

1 **Title:** *Virtually the same? Evaluating the effectiveness of remote undergraduate research experiences*

2

3 **Authors:** Riley A. Hess¹, Olivia A. Erickson², Rebecca B. Cole¹, Jared M. Isaacs², Silvia Alvarez-Clare³,
4 Jonathan Arnold⁴, Allison Augustus-Wallace⁵, Joseph C. Ayoob⁶, Alan Berkowitz⁷, Janet Branchaw⁸,
5 Kevin R. Burgio⁷, Charles H. Cannon³, Ruben Michael Ceballos⁹, C. Sarah Cohen¹⁰, Hilary Coller¹¹, Jane
6 Disney¹², Van A. Doze¹³, Margaret J. Eggers¹⁴, Edwin L. Ferguson¹⁵, Jeffrey J. Gray¹⁶, Jean T.
7 Greenberg¹⁵, Alexander Hoffmann¹⁷, Danielle Jensen-Ryan¹⁸, Robert M. Kao¹⁹, Alex C. Keene²⁰, Johanna
8 E. Kowalko²¹, Steven A. Lopez²², Camille Mathis²³, Mona Minkara²⁴, Courtney J. Murren²⁵, Mary Jo
9 Ondrechen²², Patricia Ordoñez²⁶, Anne Osano²⁷, Elizabeth Padilla-Crespo²⁸, Soubantika Palchoudhury²⁹,
10 Hong Qin³⁰, Juan Ramírez-Lugo³¹, Jennifer Reithel³², Colin A. Shaw³³, Amber Smith³⁴, Rosemary J.
11 Smith^{32,35}, Fern Tsien³⁶, Erin L. Dolan²

12 1. Department of Psychology, University of Georgia, Athens, GA 30602

13 2. Department of Biochemistry & Molecular Biology, University of Georgia, Athens, GA 30602

14 3. Center for Tree Science, The Morton Arboretum, Lisle, IL 60532

15 4. Department of Genetics, University of Georgia, Athens, GA 30602

16 5. Department of Medicine & Office of Diversity & Community Engagement, Louisiana State
17 University Health Sciences Center, New Orleans, LA 70112

18 6. Department of Computational and Systems Biology, University of Pittsburgh School of
19 Medicine, Pittsburgh, PA 15260

20 7. Education Department, Cary Institute for Ecosystem Studies, Millbrook, NY 12545

21 8. WISCIENCE and the Department of Kinesiology, University of Wisconsin-Madison, Madison,
22 WI 53706

23 9. Department of Biological Sciences, University of Arkansas, Fayetteville, AR 72701

24 10. Department of Biology, Estuary and Ocean Science Center, San Francisco State University, San
25 Francisco, CA 94132

26 11. Department of Molecular, Cell and Developmental Biology, University of California Los
27 Angeles, Los Angeles, CA 90095

28 12. Community Environmental Health Laboratory, Mt. Desert Island Biological Laboratory,
29 Salisbury Cove, ME 04672

30 13. Department of Biomedical Sciences, University of North Dakota, Grand Forks, ND 58202

31 14. Department of Microbiology and Cell Biology, Montana State University, Bozeman, MT 59717

32 15. Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL
33 60637

34 16. Department of Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, MD
35 21218

36 17. Department of Microbiology, Immunology, and Molecular Genetics, and Institute for
37 Quantitative and Computational Biosciences, University of California Los Angeles, Los Angeles,
38 CA, 90095

39 18. Department of Math and Sciences, Laramie County Community College, Cheyenne, WY 82007

40 19. Science Department, College of Arts and Sciences, Heritage University, Toppenish, WA 98948

- 41 20. Department of Biology, Texas A&M University, College Station, TX 77840
- 42 21. Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015
- 43 22. Department of Chemistry & Chemical Biology, Northeastern University, Boston, MA 02115
- 44 23. American Psychiatric Association, Washington, DC 20024
- 45 24. Department of Bioengineering, Northeastern University, Boston, MA 02115
- 46 25. Department of Biology, College of Charleston, Charleston, SC, 29424
- 47 26. Department of Computer Science, University of Puerto Rico Río Piedras, San Juan, PR 00925
- 48 27. Department of Natural Sciences, Bowie State University, Bowie, MD 20715
- 49 28. Department of Science and Technology, Inter American University of Puerto Rico – Aguadilla,
- 50 Aguadilla, PR 00605
- 51 29. Chemical and Materials Engineering, University of Dayton, Dayton, OH 45469
- 52 30. Department of Computer Science and Engineering, Department of Biology, Geology, and
- 53 Environmental Science, University of Tennessee at Chattanooga, Chattanooga, TN 37403
- 54 31. Department of Biology, University of Puerto Rico Río Piedras, San Juan, PR 00925
- 55 32. Rocky Mountain Biological Laboratory, PO Box 519, Crested Butte, CO 81224
- 56 33. Undergraduate Scholars Program and Department of Earth Sciences, Montana State University,
- 57 Bozeman, MT 59717
- 58 34. Wisconsin Institute for Science Education and Community Engagement, University of
- 59 Wisconsin–Madison, Madison, WI 53706
- 60 35. Department of Biological Sciences, Idaho State University, Pocatello, ID 83209
- 61 36. Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA
- 62 70112

63

64 * **Author for Correspondence:** Erin L. Dolan, University of Georgia, Biochemistry & Molecular
65 Biology Department, B210B Davison Life Sciences Building, Athens, GA, 30602; Email:
66 eldolan@uga.edu; Phone: 540-250-3073

67

68 **Keywords:** undergraduate research, evaluation, remote, COVID-19, online

69

70

71 **ABSTRACT**

72 In-person undergraduate research experiences (UREs) promote students' integration into careers in life
73 science research. In 2020, the COVID-19 pandemic prompted institutions hosting summer URE programs
74 to offer them remotely, raising questions about whether undergraduates who participate in remote
75 research can experience scientific integration. To address this, we investigated indicators of scientific
76 integration for students who participated in remote life science URE programs in summer 2020. We found
77 that these students experienced gains in their scientific self-efficacy and scientific identity similar to
78 results reported for in-person UREs. We also found that these students perceived high benefits and low
79 costs of doing research at the outset of their programs, and their perceptions did not change despite the
80 remote circumstances. Yet, their perceptions differed by program, indicating that programs differentially
81 affected students' perceptions of the costs of doing research. Finally, we observed that students with prior
82 research experience made greater gains in self-efficacy and identity, as well as in their perceptions of the
83 alignment of their values with those of the scientific community, in comparison to students with no prior
84 research experience. This finding suggests that additional programming may be needed for
85 undergraduates with no prior experience to benefit from remote research.

86 INTRODUCTION

87 Undergraduate research experiences (UREs) are critical for shaping students' decisions regarding whether
88 to pursue graduate education and research careers in the life sciences (Gentile et al., 2017). Although
89 UREs vary widely in duration and structure, they share some common characteristics (Gentile et al.,
90 2017). Typically, undergraduate researchers join faculty members' research groups to collaborate in or
91 carry out some aspect of their research. Undergraduates are guided in their research by a more
92 experienced researcher, such as a graduate student, postdoctoral associate, or faculty member, who is
93 typically called their "research mentor" (Aikens et al., 2016; Joshi et al., 2019). During UREs, students
94 are expected to engage in the practices of the discipline, including collecting and analyzing data,
95 interpreting results, troubleshooting and problem solving, collaborating with other researchers, and
96 communicating findings both orally and in writing (Gentile et al., 2017). Often, undergraduate researchers
97 assume increasing ownership of their research over time, taking on greater responsibility and autonomy in
98 their work as they gain experience and expertise.

