students are experiencing scientific self-efficacy growth, students perceived themselves to be less capable in their computational skills.

Students with Lower Starting Levels Grew in Their Scientific Identity. As a group, students began their UREs reporting a higher level of scientific identity than scientific self-efficacy ($M = 4.64$, $SD = 0.92$, $\kappa_i = 3.87$), and those with more prior research experience began their UREs reporting greater scientific identity ($\beta_i = 0.15$, $SE = 0.04$, $p < 0.001$). As a group, students did not grow significantly in their scientific identity from pre- to post-URE ($p = 0.074$). Rather, students with lower starting levels experienced more growth in their scientific identity than those with higher starting levels ($\Phi_{21} = -0.21$, $SE = 0.05$, $p < 0.0001$). Students differed slightly in their starting levels of scientific identity based on their programs ($\beta_i = 0.02$, $SE = 0.01$,

$p = 0.018$), but they did not differ in their identity growth based on their programs ($p = 0.957$).

Students with Lower Starting Levels Grew in Their Values Alignment. Students began their UREs reporting high levels of values alignment ($M = 5.29$, $SD = 0.68$, $\kappa_i = 5.14$). Collectively, students did not change in their values alignment from pre- to post-URE ($p = 0.261$). Yet students with lower starting levels of values alignment grew more in their values alignment compared with those who started with higher levels ($\Phi_{21} = -0.14$, $SE = 0.04$, $p < 0.0001$). Students with more prior research experience reported slightly higher levels of values alignment at the start of their UREs ($\beta_i = 0.08$, $SE = 0.03$, $p = 0.009$), although prior experience alone did not predict changes in their values alignment ($p = 0.697$). Finally, students did not differ in their starting levels of values alignment or changes in their values alignment based on their programs ($p = 0.759$ and $p = 0.166$, respectively).

Students with Lower Starting Levels Increased Their Intentions to Pursue Graduate School and Research Careers. Students began their UREs already intending to

attend graduate school ($M = 4.36$, $SD = 0.79$) and pursue a research career ($M = 4.21$, $SD = 0.84$), and their intentions as a group did not change pre- to post-URE ($p = 0.886$). Again, students with lower starting intentions experienced more growth in their intentions compared with those who started with higher levels ($\Phi_{21} = -0.09$, $SE = 0.03$, $p < 0.0006$). Students with more prior research experience reported slightly higher intentions at the start of their UREs ($\beta_1 = 0.08$, $SE = 0.03$, $p = 0.017$), although prior experience alone did not predict changes in their intentions ($p = 0.815$). Finally, students did not differ in their starting levels of intentions or changes in their intentions based on their programs ($p = 0.635$ and $p = 0.957$, respectively).

Perceptions of Benefits and Costs

Collectively, students who participated in remote UREs did not change their perceptions of the benefits and cost of doing research from pre- to post-URE. Yet students with lower starting perceptions of the benefits and costs of doing research grew more in their perceptions of both. We report the specific results for each outcome below and in Tables 3 and 5.

Students with Lower Starting Levels Grew in Their Enjoyment of Research, Personal Importance of Research, and Utility Values of Research. On average, students began their UREs at a very high level of enjoyment ($M = 5.17$, $SD = 0.88$, $\kappa_1 = 4.49$) and did not change in their enjoyment pre- to post-URE ($p = 0.381$). Students with more prior research experience started at a slightly higher level of enjoyment of research ($\beta_1 = 0.18$, $SE = 0.04$, $p < 0.0001$), while students with lower starting levels of enjoyment grew more in their enjoyment of research ($\Phi_{21} = -0.18$, $SE = 0.05$, $p < 0.001$).

Collectively, students also began their UREs perceiving a high level of the personal importance of doing research ($M = 5.31$, $SD = 0.71$, $\kappa_1 = 4.85$), and this did not change pre- to post-URE ($p < 0.748$). However, students with lower starting levels experienced more growth in their personal importance of research than those with higher starting levels ($\Phi_{21} = -0.13$, $SE = 0.05$, $p < 0.010$). Students did not differ in their starting levels or growth of personal importance of research based on their programs ($p = 0.039$ and $p = 0.397$, respectively) or their prior research experience ($p = 0.115$ and $p = 0.821$, respectively).

