

A Comparison of Internal Dispositions and Career Trajectories after Collaborative versus Apprenticed Research Experiences for Undergraduates

Kyle J. Frantz,** Melissa K. Demetrikopoulos,[§] Shari L. Britner,^{||} Laura L. Carruth,[‡] Brian A. Williams,[¶] John L. Pecore,[#] Robert L. DeHaan,^{Ⓐ†} and Christopher T. Goode**

[†]Neuroscience Institute, Georgia State University, Atlanta, GA 30302-5030; [§]Institute for Biomedical Philosophy, Dunedin, FL 34697; ^{||}Department of Teacher Education, Bradley University, Peoria, IL 61625; [‡]Early Childhood and Elementary Education, Georgia State University, Atlanta, GA 30302-3980; [¶]Department of Teacher Education and Educational Leadership, University of West Florida, Pensacola, FL 32514; [Ⓐ]Division of Educational Studies, Emory University, Atlanta, GA 30322; ^{**}Department of Psychology, Georgia State University, Atlanta, GA 30302-5010

ABSTRACT

Undergraduate research experiences confer benefits on students bound for science, technology, engineering, and mathematics (STEM) careers, but the low number of research professionals available to serve as mentors often limits access to research. Within the context of our summer research program (BRAIN), we tested the hypothesis that a team-based collaborative learning model (CLM) produces student outcomes at least as positive as a traditional apprenticeship model (AM). Through stratified, random assignment to conditions, CLM students were designated to work together in a teaching laboratory to conduct research according to a defined curriculum led by several instructors, whereas AM students were paired with mentors in active research groups. We used pre-, mid-, and postprogram surveys to measure internal dispositions reported to predict progress toward STEM careers, such as scientific research self-efficacy, science identity, science anxiety, and commitment to a science career. We are also tracking long-term retention in science-related career paths. For both short- and longer-term outcomes, the two program formats produced similar benefits, supporting our hypothesis that the CLM provides positive outcomes while conserving resources, such as faculty mentors. We discuss this method in comparison with course-based undergraduate research and recommend its expansion to institutional settings in which mentor resources are scarce.

INTRODUCTION

Demographic characteristics of the U.S. biomedical research workforce fail to mirror the diversity in the U.S. general population. Several racial and ethnic minority groups, as well as individuals from disadvantaged economic and educational backgrounds, have been and continue to be severely underrepresented in fields related to science, technology, engineering, and mathematics (STEM; Chubin *et al.*, 2010; National Science Foundation [NSF], 2015). For example, the proportion of African Americans earning doctoral degrees in STEM fields is less than half of the proportion found in the U.S. population (NSF, 2015), and the proportion of Hispanic or Latino/a STEM PhD holders is less than one-third. Women also remain underrepresented at the highest levels of the STEM workforce (NSF, 2015). Such subpopulations are known as underrepresented groups (URGs) in STEM. In addition to, or perhaps because of, this relative lack of diversity in STEM, the U.S. continues to lag behind other nations in the proportion of students receiving undergraduate degrees in the natural sciences and

Elizabeth Ambos, *Monitoring Editor*

Submitted June 29, 2016; Revised September 27, 2016; Accepted October 26, 2016
CBE Life Sci Educ March 1, 2017 16:ar1
DOI:10.1187/cbe.16-06-0206

[†]Deceased.

*Address correspondence to: Kyle J. Frantz (kfrantz@gsu.edu).

© 2017 K. J. Frantz *et al.* CBE—Life Sciences Education © 2017 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 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

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

engineering, and falls behind on several measures of innovation (National Research Council [NRC], 2003, 2007).