99

100 In 2020, the COVID-19 pandemic caused massive disruptions of research, slowing or stopping research
101 altogether at colleges and universities across the country (Korbel & Stegle, 2020; Redden, 2020). Summer
102 URE programming was not spared from these effects. In 2019, there were 125 NSF-funded URE Sites in
103 the biological sciences; in summer 2020, 80% of Sites were cancelled (Sally O'Conner, NSF Program
104 Manager for BIO REU Sites, personal communication). Remarkably, about 20% of the Sites opted to
105 proceed with their summer 2020 programs. The programs that opted to proceed were modified to operate
106 on an entirely remote basis. Research projects had to be modified, or changed entirely, to accommodate a
107 remote format. These modifications typically included a shift from experimental, laboratory, and field-
108 based research and techniques to research questions or problems that could be addressed using
109 computational and analytical approaches. Additionally, program leaders and research mentors were
110 tasked with adapting their typical program timelines, meeting schedules, communication platforms, and
111 curricula (e.g., seminars, workshops) to an online format.

112

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

126

127

128 THEORETICAL FRAMEWORK

129 For the most part, UREs have been designed to allow students to explore research as a path for further
130 education and careers (Gentile et al., 2017; Laursen et al., 2010; Lopatto & Tobias, 2010). Multiple
131 theories related to career development and decision-making have been used to explore and explain the
132 outcomes students realize from participating in research. For example, Estrada, Hernandez, and
133 colleagues carried out a series of studies framed by the Tripartite Integration Model of Social Influence
134 (TIMSI), arguing that three social factors influence students' integration into the scientific community
135 (Estrada et al., 2011; Hernandez et al., 2018). Specifically, their research has shown that students'
136 scientific self-efficacy, scientific identity, and perceptions of the alignment between their personal values
137 and the values of the scientific community predict whether students engage in research experiences
138 (Estrada et al., 2011). Furthermore, students' engagement in research increases their scientific self-
139 efficacy, which in turn positively influences their scientific identity (Robnett et al., 2015). Thus, from an
140 empirical perspective, research experiences can stimulate a positive feedback loop through which students
141 develop their research skills, feel more capable of performing research, identify and share values with the
142 research community, and choose to continue in research (Hernandez et al., 2020). Theoretically, the
143 TIMSI illustrates how research experiences embed students in the social environment of a research group,
144 thereby promoting their integration into the scientific community (Hernandez et al., 2020).

145
146 It is unclear whether remote research affords the same social environment for students to carry out
147 research as an in-person experience. For example, the types of research activities that can be done at a
148 distance are more limited, which may limit students' development of research skills and, in turn, their
149 scientific self-efficacy. The extent to which research mentors can provide in-the-moment guidance to help
150 students overcome challenges is also likely to be limited because they are not working side by side. This
151 may affect the extent to which students are successful in their research tasks, which could stymie their
152 scientific self-efficacy development. Furthermore, students may feel less engaged in the social
153 environment of their research group because their interactions are more time- and space-limited. This may
154 in turn limit their feelings of being part of the research community, thereby limiting their scientific
155 identity development. Thus, it is reasonable to question whether remote UREs would foster the same level
156 of scientific integration as in-person UREs.

157
158 Prior research has also used Expectancy-Value Theory (EVT) (Eccles & Wigfield, 2002) as a framework
159 for examining students' value of UREs as a predictor of their motivation to continue in research (Ceyhan
160 & Tillotson, 2020). Expectancy-Value Theory posits that an individual's expectations about the degree to
161 which they will be successful in a task (i.e., their self-efficacy) and their perceptions of the value of the
162 task influence their motivation to engage in the task in the future (Eccles & Wigfield, 2002). From this
163 theoretical perspective, one would expect undergraduates to decide whether to pursue graduate education
164 or research careers based on whether they perceived they were sufficiently competent and whether doing
165 research would provide sufficient value over costs. Value can take the forms of being personally
166 interesting (intrinsic value), being useful (utility value), and providing prestige or respect (attainment
167 value) (Eccles & Wigfield, 2002). Cost can be experienced in terms of effort spent, emotional or
168 psychological tolls, or missed opportunities (Ceyhan & Tillotson, 2020).

169

170 Work from Ceyhan & Tillotson (2020) indicates that undergraduates express intrinsic and utility value as
171 well as opportunity costs of in-person research. However, students may experience remote research
172 differently, ascribing different values and costs to research and differing in their motivation to continue
173 research in the future. For example, students carrying out research remotely may not be responsible for
174 the hands-on collection of their data, which may limit their interest in the work (i.e., less intrinsic value).
175 In contrast, students may perceive greater utility value because they learn computational skills that are
176 useful in a variety of career paths and in high demand among employers. In addition, students may
177 perceive less opportunity cost of doing remote research because of its inherent flexibility (e.g., no need to
178 physically relocate, options to schedule research tasks around other personal demands).

179

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

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

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

194

195 **METHODS**

196 Here we describe the results of a single-arm, comparative study. We collected data using established
197 survey measures of the constructs of interest, which we administered before and after students
198 participated in a remote research program. We evaluated the measurement models, ultimately grouping
199 values- and cost-related data into a higher order measurement model based on our results. Then we
200 evaluated the fit of the data to a series of five multilevel random intercepts models to identify changes in
201 our constructs of interest. The results reported here are part of a larger study of remote UREs that was
202 reviewed and determined to be exempt by the University of Georgia Institutional Review Board
203 (STUDY00005841, MOD00008085).

204

205 **Context and Participants**

206 We contacted the 25 institutions that planned to host remote research programs during summer 2020
207 (Sally O'Connor, personal communication) to invite them to collaborate in this study. A total of 23
208 programs hosted by 24 research institutions in 18 states and 1 U.S. territory agreed to participate by
209 distributing study information to their summer 2020 cohort of undergraduate researchers. The sample
210 included 5 non-degree granting research institutes as well as 3 masters universities, 1 doctoral university,
211 2 high research activity universities, and 11 very high research activity universities according to the

212 Carnegie Classification of Institutes of Higher Education. Three universities were classified as Hispanic
213 Serving Institutions. At the time of enrollment, undergraduate researchers did not yet know that their
214 summer programs would take place remotely. One institution did not have the capacity to host their
215 complete program remotely, so they partnered with another institution to host a joint program.
216 Additionally, one of the 24 institutions offered two distinct programs funded from different sources. We
217 treated them as a single program because the participating students, their research projects, and the
218 program activities were quite similar. In total, 307 students received the recruitment email and study
219 information. This number includes students (n=27) who participated primarily in-person who were later
220 excluded from the analysis. A total of 227 remote students in 22 programs (average group size= \sim 12)
221 completed both the pre and postsurvey. The average program duration was \sim 9 weeks; detailed duration
222 data can be found in Table 1.

223

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

234

235 **Data Collection**

236 We surveyed students twice using the secure survey service Qualtrics: at the beginning of their program
237 (presurvey or Time 1) and after all program activities had been completed (postsurvey or Time 2).
238 Students participating in programs that offered pre-program workshops were asked to complete the initial
239 survey before engaging in these workshops. Students were sent emails with the final survey within a week
240 of finishing their URE programs with up to two reminders. Monetary incentives were not offered. Only
241 students who completed both surveys were included in the sample (Table 2). The survey measures are
242 described briefly here and included in their entirety in the Supplemental Materials (Tables S1-S3).