Similarly, students as a group started their UREs with very positive perceptions of the job, life, and social utility aspects of research (job utility: $M = 5.54$, $SD = 0.67$, $\kappa_1 = 5.18$; life utility: $M = 5.12$, $SD = 0.78$, $\kappa_1 = 4.90$; social utility: $M = 5.28$, $SD = 0.72$, $\kappa_1 = 5.06$), and this did not change pre- to post-URE (p values > 0.70). However, students with lower starting levels experienced more growth in their utility perceptions than those with higher starting levels (job utility: $\Phi_{21} = -0.16$, $SE = 0.06$, $p < 0.004$; life utility: $\Phi_{21} = -0.09$, $SE = 0.04$, $p < 0.019$; social utility: $\Phi_{21} = -0.13$, $SE = 0.05$, $p < 0.015$). Students differed very slightly in their starting perceptions of job utility based on their programs ($\Phi_{21} = 0.02$, $SE = 0.01$, $p < 0.006$), but did not differ in their starting levels or growth of their utility perceptions based on their prior research experience (p values > 0.20).

Students Did Not Change in Their Intrinsic Value of Research, Regardless of Their Starting Levels. Students began their UREs perceiving a high level of intrinsic value of doing research ($M = 5.43$, $SD = 0.69$, $\kappa_1 = 4.82$), and this did not change pre- to

post-URE ($p < 0.939$). Students differed slightly in their starting levels of intrinsic value of research based on their programs ($\beta_1 = 0.02$, $SE = 0.01$, $p < 0.014$) and their prior research experiences ($\beta_1 = 0.09$, $SE = 0.03$, $p < 0.006$). Contrary to other outcomes, students did not differ in their growth in intrinsic value based on their starting levels ($p = 0.240$).

Students with Lower Starting Levels Grew in Their Perceptions of the Costs of Research. On average, students began their UREs reporting a moderate level of perceived costs ($M = 3.47$, $SD = 1.28$, $\kappa_1 = 2.48$), and this did not change pre- to post-URE. Yet students with lower starting costs perceptions experienced more growth in their costs perceptions than those with higher starting levels ($\Phi_{21} = -0.32$, $SE = 0.11$, $p < 0.003$). Students differed very slightly in their starting cost perceptions based on their programs ($\beta_1 = 0.05$, $SE = 0.02$, $p < 0.001$), but not on their prior research experiences ($p = 0.925$). In addition, students' cost perceptions did not change in ways that related to their programs or prior research experiences ($p = 0.042$ and $p = 0.641$, respectively).

DISCUSSION

In this study, we first sought to determine whether undergraduates who engage in remote research programs experienced research-related social influence in terms of gains in their self-efficacy, scientific identity, and values alignment (research question 1). We found that students in remote UREs experienced some level of integration into the scientific community despite the remote circumstances (Estrada et al., 2011; Adedokun et al., 2013; Robnett et al., 2015; Frantz et al., 2017). Specifically, students who completed remote UREs experienced significant gains in their scientific self-efficacy, and these gains were due to their research experiences and not to their particular URE programs. Even students who had prior research experience grew in their scientific self-efficacy. This result might be attributable to additional research experience building students' confidence in their research skills, regardless of how much research they have done before. Alternatively, students' self-efficacy growth may be due to the fact that remote research requires different skill sets than in-person projects (e.g., using particular software, writing code; Erickson et al., 2022). Indeed, students started their UREs reporting less confidence in their computational skills than in their other research-related skills. It is unclear whether students' initial uncertainty about their computational skills is specific to remote research or unique to the last-minute shift from away from bench or field research. As a reminder, most of the students in this study were accepted into their programs *before* decisions were made to offer programs remotely. Regardless, students perceived that they developed their computational skills even though they were researching remotely.