Undergraduates in Research

Researchers attempting to broaden participation in STEM research careers and to improve U.S. standing in STEM worldwide have emphasized the importance of engaging undergraduate students in research (NRC, 2003; Hofstein and Lunetta, 2004; Espinosa, 2011). An authentic, undergraduate research experience (URE) can confer enormous and lasting benefits on undergraduates. Indirect measures, such as self-reported curiosity, independent learning, and confidence, as well as direct measures, such as retention, graduation, and course grades, improve with undergraduates' participation in research (reviewed by Osborn and Karukstis, 2009). UREs also produce measurable improvements in research and communication skills, personal and professional gains, and, especially, increased science identity (Seymour *et al.*, 2004; Russell *et al.*, 2007; for a review, see Sadler *et al.*, 2010). Furthermore, UREs promote growth in the personal and professional precursors to career success in STEM, such as tolerance for obstacles, communication, and creativity (Lopatto, 2009). Finally, UREs increase interest in STEM careers, and participation in UREs predicts that students will take the steps necessary to attain these careers (e.g., application to and matriculation into PhD programs; Russell *et al.*, 2007). Of significant note, these well-documented gains are just as robust in students from URGs as they are among students from groups that are well-represented in STEM (Lopatto, 2004). Indeed, students from URGs who engage in research during the academic year are more likely to pursue a STEM PhD, even after controlling for educational background, intended major, and parents' education (Carter *et al.*, 2009).

Several researchers have considered the mechanisms through which UREs confer benefits on students. The theoretical model that has come to the foreground most frequently is based on a construct central to Bandura's (1977) social cognitive theory: self-efficacy. Defined as confidence in one's own ability to carry out a specific task or perform within a specific domain of skills, self-reported self-efficacy in domains such as science or mathematics is predictive of future behavior in those domains, including career choice and taking steps necessary to attain those careers (Betz and Hackett, 1983; Chemers *et al.*, 2001; Bakken *et al.*, 2006). Moreover, self-efficacy for specific career-related skills and tasks forms the basis for social cognitive career theory (SCCT; Lent *et al.*, 1994).

SCCT has been applied in a variety of contexts, including academics (Chemers *et al.*, 2001). For more than 30 years, it has been used to predict success in STEM careers, and particularly to understand the underrepresentation of women and various minority groups therein (Betz and Hackett, 1983; Gainor and Lent, 1998). This theoretical model may be used to understand how UREs can be so important to students considering STEM careers, particularly students from URGs. Chemers *et al.* (2011) surveyed hundreds of undergraduates, mostly from underrepresented racial and/or ethnic groups, and found that self-efficacy for scientific research mediates the relationship between UREs and intent to persist in STEM careers. Furthermore, science identity mediates this relationship between scientific research self-efficacy and intent to persist in STEM. Thus, a research experience may increase scientific research self-efficacy, which

strengthens science identity, which in turn promotes commitment to a science career (Chemers *et al.*, 2011). Other explorations on the role of internal dispositions in career decision making vary in specific sequences of growth and progress, but students' views of themselves appear critical in most studies. In fact, the predictive ability of reported internal dispositions may extend beyond traditional and purportedly objective predictors of career success, such as grades and standardized test scores (e.g., McGee and Keller, 2007). Their importance is consistent among demographic groups, suggesting that continued lack of diversity in STEM could result from a lack of opportunity to engage in real research among students from URGs.

Typical UREs involve the research apprenticeship of a student under a faculty mentor. Mentees are often assigned to ongoing projects under the direct supervision of a faculty member, graduate student, or postdoctoral fellow in a one-to-one relationship. Many UREs occur in summer programs such as NSF Research Experiences for Undergraduates. Other programs hosted at the home institution may extend into the academic year, resulting in progressively increasing ownership of and responsibility for the project by the mentee. While beneficial for those involved in these paired relationships, this strategy does limit access to research, particularly at institutions that focus on undergraduate education rather than on high-intensity research endeavors by faculty. Minimal research availability may be explained by faculty workload and reward structures emphasizing instruction over research, lack of research infrastructure, and related low levels of research funding. Many such institutions are community colleges, historically Black colleges and universities, Hispanic-serving institutions, and others that also enroll significant numbers of students from URGs. This disproportionately limits the number of available research opportunities for students from URGs, thereby potentially contributing directly to the lack of diversity among biomedical researchers.