243

244 **Scientific Self-Efficacy.** Scientific self-efficacy is the extent to which students are confident in their
245 ability to carry out various science research practices, such as developing a hypothesis to test. We used a
246 9-item Scientific Self-Efficacy measure that was a combination of 7 published items (Chemers et al.,
247 2011; Estrada et al., 2011) and 2 items (“Use computational skills” and “Troubleshoot an investigation or
248 experiment”) that we authored based on input from the directors of the URE programs in this study. These
249 items were intended to more fully capture the forms of scientific self-efficacy students could develop by
250 engaging in remote research (see Table S1 in Supplemental Materials for items). Response options ranged
251 from 1 (“not confident”) to 6 (“extremely confident”). Responses were averaged into a single score, with
252 higher scores indicating higher levels of scientific self-efficacy.

253

254 **Scientific Identity.** Scientific identity is the extent to which students see themselves as scientists and as
255 members of the scientific community. We used a 7-item Scientific Identity measure using 7 published
256 items (Chemers et al., 2011; Estrada et al., 2011) (see Table S2 in Supplemental Materials for items). An
257 example item is “I have a strong sense of belonging to the community of scientists.” Response options
258 ranged from 1 (“strongly disagree”) to 6 (“strongly agree”). Responses were averaged into a single score,
259 with higher scores indicating higher levels of scientific identity.

260

261 **Values Alignment.** Science values alignment is the extent to which students see their personal values as
262 aligning with values of the scientific community. We used a published 4-item Values Alignment measure
263 (Estrada et al., 2011), the structure of which was based upon the Portrait Value Questionnaire (Schwartz
264 et al., 2001) (see Table S3 in Supplemental Materials for items). Response options ranged from 1 (“not
265 like me”) to 6 (“extremely like me”). An example item is “A person who thinks it is valuable to conduct
266 research that builds the world's scientific knowledge.” Responses were averaged into a single score, with
267 higher scores indicating higher a higher degree of alignment between the student’s values and the values
268 of the scientific community.

269

270 **Intrinsic Value.** Intrinsic value refers to how much students find research personally interesting and
271 enjoyable. We adapted a published 6-item intrinsic value measure (Gaspard, Dicke, Flunger, Schreier, et
272 al., 2015) (see Table S3 in Supplemental Materials for items). Response options ranged from 1 (“strongly
273 disagree”) to 6 (“strongly agree”). An example item is “Research is fun to me.” Responses were averaged
274 into a single score, with higher scores indicating higher levels of intrinsic value.

275

276 **Personal Importance.** Personal importance (also known as attainment value) refers to the importance
277 that students place on doing well in research, including how relevant doing well in research is for their
278 identity. We adapted a 3-item personal importance measure (Gaspard, Dicke, Flunger, Schreier, et al.,
279 2015) (see Table S3 in Supplemental Materials for items). Response options ranged from 1 (“strongly
280 disagree”) to 6 (“strongly agree”). An example item is “Research is very important to me personally.”
281 Responses were averaged into a single score, with higher scores indicating higher levels of personal
282 importance.

283

284 **Utility Value.** Although EVT conceptualizes utility value as a single construct, work from Gaspard and
285 others has shown that students perceive different forms of utility from their educational experiences, such
286 as utility for their future careers or for helping their community (Gaspard, Dicke, Flunger, Brisson, et al.,
287 2015; Gaspard, Dicke, Flunger, Schreier, et al., 2015; Thoman et al., 2014). Thus, we chose to use a
288 measure of utility value that included multiple dimensions: social, job, and life utility (see Table S3 in
289 Supplemental Materials for items). Social utility refers to students’ perceptions of how useful the ability
290 to do research would be for their communities. We adapted 3 social utility items (Gaspard, Dicke,
291 Flunger, Schreier, et al., 2015), such as “Being well versed in research will prepare me to help my
292 community.” Job utility refers to students’ perceptions of how useful the ability to do research would be
293 in the context of a workplace. We adapted 3 job utility items (Gaspard, Dicke, Flunger, Schreier, et al.,
294 2015), such as “The skills I develop in research will help me be successful in my career.” Life utility
295 refers to students’ perceptions of how useful the ability to do research would be for their everyday lives.
296 We adapted 3 life utility items (Gaspard, Dicke, Flunger, Schreier, et al., 2015), such as “Research comes

297 in handy in everyday life.” For all utility items, the response options ranged from 1 (“strongly disagree”)
298 to 6 (“strongly agree”). Responses were averaged into a single score, with higher scores indicating higher
299 levels of utility value.

300

301 **Cost.** Cost is the extent to which students perceive research as requiring them to make sacrifices. We
302 adapted the 3-item cost scale (Gaspard, Dicke, Flunger, Schreier, et al., 2015) (see Table S3 in
303 Supplemental Materials for items). Response options ranged from 1 (“strongly disagree”) to 6 (“strongly
304 agree”). An example item is “I have to give up a lot to do well in research.” Responses were averaged into
305 a single score, with higher scores indicating higher levels of perceived cost of engaging in research.

306

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

313

314 **Previous Research Experience.** In order to better characterize the study sample and explore possible
315 differential effects of remote research experiences for students with different levels of research
316 experience, we asked students how much research experience they had prior to participating in the study.
317 Response options included: None, one semester or summer, two semesters or summers, three semesters or
318 summers, and more than three semesters or summers.

319

320 **Data Analysis**

321 Following the Anderson and Gerbing (1988) two-step approach, we first tested a confirmatory
322 measurement model before fitting our structural models. Our confirmatory measurement model specifies
323 the relationships between survey items and the latent variables they represent. Our structural models
324 estimate the effect of participating in a remote research program on student outcomes. To attain optimum
325 model fit for our measurement model, we followed an iterative process of model specification using
326 confirmatory factor analysis (CFA). To test our structural model, we used a multilevel modeling approach
327 because the data are clustered such that students are nested within programs. All analyses were conducted
328 in R version 4.0.1 and RStudio using lme4 (linear mixed effects modeling) and lavaan (latent variable
329 modeling) (Bates et al., 2014; Rosseel, 2012). Fixed-effect only models were estimated with maximum
330 likelihood estimation and mixed-effect models were estimated with restricted maximum likelihood
331 estimation, as is recommended by Theobald (2018). Conditional R^2 values, which take into account the
332 variance of both the fixed and random effects, were calculated using the MuMIn package for model
333 averaging (Bartoń, 2020). Random and fixed effects for each model, as well as AIC and R^2 values, are
334 reported.

335

336 **Assessment of Measurement Models.** We used several fit indices to assess how adequately our CFA
337 models reproduced their variance-covariance matrices. First, we report a chi squared test (χ^2) for each
338 model. Chi square is highly sensitive to misfit because it has strong assumptions, including that there is

339 no kurtosis in the data, which is a measure of the “tailedness” of the probability distribution of a real-
340 valued random variable (Kline, 2015). However, a significant chi square indicates misfit to some degree
341 (Credé & Harms, 2019), and so it is best practice to report it. We also include the root mean square error
342 of residuals (RMSEA), which approximates how well the model estimates the population covariance
343 matrix while favoring more parsimonious models. Higher values of RMSEA indicate poorer fit. Hu and
344 Bentler (1999) recommend an RMSEA cutoff value of 0.06. In addition, we chose to include the
345 standardized root mean square residual (SRMSR/SRMR) because it is sensitive to mis-specified
346 covariance structures. This means that a high SRMR value (greater than 0.08) in the absence of other
347 indications of misfit may indicate that the factor structure is mis-specified. Finally, we consider the
348 Tucker-Lewis index (TLI) and comparative fit index (CFI), which are incremental fit measures, meaning
349 that they compare model fit to the worst possible model. Higher values indicate better fit. Because TLI
350 and CFI are sensitive to mis-specified factor loadings, they are useful for evaluating the appropriateness
351 of survey items as representative of their latent variable (Hu & Bentler, 1999). A value of 0.90 or above is
352 recommended (Hu & Bentler, 1999).