The self-efficacy growth experienced by students in this study resembled the growth observed in a number of longitudinal studies of in-person UREs. For instance, Robnett and colleagues (2015) studied students who completed in-person UREs at colleges and universities across the country. The positive effects they observed took place over a period of four semesters of in-person research, while the positive effects we observed occurred in a much shorter period—an average of about 9 weeks—in entirely remote research. This result may be due to the intensity of the summer experience (~35–40 hours per

TABLE 5. Students in remote UREs differ in their perceptions of the benefits and costs of doing research based on their initial perceptions

Outcome	Parameter ^a	β	SE	z	p
Enjoyment	Starting level	4.49	0.27	16.79	0.000
	Growth	0.20	0.23	0.88	0.381
	Starting level by program	0.01	0.01	0.91	0.362
	Growth by program	-0.01	0.01	-0.66	0.508
	Starting level based on prior experience	0.18	0.04	4.10	0.000
	Growth based on prior experience	-0.03	0.04	-0.84	0.403
Intrinsic value	Growth based on starting level	-0.18	0.05	-3.36	0.001
	Starting level	4.82	0.20	24.59	0.000
	Growth	0.02	0.24	0.08	0.939
	Starting level by program	0.02	0.01	2.46	0.014
	Growth by program	-0.01	0.01	-0.63	0.526
	Starting level based on prior experience	0.09	0.03	2.75	0.006
Personal importance	Growth based on prior experience	0.00	0.04	-0.11	0.911
	Growth based on starting level	-0.06	0.05	-1.17	0.240
	Starting level	4.85	0.21	22.77	0.000
	Growth	0.08	0.25	0.32	0.748
	Starting level by program	0.02	0.01	2.06	0.039
	Growth by program	-0.01	0.01	-0.85	0.397
Job utility	Starting level based on prior experience	0.06	0.04	1.58	0.115
	Growth based on prior experience	0.01	0.04	0.23	0.821
	Growth based on starting level	-0.13	0.05	-2.56	0.010
	Starting level	5.18	0.18	28.31	0.000
	Growth	-0.05	0.22	-0.24	0.809
	Starting level by program	0.02	0.01	2.75	0.006
Life utility	Growth by program	-0.01	0.01	-1.11	0.266
	Starting level based on prior experience	0.02	0.03	0.80	0.425
	Growth based on prior experience	0.05	0.04	1.41	0.159
	Growth based on starting level	-0.16	0.06	-2.92	0.004
	Starting level	4.90	0.19	25.51	0.000
	Growth	0.06	0.21	0.28	0.783
Social utility	Starting level by program	0.02	0.01	2.27	0.023
	Growth by program	-0.01	0.01	-1.82	0.068
	Starting level based on prior experience	0.01	0.03	0.23	0.815
	Growth based on prior experience	0.06	0.04	1.58	0.114
	Growth based on starting level	-0.09	0.04	-2.34	0.019
	Starting level	5.06	0.22	23.43	0.000
Cost	Growth	-0.06	0.26	-0.22	0.824
	Starting level by program	0.01	0.01	1.27	0.204
	Growth by program	-0.01	0.01	-1.07	0.283
	Starting level based on prior experience	0.02	0.04	0.51	0.614
	Growth based on prior experience	0.05	0.04	1.27	0.203
	Growth based on starting level	-0.13	0.05	-2.44	0.015
Cost	Starting level	2.48	0.38	6.55	0.000
	Growth	0.51	0.36	1.43	0.152
	Starting level by program	0.05	0.02	3.23	0.001
	Growth by program	-0.02	0.01	-2.03	0.042
	Starting level based on prior experience	0.01	0.07	0.09	0.925
	Growth based on prior experience	-0.03	0.07	-0.47	0.641
	Growth based on starting level	-0.32	0.11	-3.02	0.003

^aWe interpret the intercept fixed effect (κ_1) as the level at which students start their UREs (starting level); the slope fixed effect (κ_2) as students' growth from pre- to post-URE (growth); intercept of program and prior research experience variables (β_1) as starting level by program and starting level based on prior experience, respectively; the slope of program and prior research experience variables (β_2) as students' growth by program and growth based on prior experience, respectively; and the correlation of the random intercept and slope (Φ_{12}) as an indicator of whether students experienced different growth based on starting level on a variable. A positive correlation indicates that students starting at a higher level grew more pre- to post-URE, and a negative correlation indicates that students starting at a higher value grew less. Significant results are bolded.

week) versus the less intense, more protracted nature of academic year UREs. Frantz and colleagues have observed similar self-efficacy growth among students in a 10-week summer program, providing additional evidence that shorter, intensive experiences are similarly effective in building students' confi-

dence in their ability to be successful in science research compared with longer, less intense programs (Frantz *et al.*, 2017). Estrada and colleagues (2018) also studied the effects of UREs on the self-efficacy of a cohort of underrepresented minority students in their junior and senior years. Similar to our results, their

findings indicated that in-person UREs had a small but significant positive effect on students' self-efficacy.