As positive outcomes from undergraduate research are detailed and limited access for key undergraduate populations are described, however, outstanding solutions to the problem are emerging (e.g., Cejda and Hensel, 2009; Perna *et al.*, 2009). For example, course-based undergraduate research experiences (CUREs) offer an alternative to the traditional, mentored apprenticeship. Over the past 12 years, the demand for UREs and a national call for the integration of inquiry-based methods into science curricula (NRC, 2003; Lopatto *et al.*, 2008; American Association for the Advancement of Science, 2011; President's Council of Advisors on Science and Technology, 2012) have resulted in creation and adoption of a wide variety of CUREs around the nation. They typically facilitate student-driven inquiry and discovery, enabling students to generate or shape a research question that is either unique to a team or class or part of a national model. Some prime examples in bioscience include the Phage project (Hatfull *et al.*, 2006), the Freshman Research Initiative at the University of Texas at Austin (Simmons, 2014), and the Center for Authentic Science Practice in Education (Weaver *et al.*, 2007). A national network of CUREs funded by NSF has facilitated the implementation and evaluation of CUREs in the United States (CUREnet, n.d.).

Participation in CUREs confers many of the same benefits that result from the traditional, mentored, research apprenticeships described earlier (Lopatto *et al.*, 2008; Shaffer *et al.*, 2010; Thiry *et al.*, 2012). It may therefore serve as a means to

level the playing field for students from URGs (Banger and Brownell, 2014). For example, CUREs increase the extent to which students self-identify as scientists, promote persistence toward STEM careers, improve understanding of both discipline-specific content and the nature of science, and even improve graduation rates (Shaffer *et al.*, 2010; Harrison *et al.*, 2011; Thiry *et al.*, 2012; Beck *et al.*, 2014; Rodenbusch *et al.*, 2016). Moreover, CUREs within laboratory courses do more to promote research self-efficacy and persistence in STEM careers than traditional “cookbook” laboratory research assignments (Russell and Weaver, 2011; Brownell *et al.*, 2012).

Lacking in current education research on UREs in general and CUREs in particular, however, are sufficient studies that feature direct comparisons of CUREs with traditional, mentored apprenticeships. Studies that involve random assignment of participants to one or the other curricular model appear to be entirely absent. Some of the few existing direct comparisons do provide promising information on the ability of CUREs to provide viable and effective alternatives to traditional apprenticeship UREs. For example, students who participated in the multi-institution Genomics Education Partnership (CURE) reported similar levels of independent thinking and motivation to learn compared with those who participated in a traditional summer URE (Shaffer *et al.*, 2010). They even described themselves as “active learners” to the same degree. CURE and traditional URE participants also reported similar learning gains, increased interest in their disciplines, and satisfaction with their respective experiences (Shapiro *et al.*, 2015). On tests of higher-order cognitive skills, the CURE students also caught up from an initial lower precourse baseline to reach the same level as students in the URE, suggesting that a CURE can help close achievement gaps. In a CURE centered on a rain forest expedition, computational linguistic analysis of student interview data revealed that students indicated greater emotional commitment and personal connection with their projects than did a control group engaged in a traditional URE (Hanauer *et al.*, 2012). However, none of these three direct comparisons involved random assignment; participants either self-selected into CUREs versus traditional UREs, or comparisons were made using existing data. Rigorous evaluation of CUREs requires controlled comparisons akin to randomized control trials in clinical research (Auchincloss *et al.*, 2014). In fact, Linn *et al.* (2015) directly urged researchers to compare CUREs and UREs with appropriate controls, specifically including random assignment to curriculum models. Corwin *et al.* (2015) called for deeper consideration of how CUREs exert their positive impact. The present study helps to answer this call.