353

354 In addition to fit indices, we evaluated the appropriateness of our measurement models based on factor
355 loadings and coefficient alpha values (see Tables S1-S3 in Supplemental Materials for factor loadings).
356 Factor loadings indicate the extent to which each survey item reflects its respective latent variable. A
357 minimum factor loading of 0.40 is recommended (Bandalos, 2018). Coefficient alpha is a measure of
358 internal consistency, or the degree or item correlation within the factor. Coefficient alpha values were
359 similar across timepoints; we report values that include both timepoints for each measure. Ultimately, we
360 balanced evidence from fit indices, factor loadings, and alpha values to determine our final measurement
361 models.

362

363 **Scientific Self-Efficacy.** The scientific self-efficacy scale demonstrated high internal reliability ($\alpha=0.92$).
364 Fit of the model was acceptable, $\chi^2(27)=140.839$ ($p<0.001$), RMSEA=0.137, SRMR=0.050, CFI=0.912,
365 TLI=0.883, although RMSEA is substantially higher than the recommended value of 0.05 and TLI is
366 slightly lower than the recommended value of 0.90. Given the high alpha value, the high factor loadings
367 (0.45-0.87), and the use of this scale in the study of other UREs, we opted to proceed as is with the
368 measure as a single factor. Item 2 (“Use computational skills [software, algorithms, and/or quantitative
369 technologies]”) produced a factor loading much lower than the second lowest factor loading (0.45 vs.
370 0.66). This result suggests that students responded differently to this item. However, removing this item
371 did not result in improved model fit, $\chi^2(28)=110.981$ ($p<0.001$), RMSEA=0.142, SRMR=0.043,
372 CFI=0.926, TLI=0.896. Moreover, we felt that this item captured information relevant to students’ remote
373 research experiences. Thus, we moved forward with the complete scientific self-efficacy measure as it
374 was administered to students.

375

376 **Scientific Identity.** The scientific identity scale also demonstrated high internal reliability ($\alpha=0.90$).
377 However, RMSEA, CFI, and TLI indicated poor model fit, $\chi^2(14)=176.429$ ($p<0.001$), RMSEA=0.228,
378 SRMR=0.096, CFI=0.792, TLI=0.688, with no clear cause of the model misfit. We attempted to remove
379 items and test a two-factor structures with no improvement in model fit. Thus, the factor structure of
380 scientific identity is still uncertain and may be sample dependent. Given the high alpha value, the high

381 factor loadings (0.52-0.90), and the use of this scale in the study of other UREs, we opted to proceed with
382 the measure as a single factor.

383

384 **Values and Cost.** We began by testing the factor structure of values with seven factors: values alignment,
385 intrinsic value, personal importance, cost, social utility, job utility, and life utility. Overall, factor loadings
386 were higher than the recommended minimum value of 0.40 (Bandalos, 2018), ranging from 0.473 to
387 0.949. Despite high factor loadings, model fit statistics indicated poor fit ($\chi^2(254)=747.528$ ($p<0.001$),
388 RMSEA=0.094, SRMR=0.090, TLI=0.816, CFI=0.844). Most factor correlations between the seven
389 factors were moderate to low; however, the factor correlation between intrinsic value and personal
390 importance was high ($r=0.848$, $p<0.001$). Therefore, we evaluated our values factor for sources of misfit.
391 Based on item content and factor loadings in our seven-factor model, intrinsic value appeared to be two
392 separate variables. The content of the first three items refers to enjoyment of research (e.g., “Research is
393 fun for me”) whereas the last three items are more value-oriented (e.g., “Performing well in research is
394 important to me”). In addition, factor loadings were stronger for the first three items (0.91, 0.95, 0.87)
395 than for the later three items (0.60, 0.57, 0.47). The differences in the strength between the first and
396 second half of the items suggests that the intrinsic value factor may be better represented as two factors.
397 Indeed, when we split this factor in two, factor loadings for the second half of items (intrinsic 2) increased
398 substantially (0.79, 0.89, 0.77), as did model fit ($\chi^2(247)=477.332$ ($p<0.001$), RMSEA=0.065,
399 SRMR=0.055, CFI=0.927, TLI=0.912).

400

401 **Higher-Order Confirmatory Factor Analysis.** To address concerns about measurement model fit and to
402 simplify the interpretation of our structural model analyses, we conducted a higher-order CFA.
403 Statistically, a “higher-order factor” models the covariance between two or more “lower-order factor(s),”
404 which are seen as manifestations of the higher-order factor. Higher-order factors are useful because they
405 tend to have higher predictive validity compared to narrower factors (Credé & Harms, 2015). They also
406 help address high inter-factor correlations (Table S5). High factor correlations ($r > 0.70$) are problematic
407 because they indicate too much overlap between constructs for them to be meaningfully different from
408 one another. Collapsing factors into one higher-level factor addresses this concern.

409

410 Because values alignment did not correlate highly ($r > 0.70$) with any other value-related factors, we
411 opted to represent values alignment under its own higher-order factor, HO1. Personal importance strongly
412 correlated with intrinsic 1 and intrinsic 2, thus we chose to represent personal importance, intrinsic 1, and
413 intrinsic 2 with a higher order factor, HO2. Because cost did not correlate strongly with other values-
414 related factors, we represented it with the higher-order factor HO3. Finally, we group together the three
415 forms of utility (social, job, life), based on their higher correlations and conceptual similarity, under HO4.
416 Although the fit of this four-factor, higher-order model was good according to fit indices (χ^2
417 (263)=525.357, $p<0.0001$, RMSEA=0.067, SRMR=0.068, CFI=0.917, TLI=0.906), there were two
418 Heywood cases (i.e., impossible factor loadings). The standardized loading for life utility onto HO4 was
419 1.010 and the standardized loading for personal importance on HO2 was 1.001. Furthermore, life utility
420 demonstrated a negative variance (-0.019), as did personal importance (-0.002) indicating misfit. HO2
421 was highly correlated with HO4 ($r=0.722$), so we decided to collapse the HO2 and HO4 factors to
422 eliminate this source of misfit.

423

424 Ultimately, we fit a values and cost model that contained three higher-order factors: “Alignment” or HO1
425 represents students’ perceptions of values alignment, “Perceived Benefits” or HO2 represents students’
426 perceptions of the intrinsic value, personal importance, and utility of engaging in research, and “Perceived
427 Costs” or HO3 represents students’ perceptions of the costs of engaging in research. For readability, we
428 refer to these factors as alignment, perceived benefits, and perceived costs. Fit of this model was
429 acceptable ($\chi^2(266)=577.278, p<0.0001, RMSEA=0.073, SRMR=0.080, CFI=0.902, TLI=0.889$), and it
430 eliminated the Heywood cases and negative factor variance. Thus, we decided to move forward with this
431 three-factor model (Figure 1; see Table S4 for higher-order factor loadings and Table S5 for higher-order
432 factor correlations).

433

434 **Assessment of Structural Models.** Given the exploratory nature of research on remote URE
435 programming, we tested three models for each student outcome variable. This approach allowed us to
436 estimate the effects of completing a remote URE to answer our research questions, and to explore whether
437 the program in which students completed their remote URE and their level of prior research experience
438 influenced their outcomes.