Students in our study only experienced changes in their scientific identity, values alignment, or intentions to pursue graduate education or research careers if they started their remote UREs with lower levels of these indicators. In addition, these students made relatively larger gains in self-efficacy, lesser gains in scientific identity, and even more modest gains in values alignment and graduate and career intentions. This pattern of effect sizes resembles those observed in studies of in-person UREs (Robnett *et al.*, 2015; Frantz *et al.*, 2017; Hernandez *et al.*, 2020), indicating that students are experiencing remote UREs similarly to in-person UREs but perhaps to a lesser extent. Indeed, our results differ from those observed for in-person UREs, several of which have documented a positive influence of UREs on students' scientific identity regardless of their starting point. Thus, remote UREs appear to be productive environments for advancing students' scientific integration, but primarily for students who do not already perceive themselves as already integrated into the scientific community. Notably, students' starting levels on indicators of integration were more predictive of growth than their prior research experiences. This result is consistent with observations that UREs can vary widely in implementation (Gentile *et al.*, 2017) and students' experiences differ, even within the same program (Cooper *et al.*, 2019; Limeri *et al.*, 2019; Erickson *et al.*, 2022). Thus, programs and researchers should be cautious about assuming that students who report engaging in research for a similar number of terms have comparable experiences or realized similar outcomes.

In keeping with the EVT of motivation (Barron and Hulleman, 2015), we also sought to explore the extent to which undergraduates in remote research programs shifted their perceptions of the benefits and costs of doing research (research question 2). Students in this study already perceived high benefits and low costs of research when they started their remote research, and their perceptions did not change. It is encouraging that the challenges of remote research did not, on average, dissuade students from the benefits of doing research and did not magnify their cost perceptions. In fact, students who started their remote UREs with lower levels increased slightly in their enjoyment, personal importance, and utility values of doing research. Thus, to some extent, remote UREs are useful for students to weigh the benefits of doing research if they do not already perceive high benefits. We were unable to find any quantitative studies of undergraduate researchers' perceptions of the benefits and costs of doing research with which to compare our results. Qualitative research from Ceyhan and Tillotson (2020) indicates that undergraduates express intrinsic value, which includes both interest and enjoyment, utility value, and opportunity costs of in-person research. Our findings are consistent with these results and offer additional insight that two facets of intrinsic value, namely students' enjoyment of and interest in research, can be empirically distinguished. Our results are consistent with the notion students with high interest in research self-select into summer research programs and do not on average change in their interest, but can experience changes in other research-related values.

Notably, students who started their remote UREs with lower cost perceptions also increased in these perceptions. Students

perceiving greater benefits *and* costs of doing research seems counterintuitive, yet this effect was also observed by Ceyhan and Tillotson (2020) in their study of in-person UREs. Students may be developing a deeper or more sophisticated understanding of what research is and what doing research entails, which enables them to recognize more and different benefits as well as more costs. Our measurement model assessment results (in the Supplemental Material) support this idea, because the factor loadings for benefits and costs items increase pre- to post-URE and the measures show configural invariance but not factorial invariance. In other words, students appear to be perceiving the items differently after they complete their UREs than before.

LIMITATIONS

There are several limitations of this study that should be considered in interpreting the results. The main limitation is that we designed the study as a single-arm, comparison study; no comparison group of students completing UREs in-person was included because of the circumstances caused by COVID-19. It may be that students who opted to participate in a remote URE were particularly primed for success or that mentors and URE program directors put forth additional effort to ensure a positive experience. It also may be that students were grateful to have any meaningful experience in the midst of the pandemic lockdown and thus responded more favorably than would otherwise be the case. Future research should directly compare remote versus in-person UREs, ideally using random assignment to one or the other format with students who are willing to do either. Our results provide at least some evidence of the benefits of remote research, which mitigates the ethical concerns associated with such a study.

Another limitation is that we did not collect program-level data that would allow us to connect student outcomes to program features or activities. Future research should explore how to systematically characterize URE elements in a way that allows such connections to be explored and tested. Although some efforts have been made to characterize what undergraduates do during research (Robnett *et al.*, 2015), these efforts do not capture program-level elements that are likely to influence student experiences and outcomes from UREs (Erickson *et al.*, 2022).