Behavioral Research Advancements in Neuroscience, the BRAIN Program

As described previously (Frantz *et al.*, 2006; Britner *et al.*, 2012; Goode *et al.*, 2012), we created a novel program that compares elements of CUREs with those of traditional, mentored URE apprenticeships in the context of a summer training program in neuroscience. Behavioral Research Advancements in Neuroscience (BRAIN) is a 10-week, paid, intensive summer research program with a competitive application process. The overall goals of the BRAIN program are to engage undergraduates in research and to ignite and sustain serious interest in science-related careers. We created two program models with the intent

of directly comparing a novel, team-based, collaborative learning model (CLM) with the traditional, mentored apprenticeship (AM). Perhaps most importantly for the present report, students admitted to the BRAIN program were randomly assigned to either the CLM or the AM. After a few days to allow students to move in and adjust to the local environment, the program curriculum began with 1 week of intensive classroom instruction in basic neuroscience, shared by all participants, followed by 9 weeks of neuroscience laboratory research in either the CLM or the AM group. The introductory classroom instruction addressed cellular and molecular neuroscience and behavioral neuroscience, using activities, lectures, and hands-on mini-experiments (~9:00 am to 5:00 pm daily over 5 or 6 days). During the subsequent 9 weeks, all participants were expected to work 35 hours per week in their laboratory settings. They also attended weekly 4-hour professional development workshops on topics including science ethics, science writing, poster presenting, diversity in science career opportunities, graduate school preparation, stress reduction, and time management. The program culminated in the preparation of a written report (in the form of a mini-research proposal for the CLM or a journal article for the AM) and preparation of a research poster to be presented and judged at a closing research symposium. On successful completion of program requirements, each participant received a stipend of \$3000 or \$4000 paid in increments.

Participants in the CLM all convened in a single dedicated laboratory (with neighboring seminar rooms and computers) to engage in various research techniques using an invertebrate animal model (red swamp crayfish [*Procambarus clarkii*]). This species was chosen due to the extensive body of literature available on its cellular and molecular mechanisms of behavior, the relative simplicity of its nervous system, ease of care, and low-level Institutional Animal Care and Use Committee oversight. Approximately eight instructors were deployed over 8 weeks for the CLM (e.g., two faculty, three postdocs, three graduate students), with two or three present at any given time. They led demonstrations and experiments that required participants to use the following techniques: observation of animal behavior, anatomical dissection, histological staining, electrophysiological recording, RNA extraction from nervous tissue, quantitative PCR, and protein detection. During the first 5 weeks in the CLM, daily activities generally consisted of 1- to 2-hour introductions to new material (via lecture, demonstration, and discussion related to assigned readings), review of protocols, and initiation of experimentation in self-selected teams of two to four participants, with assistance from instructors. Although all research teams used similar techniques in a given week, their specific experimental questions were based on individual team interests. During the last 3 weeks in the CLM, each team designed and conducted its own pilot investigation on a unique topic chosen by team members. Usually, only one mentor was present during this period, but several instructors and mentors reviewed ideas, read research proposals, provided guidance, and assisted with data collection during individual team meetings, consultations, and progress updates attended by all CLM participants. Weekly “journal clubs” facilitated comprehension of peer-reviewed articles on crayfish neurobiology.

Participants in the traditional AM joined new or ongoing research projects in more than 30 different laboratories at five local research institutions. BRAIN program administrators

exerted no influence over the nature of their research experiences, except that mentors were recruited based on nominations from the community that they traditionally provide strong research opportunities for undergraduates. Based on submission of weekly time sheets signed by mentors, participants fulfilled the expectation to conduct research activities for 35 hours per week, but daily schedules were designed individually by participants and mentors as they deemed fit for the diverse research paradigms, laboratories, and institutions that comprised the apprenticeship experiences.

In a pilot study (Frantz *et al.*, 2006), we compared the CLM with the AM summer research curricular models to determine whether students demonstrated gains in either or both models. Confidence with neuroscience concepts and confidence with research skills both increased significantly from the beginning to the middle of the program and from the middle to end with no significant differences between program models. While attitudes toward science remained flat for both subgroups, attitudes toward neuroscience improved significantly, also without significant differences across models. Expanding on this strong foundation, the current report summarizes a full test of whether the CLM confers the same benefits as the AM on undergraduate researchers. We conducted a 4-year study with additional 4- to 7-year follow-up of an undergraduate student population from around the nation, with a large proportion of participants from URGs. The specific URGs of focus in the present report are racial and ethnic minority groups underrepresented in STEM. Participants were randomly assigned to one of the two program models, and we used a mixed, qualitative and quantitative methodology to test whether the CLM produced outcomes at least as positive as the AM. Here, we include the results of our quantitative investigation; partial preliminary results with the first three cohorts were previously published, including in-depth explanations of the instruments used in assessment (Goode *et al.*, 2012) and an in-depth qualitative case study of four participants (Britner *et al.*, 2012).