439

440 **Model 1.** This model allowed us to estimate the effects of completing a remote URE and to explore
441 program-level effects. Specifically, there were multiple students in each program, which means that
442 students’ experiences within programs are not independent of one another (i.e., data are clustered).
443 Therefore, Model 1 includes program as a random effect such that each grouping factor has its own
444 random intercept, meaning that each program’s level of our five latent variables starts at a different point
445 on the y-axis. It also includes a fixed effect of the URE. Thus, Model 1 can be stated as:

446

$$Y_{si}=(\beta_0 + b_{S,0s}) + \beta_1 X_i + e_{si}$$

447 In this model, X_i is our predictor variable, time, which takes on a value of 0 or 1 depending on whether i is
448 at time 1 (pre-program) or time 2 (post-program). e_{si} represents error. β_0 is the fixed effect of the slope, β_1
449 is the fixed effect of the intercept, and $b_{S,0s}$ are the random intercepts. Here we report the syntax used to
450 run our multilevel regression models. “Student outcome variable” represents each dependent variable,
451 “Time” represents the measurement timepoint, and “Program” represents the program where the student
452 participated in their URE. Program is treated as a categorical variable and the student outcome variable
453 and time are treated as continuous variables. The model syntax is as follows:

454

```
Model 1 <- lmer(Student outcome variable ~ Time + (1|Program))
```

455

456 **Model 2.** Students began their UREs with different levels of research experience, which could account for
457 variance in our dependent variables. Thus, we also included prior research experience as a fixed effect in
458 our models. Prior research experience is treated as a categorical variable. Thus, Model 2 can be stated as:

459

$$Y_{si}=(\beta_0 + \beta_{01} + b_{S,0s}) + \beta_1 X_i + e_{si}$$

460 Note that this model is the same as Model 1, but with the addition of a fixed intercept for research
461 experience, β_{01} . The model syntax may be written as:

462

```
Model 2 <- lmer(Student outcome variable ~ Time + Research Experience + (1|Program))
```

463

464 **Model 3 Equation.** Model 2 estimates the amount of variance accounted for by prior research experience.
465 However, it does not estimate the relative importance of different levels of research experience. In other
466 words, do more experienced researchers or less experienced researchers have more to gain from the URE?
467 To answer this question, we estimated Model 3, which includes research experience as a random
468 intercept. Thus, Model 3 can be stated as:

$$469 \quad Y_{si} = (\beta_0 + b_{S,0s} + b_{S,01s}) + \beta_1 X_i + e_{si}$$

470 Model 3 is the same as Model 1, but with the addition of an additional random intercept of research
471 experience, $b_{S,01s}$. The model syntax is as follows:

```
472 Model 3 <- lmer(Student outcome variable ~ Time + (1|Research experience) + (1|Program))
```

473

474 **Comparing Models with Akaike's Information Criteria (AIC).** In order to identify the most explanatory
475 and parsimonious models, we chose to compare fit between models using Akaike's information criteria
476 (AIC). AIC is a fit index that weights how well the model fits the data, while adding a penalty for the
477 number of parameters in the model. This penalty favors more parsimonious models, thereby balancing the
478 likelihood function given the observations and number of parameters. Smaller AIC values indicate better-
479 fitting models (Theobald, 2018). A difference of 2 or greater is necessary for establishing significantly
480 different AIC values (Burnham & Anderson, 2002).

481

482 We tested a total of three models for each student outcome and compared AIC values among them. For
483 each dependent variable, we began by testing a mixed effects model with a fixed effect of the URE and a
484 random intercept for the program (Model 1). Next, we added in a fixed effect of students' prior research
485 experience (Model 2). Finally, we tested this same model with a random effect of prior research
486 experience instead of a fixed effect (Model 3). Model 3 had the lowest AIC values and highest R^2 values,
487 and therefore is the primary model which we interpret in the following section. We also discuss the fixed
488 effects of prior research experience from Model 2 because the fixed effects inform the strength and
489 direction of the effect of UREs on student outcomes. Because we ran all three models seven times – once
490 for each dependent variable – we implemented a study-wide Bonferroni correction to interpret $p < 0.007$ as
491 significant.

492

493 RESULTS

494 Here we report the significant results of our Model 2 and 3 analyses (see Supplemental Materials for
495 Model 1 and non-significant results). We report intercepts (β_0) as a “baseline” of where students are with
496 respect to each construct at the start of their remote URE, slopes (β_1) to identify any changes pre- to post-
497 URE and characterize the size of any effects, and percentages of variance in student outcomes explained
498 by their program and their prior experience.

499

500 Scientific Self-Efficacy

501 We found that students began their UREs with a moderate level of scientific self-efficacy (Model 3:
502 $\beta_0 = 3.62$, $SE = 0.07$, $p < 0.001$), and their self-efficacy increased significantly from pre- to post-URE (Model
503 3: $\beta_1 = 0.64$, $SE = 0.08$, $p < 0.001$) (Table 3). We observed that students' program accounted for only 3% of
504 variance in scientific self-efficacy, which indicates that differences between programs had little if any

505 effect on students' self-efficacy development. We found that students' prior research experience
506 accounted for 9% of variance in their self-efficacy growth. Students who had three semesters or summers
507 of prior research experience (Model 2: $\beta=0.65$, $SE=0.15$, $p<0.0001$) or more than three semesters or
508 summers of prior research (Model 2: $\beta=0.71$, $SE=0.13$, $p<0.0001$) experienced significant gains in
509 scientific self-efficacy. Thus, we can infer that there was a positive effect of the remote URE on students'
510 scientific self-efficacy, and the effect was stronger for more experienced students.

511

512 In analyzing the self-efficacy data, we observed that the mean score for item 2 ("Use computational skills
513 [software, algorithms, and/or quantitative technologies]") is lower than for the other items in the scale:
514 $M=3.08$ 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
515 items). This suggests that, even though students are experiencing self-efficacy growth, students perceived
516 themselves to be less capable in their computational skills.

517

518 **Scientific Identity**

519 We found that students began their UREs at a high level of scientific identity (Model 3: $\beta_0=4.72$,
520 $SE=0.13$, $p<0.001$), which increased significantly from pre to post URE (Model 3: $\beta_1=0.24$, $SE=0.08$,
521 $p=0.005$) (Table 4). Again, we observed that the students' program accounted for a small amount (5%) of
522 the variance in their scientific identity growth. We also found that students' prior research experience also
523 accounted a small amount of the variance in scientific identity (5%), with students with more than three
524 semesters or summers of research (Model 2: $\beta=0.44$, $p=0.002$) experiencing significantly greater gains in
525 their sense of scientific identity.

526

527 **Values Alignment**

528 We found that students began their UREs with a very high level of values alignment (Model 3: $\beta_0=5.33$,
529 $SE=0.07$, $p<0.001$) and did not, as a group, change in their values alignment from pre to post URE
530 (Model 3: $\beta_1=0.01$, $SE=0.07$, $p=0.856$) (see Supplemental Materials). Program did not account for any
531 variance in values alignment, and prior research experience only accounted for 1%, which indicates that
532 the program and prior research experience did not affect students' values alignment. We observed that
533 students with more than three semesters or summers of prior research experience displayed small but
534 significant gains in values alignment (Model 2: $\beta=0.35$, $SE=0.11$, $p=0.002$).

535

536 **Perceived Benefits**

537 Similar to the values alignment, we found that students began their UREs at a very high level of perceived
538 benefits (Model 3: $\beta_0=5.32$, $SE=0.07$, $p<0.001$) and did not change in their perceptions of the benefits of
539 doing research from pre to post URE (Model 3: $\beta_1=-0.06$, $SE=0.06$, $p=0.287$) (see Supplemental
540 Materials). Program did not account for any variance in perceived benefits and prior research experience
541 only accounted for 1%. We observed that students with more than three semesters and summers of prior
542 research experience displayed small gains in perceived benefits (Model 2: $\beta=0.025$, $SE=0.10$, $p=0.011$).
543 However, the p value is greater than our adjusted alpha level of $p=0.007$, indicating a non-significant
544 effect.