It may be that growth by students who started their remote UREs at higher levels of the constructs we examined was limited by the measures we used. We examined means and standard deviations, which indicated room for growth, and tested for and ruled out regression to the mean as alternative explanations for limited growth by these students (see Supplemental Material). However, we cannot rule out limitations of the measurements. For instance, our measure of scientific identity demonstrated configural invariance and high internal reliability, but the measurement model fit was only fair and did not show factorial invariance (i.e., factor loadings increased from pre- to post-URE). These results suggest that students may be changing their thinking and perceptions about research as they engage in research. Undergraduates may be shifting from thinking of themselves as a "science person" to a "science research person" as they gain more research experience. Current measures likely capture the former but not the latter. Moving forward, researchers should explore the utility of existing measures for discriminating among undergraduate students with more or less research experience and develop additional measures as needed.

Finally, there were limitations related to our sample, which was entirely comprised of biology students. Therefore, our results may be unique to the discipline. Biology research may be more or less amenable to remote research compared with other STEM disciplines. Moreover, as the full extent of the COVID-19 pandemic unfolded, students and mentors who chose to move forward with remote research may possess different personality traits or differing levels of our variables of interest (i.e., scientific identity, scientific self-efficacy) from those who opted out of remote research. Research topics themselves likely changed during the transition to accommodate the remote research arrangement, so researchers who chose to move forward with remote research may have conducted a different type of research than they originally envisioned. Finally, data were collected during a time of social unrest in the United States during Summer 2020. Awareness of social unrest and systematic racism may have affected the well-being of participants, which may have influenced their experiences in the remote URE program.

CONCLUSION

In summary, our work suggests that remote UREs can have a positive effect on student outcomes, especially their scientific self-efficacy, which has been shown to influence students' decisions to continue in science research-related career paths (Estrada *et al.*, 2011; Hernandez *et al.*, 2018). Thus, programs may wish to offer remote URE programming even though in-person research has resumed. Perhaps the greatest advantage of remote research programs is that they open doors for students who may not have the opportunity to participate in an in-person research program (Erickson *et al.*, 2022). Remote UREs can allow for more flexible scheduling and enable research participation without the additional costs and logistics of travel and lodging. Thus, remote programs may be a viable method of expanding access to UREs, especially among students who may find it difficult to travel.

Although remote UREs have many advantages, their appropriateness should be evaluated on a case-by-case basis and should be considered alongside the advantages and disadvantages of in-person UREs. Our results indicate that remote UREs do not benefit all students equally. Rather, the benefits appear to be larger for students who have more to gain because they report lower levels of scientific integration and perceive fewer benefits associated with doing research. Furthermore, certain types of research (e.g., computational biology) may be more amenable to remote work (Alford *et al.*, 2017). Particular research mentors and undergraduates may be better able to navigate the unstructured nature of remote work. Certain remote research environments may be more or less accessible for different individuals, such as those who can sit and work on a computer for extended periods of time (Reinholz and Ridgway, 2021). Certain personal situations may make remote research more difficult, such as whether individuals have access to robust Internet connections and quiet workspaces (Erickson *et al.*, 2022). Finally, because students are not able to complete benchwork at home, remote UREs may aid in the development of a different skill set than in-person UREs. Thus, students may benefit from completing both types of UREs throughout their undergraduate degree programs in order to develop a wider variety of skills.

It is important to note that students in this study were *all* conducting the *entire* research experience remotely. In the future, URE programs may wish to consider hybrid designs in which some students are in person and others are remote, or in which all students participate partly in person and partly remotely. Students may experience a hybrid program quite differently than a remote program, which could influence their outcomes. We are not aware of any existing research to support the efficacy of a hybrid URE program. If such a program exists, we encourage researchers to investigate differential outcomes for in-person and remote students who are within the same URE program.

ACKNOWLEDGMENTS

We thank all of the students, faculty, and other research mentors for their willingness to proceed with remote REU programming and for sharing their experiences so that others could learn. We also thank the Social Psychology of Research Experiences and Education group members for feedback on drafts of this article. This material is based upon work supported by the NSF under grant no. DBI-2030530. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of any of the funding organization. The authors dedicate this work to all of the undergraduates seeking to do research and the individuals who provide these opportunities despite challenging circumstances.

REFERENCES

- Adedokun, O. A., Bessenbacher, A. B., Parker, L. C., Kirkham, L. L., & Burgess, W. D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self-efficacy. *Journal of Research in Science Teaching*, 50(8), 940–951. <https://doi.org/10.1002/tea.21102>
- Aikens, M. L., Sadselia, S., Watkins, K., Evans, M., Eby, L. T., & Dolan, E. L. (2016). A social capital perspective on the mentoring of undergraduate life science researchers: An empirical study of undergraduate–postgraduate–faculty triads. *CBE—Life Sciences Education*, 15(2), ar16. <https://doi.org/10.1187/cbe.15-10-0208>
- Alford, R. F., Leaver-Fay, A., Gonzales, L., Dolan, E. L., & Gray, J. J. (2017). A cyber-linked undergraduate research experience in computational biomolecular structure prediction and design. *PLoS Computational Biology*, 13(12), e1005837. <https://doi.org/10.1371/journal.pcbi.1005837>
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Barron, K. E., & Hulleman, C. S. (2015). Expectancy-value-cost model of motivation. *Psychology*, 84, 261–271.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *LME4: Linear Mixed-Effects Models Using Eigen and S4 (R Package Version 1.1-4)*. Retrieved September 12, 2021, from <https://cran.r-project.org/web/packages/lme4/index.html>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In Klein, K. J., & Kozlowski, S. W. J. (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Hoboken, NJ: Jossey-Bass/Wiley.
- Ceyhan, G. D., & Tillotson, J. W. (2020). Early year undergraduate researchers' reflections on the values and perceived costs of their research experience. *International Journal of STEM Education*, 7(1), 1–19.
- Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, 67(3), 469–491. <https://doi.org/10.1111/j.1540-4560.2011.01710.x>