545

546 **Perceived Costs**

547 We found that students began their UREs reporting a moderate level of perceived costs (Model 3:
548 $\beta_0=3.53$, $SE=0.14$, $p<0.001$). Their perceptions of costs did not change significantly pre- to post-URE
549 (Model 3: $\beta_1=-0.05$, $SE=0.13$, $p=0.68$) (Table 5). In contrast to the other outcomes, we found that
550 program accounted for 23% of variance in students' perceptions of the cost of research, indicating that the
551 programs in this study differentially affected students' perceptions of the costs of research. Students' prior
552 research experience did not account for any variance in their perceptions of the costs of research.

553

554 **Graduate School and Career Intentions**

555 On average, we found that students started their remote UREs already intending to attend graduate school
556 (Model 3: $\beta_0=4.38$, $SE=0.08$, $p<0.001$) (see Supplemental Materials). This intention did not change pre-
557 to post-URE (Model 3: $\beta_1=0.03$, $SE=0.07$, $p=0.717$). Likewise, students' intentions to pursue a career in
558 science were high before completing the program (Model 3: $\beta_0=4.23$, $SE=0.07$, $p<0.001$) and did not
559 change significantly pre- to post-URE (Model 3: $\beta_1=0.10$, $SE=0.08$, $p=0.196$). We found that program
560 accounted for only 2% of variance in graduate school intentions and 1% of variance in career intentions,
561 which suggests that programs did not have different effects on students' graduate and career intentions.
562 Similarly, prior research experience accounted for only 1% of variance in graduate school and career
563 intentions, which suggests that different amounts of research experience did not differentially affect
564 students' intentions. We observed that students with more than three semesters or summers of research
565 experience experienced gains in graduate school intentions (Model 2: $\beta=0.26$, $SE=0.12$, $p=0.031$) and
566 career intentions ($\beta=0.29$, $SE=0.12$, $p=0.019$). However, this effect was nonsignificant with our corrected
567 $p<0.007$.

568

569 **DISCUSSION**

570 In this study, we first sought to determine whether undergraduates who engage in remote research
571 programs experienced research-related social influence in terms of gains in their self-efficacy, scientific
572 identity, and values alignment (Research Question 1). We found that students in remote UREs
573 experienced outcomes that indicated their integration into the scientific community despite the remote
574 circumstances (Adedokun et al., 2013; Estrada et al., 2011; Robnett et al., 2015). Specifically, students
575 who completed remote UREs experienced significant gains in their scientific self-efficacy, and these
576 gains were largely due to their research experience and not to their particular URE program. Students in
577 remote UREs also experienced gains in their scientific identity, although these gains were more modest
578 than their self-efficacy gains and were related to their specific program. This finding suggests that remote
579 UREs can be productive environments for students' scientific identity development, but that programs are
580 either attracting or selecting students who differ in their scientific identity or that certain programs are
581 having greater influence on students' identity development. Students in our study did not experience any
582 changes in the extent to which they perceived their personal values as aligned with the values of the
583 scientific community. Rather, students in this study already felt that their personal values were well-
584 aligned with the values of the scientific community.

585

586 We also sought to determine the extent to which students in remote UREs experienced further integration
587 into the scientific community indicated by shifts in their intentions to pursue graduate education and

588 science research-related careers. For the most part, students in this study already intended to pursue a
589 graduate degree and a career in science research, and their intentions did not change significantly from
590 pre- to post-URE. It is encouraging that the challenges of engaging in research remotely did not dissuade
591 students from pursuing graduate school and research careers. Yet, it is also important to note that remote
592 research and perhaps UREs in general may not be a lever for changing students' plans to pursue graduate
593 education or science research careers because students who seek out or are selected into these programs
594 may already be firm in their intentions.

595

596 Although students in this study gained in their scientific self-efficacy, it is worth noting that students
597 started their UREs reporting less confidence in their computational skills. It is unclear whether students'
598 initial uncertainty about their computational skills is specific to remote research or unique to the last-
599 minute shift from away from bench or field research. As a reminder, most of the students in this study
600 were accepted into their programs *before* decisions were made to offer programs remotely. Regardless, it
601 appears that the programs in this study were able to support students' development of computational
602 skills.

603

604 In keeping with the Expectancy-Value Theory of motivation, we also sought to explore the extent to
605 which undergraduates in remote research programs shifted their perceptions of the benefits and costs of
606 doing research (Research Question 2). Students in this study already perceived high benefits and low
607 costs of research when they started their remote research and their perceptions did not change. Again, it is
608 encouraging that the challenges of engaging in research remotely did not dissuade students from the
609 benefits of research or increase their perceptions of the costs. Interestingly, students' programs appeared
610 to shape their perceptions of costs of doing research. It may be that some program contexts lessened
611 students' perceptions of costs and others exacerbated students' perceptions of costs. The types of
612 institutions that hosted the remote URE programs in our sample varied widely, from masters-granting
613 institutions to high research-intensity universities to non-degree granting research institutes. It may be that
614 differences in research mentor workloads or lifestyles or institutional cultures in these different
615 environments affected students' perceptions of the costs of doing research. Alternatively, it may be that
616 contextual differences between students' own undergraduate institutions and the institution that hosted
617 their remote URE program are influencing students' cost perceptions. Indeed, Duckworth and Yeager
618 (2015) have argued about the importance of considering context dependency of some measures. For
619 instance, a student who has done research at a more teaching-intensive institution and then participates in
620 a summer URE at a highly research-intensive university, or vice versa, may shift substantially in their
621 perceptions of what doing research entails and thus what opportunity costs they might experience if they
622 choose to continue in research.

623

624 **How much research experience is enough?**

625 Our results indicate that students with more prior research experience benefited more from remote UREs
626 compared to students with less research experience. Students with the most prior experience reported the
627 most substantial gains in scientific self-efficacy, scientific identity, values alignment, and graduate and
628 career intentions. This finding suggests that students may need at least two or three terms of research
629 experience before they start to realize positive gains from a summer remote URE. There are several
630 plausible explanations for why more experienced researchers realize greater gains in scientific self-

631 efficacy and scientific identity. One possibility is that self-efficacy functions as a positive feedback loop
632 or virtuous cycle. As students gain more experience, they become better at research and are willing to try
633 more things and put forth effort. As a result, they experience more research success and thus become
634 more confident in their abilities to do research. Alternatively, it may be that students who seek out
635 additional research experiences are primed to gain the most. It also may be that students with less research
636 experience do not feel efficacious and thus are less likely to seek out additional research experience,
637 thereby exerting a selection effect. This result provides at least some evidence that, if remote URE
638 programming continues, less experienced students should be prioritized for in-person UREs and more
639 experienced researchers should be prioritized for remote UREs. Alternatively, remote UREs could
640 develop and evaluate additional program elements aimed at better supporting of novice researchers.

641

642 **Comparison to In-Person UREs**

643 Overall, we found that students in this study realized scientific self-efficacy growth that resembled the
644 growth observed by Robnett and colleagues (2015) in their longitudinal study of students who completed
645 in-person UREs at colleges and universities across the country. Interestingly, the positive effects observed
646 by Robnett et al. (2015) took place over a period of four semesters of in-person research, while positive
647 effects we observed occurred in a much shorter period – an average of about nine weeks – in entirely
648 remote research. This result may be due to the intensity of the summer experience (~35-40 hours per
649 week) versus the less intense, more protracted nature of academic year UREs. Alternatively, the remote
650 nature of the programs in this study may have prompted mentors and program leadership to engage more
651 regularly or intentionally with students to ensure they can engage and make progress at a distance. In
652 addition, remote programming may have selected, intentionally or unintentionally, for mentors who were
653 most invested in undergraduate research and undergraduate researchers who were particularly primed to
654 invest time and effort, thereby maximizing the likelihood of students' experiencing favorable outcomes.

655

656 Our results differed to some extent from other longitudinal studies of in-person UREs. Estrada et al.
657 (2018) studied the effects of UREs on the self-efficacy, scientific identity, and values alignment of a
658 cohort of underrepresented minority students in their junior and senior years. Similar to our results, they
659 found that in-person UREs had a small but significant, positive effect on scientific self-efficacy and
660 scientific identity of these more advanced students. In contrast, they found that in-person UREs also had a
661 small but significant, positive effect on students' values alignment. Hernandez and colleagues (2020)
662 tracked a cohort of STEM students from historically well-represented backgrounds at a research-
663 intensive, public university throughout their four years of college. They also found that in-person research
664 experiences positively predicted scientific self-efficacy and scientific identity but failed to predict values
665 alignment among advanced students, similar to our results. In contrast to our results, Hernandez and
666 colleagues (2020) observed self-efficacy and identity growth among first- and second-year students. It is
667 possible that semester-long (or longer) research experiences are needed to promote these outcomes for
668 less experienced researchers. This would suggest that more experienced researchers are better suited for
669 summer research experiences. Alternatively, the benefits of engaging in undergraduate research early on
670 might not be evident until later in college. As Hernandez and colleagues (2020) note, early social
671 integration through mentorship and research experience exerts a reciprocal longitudinal influence on
672 future engagement with the scientific community.

673

674 **LIMITATIONS**

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

685
686 Another limitation relates to our measure of scientific identity, which demonstrated high internal
687 reliability based on coefficient alpha but suboptimal model fit. Moving forward, researchers should seek
688 to improve this measure by modifying item content and collecting additional validity evidence, including
689 its utility for discriminating among undergraduate students with more or less research experience. More
690 robust frameworks may be needed to better operationalize scientific identity, such as the Carlone and
691 Johnson framework, which conceptualizes scientific identity as a combination of social performance, self-
692 recognition as a “science person,” and knowledge and understanding of science content (Carlone &
693 Johnson, 2006).

694
695 Finally, there were limitations related to our sample, which was entirely comprised of biology students.
696 Therefore, our results may be unique to the discipline. Biology research may be more or less amenable to
697 remote research compared to other STEM disciplines. Moreover, as the full extent of the COVID-19
698 pandemic unfolded, students and mentors who chose to move forward with remote research may possess
699 different personality traits or differing levels of our variables of interest (i.e., scientific identity, scientific
700 self-efficacy) from those who opted out of remote research. Research topics themselves likely changed
701 during the transition to accommodate the remote research arrangement, so researchers who chose to move
702 forward with remote research may have conducted a different type of research than they originally
703 planned on. Lastly, data were collected during a time of social unrest in the United States during summer
704 of 2020. Awareness of social unrest and systematic racism may have affected the well-being of
705 participants, which may have influenced their experience in the remote URE program.

706 707 **CONCLUSION**

708 Perhaps the greatest advantage of remote research programs is that they open doors for students who may
709 not have the opportunity to participate in an in-person research program (Erikson *et al.*, in press). Remote
710 UREs can allow for more flexible scheduling and enable research participation without the additional
711 costs and logistics of travel and lodging. Thus, remote programs may be a viable method of expanding
712 access to UREs, especially among students who may find it difficult to travel. Although remote UREs
713 have many advantages, their appropriateness should be evaluated on a case-by-case basis and should be
714 considered alongside the advantages and disadvantages of in-person UREs. For example, certain types of
715 research (e.g., computational biology) may be more amenable to remote work. Particular research
716 mentors and undergraduates may be better able to navigate the unstructured nature of remote work.

717 Certain remote research environments may be more or less accessible for different individuals, such as
718 those who can sit and work on a computer for extended periods of time (Reinholz & Ridgway, 2020).
719 Certain personal situations may make remote research more difficult, such as whether individuals have
720 access to robust internet connections and quiet workspaces (Erikson *et al.*, in press). Finally, because
721 students are not able to complete bench work at home, remote UREs may aid in the development of a
722 different skillset than in-person UREs. Thus, students may benefit from completing both types of UREs
723 throughout their undergraduate degree in order to develop a wider variety of skills.

724

725 In summary, our work suggests that remote UREs can have a positive effect on student outcomes, but
726 they do not benefit all students equally. The benefits of remote UREs are larger for more experienced
727 researchers compared to less experienced researchers. Given that more experienced researchers benefitted
728 more from remote UREs compared to less experienced researchers, institutions may wish to prioritize
729 selection of less experienced researchers into in-person programs and more experienced researchers into
730 remote or hybrid programs. This would provide less experienced researchers with the supervision and
731 guidance needed to grow while allowing more freedom and flexibility to experienced researchers.
732 Institutions should also consider further developing programming to better meet the needs of novice
733 researchers.

734

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

742

743 **ACKNOWLEDGEMENTS**

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

752

753

754 **REFERENCES**

755

756 Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and
757 recommended two-step approach. *Psychological Bulletin*, 103(3), 411-423.

758 Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford
759 Publications.

760 Bartoń, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. [https://cran.r-](https://cran.r-project.org/web/packages/MuMIn/index.html)
761 [project.org/web/packages/MuMIn/index.html](https://cran.r-project.org/web/packages/MuMIn/index.html)

762 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). LME4: Linear Mixed-Effects Models Using
763 Eigen and S4. R Package version 1.1-4 <https://cran.r-project.org/web/packages/lme4/index.html>

764 Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference, a practical*
765 *information-theoretic approach (2nd ed)*. Springer.

766 Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of
767 color: Science identity as an analytic lens. *Journal of Research in Science Teaching: The Official*
768 *Journal of the National Association for Research in Science Teaching*, 44(8), 1187-1218

769 Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy
770 and identity in science career commitment among underrepresented minority students. *Journal of*
771 *Social Issues*, 67(3), 469-491.

772 Credé, M., & Harms, P. D. (2015). 25 years of higher-order confirmatory factor analysis in the
773 organizational sciences: A critical review and development of reporting recommendations.
774 *Journal of Organizational Behavior*, 36(6), 845–872. <https://doi.org/10.1002/job.2008>

775 Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than
776 cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
777 <https://doi.org/10.3102/0013189X15584327>

778 Erikson, O.A. et al. (in press). “How do we do this at a distance?!” A descriptive study of remote
779 undergraduate research programs during COVID-19. *CBE—Life Sciences Education*.

780 Estrada, M., Woodcock, A., Hernandez, P. R., & Schultz, P. W. (2011). Toward a Model of Social
781 Influence that Explains Minority Student Integration into the Scientific Community. *Journal of*
782 *Educational Psychology*, 103(1), 206–222. <https://doi.org/10.1037/a0020743>

783 Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015).
784 More value through greater differentiation: Gender differences in value beliefs about math.
785 *Journal of Educational Psychology*, 107(3), 663–677. <https://doi.org/10.1037/edu0000003>

786 Hernandez, P. R., Agocha, V. B., Carney, L. M., Estrada, M., Lee, S. Y., Loomis, D., ... & Park, C. L.
787 (2020). Testing models of reciprocal relations between social influence and integration in STEM
788 across the college years. *PloS one*, 15(9), 1-30. <https://doi.org/10.1371/journal.pone.0238250>

789 Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:
790 Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary*
791 *Journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>

792 Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.