- Cooper, K. M., Gin, L. E., Akeeh, B., Clark, C. E., Hunter, J. S., Roderick, T. B., ... & Pfeiffer, L. D. (2019). Factors that predict life sciences student persistence in undergraduate research experiences. *PLoS ONE*, 14(8).
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Erickson, O. A., Cole, R. B., Isaacs, J. M., Alvarez-Clare, S., Arnold, J., Augustus-Wallace, A., ... & Dolan, E. L. (2022). "How do we do this at a distance?!" A descriptive study of remote undergraduate research programs during COVID-19. *CBE—Life Sciences Education*, 21(1), ar1. <https://doi.org/10.1187/cbe.21-05-0125>
- Estrada, M., Hernandez, P. R., Schultz, P. W., & Herrera, J. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sciences Education*, 17(1), ar9. <https://doi.org/10.1187/cbe.17-04-0066>
- Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, W. P. (2011). Toward a model of social influence that explains minority student integration into the scientific community. *Journal of Educational Psychology*, 103(1), 206–222. <https://doi.org/10.1037/a0020743>
- Fletcher, T. D. (2010). *Psychometric: Applied Psychometric Theory (R Package Version 2.3)*. Retrieved September 28, 2022, from <https://cran.r-project.org/web/packages/psychometric/index.html>
- Frantz, K. J., Demetrikopoulos, M. K., Britner, S. L., Carruth, L. L., Williams, B. A., Pecore, J. L., ... & Goode, C. T. (2017). A comparison of internal dispositions and career trajectories after collaborative versus apprenticed research experiences for undergraduates. *CBE—Life Sciences Education*, 16(1), ar1.
- Gaspard, H., Dicke, A.-L., Flunger, B., Brissson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015a). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology*, 51(9), 1226.
- Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015b). More value through greater differentiation: Gender differences in value beliefs about math. *Journal of Educational Psychology*, 107(3), 663.
- Gentile, J., Brenner, K., & Stephens, A. (2017). *Undergraduate research experiences for STEM students: successes, challenges, and opportunities*. National Academies Press. Retrieved May 17, 2017, from www.nap.edu/catalog/24622/undergraduate-research-experiences-for-stem-students-successes-challenges-and-opportunities
- Hanauer, D. I., Frederick, J., Fotinakes, B., & Strobel, S. A. (2012). Linguistic analysis of project ownership for undergraduate research experiences. *CBE—Life Sciences Education*, 11(4), 378–385. <https://doi.org/10.1187/cbe.12-04-0043>
- Hernandez, P. R., Agocha, V. B., Carney, L. M., Estrada, M., Lee, S. Y., Loomis, D., ... & Park, C. L. (2020). Testing models of reciprocal relations between social influence and integration in STEM across the college years. *PLoS ONE*, 15(9), e0238250. <https://doi.org/10.1371/journal.pone.0238250>
- Hernandez, P. R., Woodcock, A., Estrada, M., & Schultz, P. W. (2018). Undergraduate research experiences broaden diversity in the scientific workforce. *BioScience*, 68(3), 204–211. <https://doi.org/10.1093/biosci/bix163>
- 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. <https://doi.org/10.1002/sce.20173>
- Joshi, M., Aikens, M. L., & Dolan, E. L. (2019). Direct ties to a faculty mentor related to positive outcomes for undergraduate researchers. *BioScience*, 69(5), 389–397. <https://doi.org/10.1093/biosci/biz039>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY: Guilford.
- Korbel, J. O., & Stegle, O. (2020). Effects of the COVID-19 pandemic on life scientists. *Genome Biology*, 21(1), 113. <https://doi.org/10.1186/s13059-020-02031-1>
- Laursen, S., Hunter, A.-B., Seymour, E., Thiry, H., & Melton, G. (2010). *Undergraduate research in the sciences: Engaging students in real science*. San Francisco, CA: Wiley.
- Limeri, L. B., Asif, M. Z., Bridges, B. H. T., Esparza, D., Tuma, T. T., Sanders, D., ... & Dolan, E. L. (2019). "Where's My Mentor?!" Characterizing negative mentoring experiences in undergraduate life science research. *CBE—Life Sciences Education*, 18(4), ar61. <https://doi.org/10.1187/cbe.19-02-0036>
- Lopatto, D. (2003). The essential features of undergraduate research. *Council on Undergraduate Research Quarterly*, 24, 139–142.
- Lopatto, D., & Tobias, S. (2010). *Science in solution: The impact of undergraduate research on student learning*. Washington, DC: Council on Undergraduate Research.
- Marcoulides, K. M., & Yuan, K.-H. (2017). New ways to evaluate goodness of fit: A note on using equivalence testing to assess structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(1), 148–153.
- National Science Foundation. (n.d.). *Research experiences for undergraduates (REU)*. Retrieved September 12, 2021, from <https://beta.nsf.gov/funding/opportunities/research-experiences-undergraduates-reu>
- Peugh, J., & Feldon, D. F. (2020). "How well does your structural equation model fit your data?": Is Marcoulides and Yuan's equivalence test the answer? *CBE—Life Sciences Education*, 19(3), es5.
- Redden, E. (2020). "Nonessential" research has halted on many campuses. *Inside Higher*. Retrieved September 12, 2021, from www.insidehighered.com/news/2020/03/30/nonessential-research-has-halted-many-campuses
- Reinholz, D. L., & Ridgway, S. W. (2021). Access needs: Centering students and disrupting ableist norms in STEM. *CBE—Life Sciences Education*, 20(3), es8. <https://doi.org/10.1187/cbe.21-01-0017>
- Robnett, R. D., Chemers, M. M., & Zurbruggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, 52(6), 847–867. <https://doi.org/10.1002/tea.21221>
- Rossee, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36.
- Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology*, 32(5), 519–542. <https://doi.org/10.1177/0022022101032005001>
- 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. <https://doi.org/10.1002/sce.10131>
- Thiry, H., & Laursen, S. L. (2011). The role of student-advisor interactions in apprenticing undergraduate researchers into a scientific community of practice. *Journal of Science Education and Technology*, 20(6), 771–784. <https://doi.org/10.1007/s10956-010-9271-2>
- Thiry, H., Laursen, S. L., & Hunter, A.-B. (2011). What Experiences Help Students Become Scientists? A Comparative Study of Research and Other Sources of Personal and Professional Gains for STEM Undergraduates. *Journal of Higher Education*, 82(4), 357–388. <https://doi.org/10.1353/jhe.2011.0023>
- Thoman, D. B., Brown, E. R., Mason, A. Z., Harmsen, A. G., & Smith, J. L. (2014). The role of altruistic values in motivating underrepresented minority students for biomedicine. *BioScience*, bio199. <https://doi.org/10.1093/biosci/biu199>
- Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling*, 23(3), 319–330.