793 Reinholz, D. L., & Ridgway, S. W. (2021). Access needs: centering students and disrupting ableist Norms
794 in STEM. *CBE—Life Sciences Education*, 20(3), es8. <https://doi.org/10.1187/cbe.21-01-0017>

- 795 Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among
796 undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science*
797 *Teaching*, 52(6), 847–867. <https://doi.org/10.1002/tea.21221>
- 798 Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical*
799 *Software*, 48, 1-36. <https://users.ugent.be/~yrosseel/lavaan/lavaanIntroduction.pdf>
- 800 Theobald, E. (2018). Students are rarely independent: When, why, and how to use random effects in
801 discipline-based education research. *CBE—Life Sciences Education*, 17(3).
802 <https://doi.org/10.1187/cbe.17-12-0280>
- 803
804

805
 806 **Table 1. Duration of URE programs.** Remote URE programs in this study varied in duration, with most
 807 being about 10 weeks long. *One program had staggered end dates with most students engaging in
 808 research for 9 weeks.

809

Duration in Weeks	Number of Programs
5	1
8	3
9	4*
10	12
11	2

810

811

812 **Table 2. Demographics of study participants.** In total, 227 students responded to both the pre- and post-
 813 survey, including 153 women, 69 men, and 4 individuals who identified as non-binary. Note that students
 814 were able to indicate multiple races or ethnicities, so race/ethnicity counts do not sum to the total sample
 815 size. With respect to parent education level, 79 students had a parent or guardian who did not attend
 816 college. There were 45 students who indicated that they had transferred to their current institution from
 817 another college or university.

818

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

819

820

821

822

823 **Table 3. Remote UREs and prior research experience, but not program, relate to student gains in**
 824 **scientific self-efficacy.** Students reported significantly higher levels of scientific self-efficacy from pre-
 825 to post-URE. Program had a very small effect on students' scientific self-efficacy gains. Students with at
 826 least three semesters of prior research experience made larger gains in scientific self-efficacy compared to
 827 students with less prior experience.

828

		Variance	Std. Deviation			
Model 2	Random Effect					
	Program	0.03	0.18			
		Estimate	Std. Error	DF	t-value	p-value
	Fixed Effect					
	Intercept	3.28	0.11	179.47	28.94	<0.0001
	URE	0.64	0.08	417.03	8.08	<0.0001
	Research Experience 1	0.18	0.13	434.36	1.39	0.167
	Research Experience 2	0.22	0.13	437.55	1.77	0.077
	Research Experience 3	0.65	0.15	437.44	4.41	<0.0001
	Research Experience 4	0.71	0.13	437.93	5.40	<0.0001
AIC	1138.07					
R ²	0.23					
Model 3	Random Effect	Variance	Std. Deviation			
	Program	0.03	0.18			
	Research Experience	0.09	0.30			
	Fixed Effect	Estimate	Std. Error	DF	t-value	p-value
	Intercept	3.63	0.15	5.36	24.30	<0.0001
	URE	0.64	0.08	416.97	8.08	<0.0001
	AIC	1135.67				
R ²	0.24					

829

830

831 **Table 4. Remote UREs, program, and prior research experience relate to student gains in scientific**
 832 **identity.** Students reported significantly higher levels of scientific identity from pre- to post-URE.
 833 Program and prior research experience had a very small effect on students' scientific identity gains;
 834 students with at least three semesters of prior research experience made larger gains in scientific identity
 835 compared to students with less prior experience.

836

		Variance	Std. Deviation			
Random Effect						
	Program	0.05	0.23			
		Estimate	Std. Error	DF	t-value	p-value
Fixed Effect						
Model 2	Intercept	4.52	0.13	176.41	35.91	<0.0001
	URE	0.24	0.08	420.91	2.81	0.005
	Research Experience 1	-0.11	0.14	435.68	-0.77	0.443
	Research Experience 2	0.28	0.14	438.38	2.11	0.036
	Research Experience 3	0.38	0.16	438.68	2.37	0.018
	Research Experience 4	0.44	0.14	439.07	3.09	0.002
AIC	1207.63					
R^2	0.13					
		Variance	Std. Deviation			
Random Effect						
	Program	0.05	0.23			
	Research Experience	0.05	0.22			
		Estimate	Std. Error	DF	t-value	p-value
Model 3	Fixed Effect					
	Intercept	4.72	0.13	7.24	37.49	<0.0001
	URE	0.24	0.08	420.14	2.81	0.005
AIC	1203.16					
R^2	0.13					

837

838

839 **Table 5. Student perceptions of the cost of doing research vary by program, but not by current or**
 840 **prior research experience.** Students reported no change in perceptions of the cost of doing research from
 841 pre- to post-URE and no differences in cost perceptions based on their prior research experience. Program
 842 had a significant and moderate effect on students' perceptions of the cost of doing research.
 843

		Variance	Std. Deviation			
Model 2	Random Effect					
	Program	0.2354	0.4852			
		Estimate	Std. Error	DF	t-value	p-value
	Fixed Effect					
	Intercept	3.62	0.20	110.47	18.14	<0.0001
	URE	-0.06	0.13	419.87	-0.46	0.646
	Research Experience 1	-0.16	0.21	431.63	-0.75	0.452
	Research Experience 2	-0.05	0.20	436.01	-0.26	0.798
	Research Experience 3	0.03	0.24	434.76	0.14	0.887
	Research Experience 4	-0.16	0.21	436.26	-0.75	0.453
	AIC	1574.53				
	R ²	0.12				
		Variance	Std. Deviation			
Model 3	Random Effect					
	Program	0.23	0.48			
	Research Experience	0.00	0.00			
	Fixed Effect	Estimate	Std. Error	DF	t-value	p-value
	Intercept	3.54	0.14	30.10	25.41	<0.0001
	URE	-0.06	0.13	423.58	-0.46	0.65
	AIC	1563.53				
	R ²	0.12				

844

845

846

847

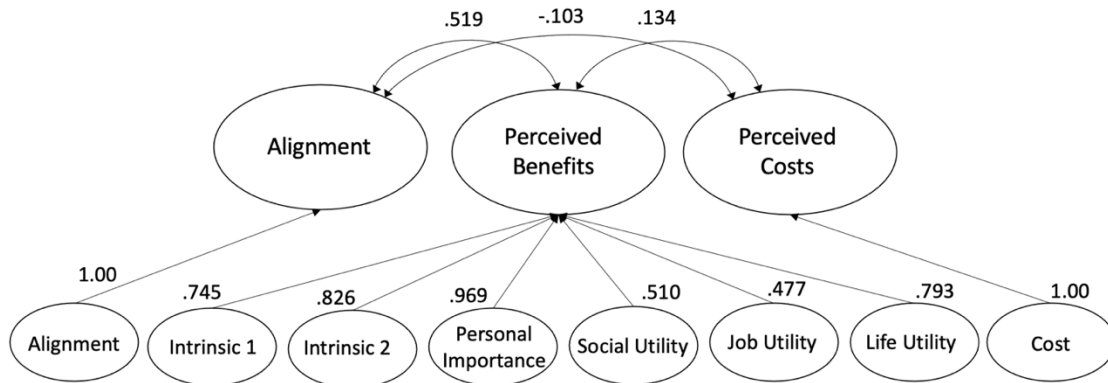
848

849 **Figure 1. Factor loadings and factor correlations for higher-order confirmatory factor analysis.**

850 Factor loadings and correlations are reported for Time 1 (pre-URE). Loadings for the higher-order factors
851 with only one lower-order construct (i.e., Alignment, Cost) will always be 1.00 and are not meaningful.
852 See Supplemental Materials for Time 2 factor loadings.

853

854



855