METHODS

Participants

Four cohorts of approximately 40 undergraduate students each were recruited from around the nation in 2009, 2010, 2011, and 2012 for a total of 155 participants. Our selection criteria favored members of demographic groups underrepresented in STEM. Self-reported ethnicity was as follows: African descent/African American (49), Asian descent/Asian American (25), Caucasian (46), Hispanic or Latino/a (18), Native American/Pacific Islander (1), other (5), and 11 initially elected not to provide racial/ethnic information. Seven participants, including the five who initially indicated “other” subsequently declared race and ethnicity, and one of those was in a group underrepresented in STEM fields. One student had a physical disability. Others may have had documented cognitive or social disabilities, or non–documented disabilities, but did not declare them on application to this program. Overall, ~45% were from demographic groups underrepresented in STEM. Almost 66% were women. Selection criteria favored research novices, and the participant population included freshmen (27), sophomores (65), juniors (46), and seniors (17) for a ratio of ~60% freshmen or sophomores and 40% juniors or seniors.

For the stratified random assignment, accepted students were categorized into groups within each cohort by race/ethnicity, sex, academic year, and descending grade point average scores. They were then randomly assigned to either the CLM or the AM. Further balance across CLM and AM was checked for distribution of in-state/out-of-state home institution, prior research experience, and number of relevant courses completed (in that order of priority), and switches to the assignments were made if they could further balance the treatment groups on those additional variables without disrupting balance on the primary variables of concern. Students were then invited into their assigned program models and were not provided with an opportunity to switch to the other model. In this manner, two balanced treatment groups $n = 76$ (CLM) and $n = 79$ (AM) were created. All data collection from these participants was conducted with approval of the Georgia State University Institutional Review Board and included appropriate informed consent.

Mentors

We recruited a diverse group of faculty members and advanced trainees (e.g., postdoctoral fellows and graduate students) as research mentors and instructors for both program models using local advertisements and targeted electronic communication each summer. To fulfill various teaching and mentoring roles in a single laboratory facility, the CLM included a series of faculty, postdoctoral fellow, and graduate student instructors with expertise in specific research methods, each participating for 1 or 2 of the 5 weeks of guided instruction. These individuals were in the laboratory facility with participants for approximately 8 hours per day, 5 days per week, except during professional development workshops, and with occasional evening and weekend electronic communications about lab procedures and preparations. The instructors were complemented by a leadership team of senior faculty mentors (who were also BRAIN program leaders) who met with the participants at least weekly to mentor and facilitate specifically the development of the teams and their individual research hypotheses, methods, analyses, and reporting. Finally, a single, dedicated laboratory manager served as the daily mentor with the longest duration and highest frequency of interaction with the CLM participants. This individual had more than 280 contact hours with the participants, out of the 400-hour summer program. The instructional team members varied from year to year.

Within the AM, a structured mentor-matching process preceded participant arrival to the program site. Briefly, recruited mentors provided brief descriptions of research projects to be carried out that summer, mentor names and institutions were removed, and participants were asked to read the descriptions and submit their top five choices in rank order of preference. Program administrators then provided matches based on rankings and statements of interest from the program application. Preprogram communication was encouraged, and participants met their mentors in person at a meet-the-mentor luncheon during the orientation week. Faculty mentors assigned participants to ongoing research projects within their own research teams, and BRAIN program leaders were not involved in structuring the AM research experience. Although faculty members were the primary contacts in the laboratory, it was typical for some student participants to

work closely on a daily basis with graduate students or postdoctoral fellows.

Measures and Procedure

Electronic surveys were used to measure the following internal dispositions: scientific research self-efficacy (SRSE); leadership/teamwork self-efficacy (LTSE); science identity (SCIID); science and neuroscience anxiety (SA and NA); and commitment to science careers (COMMIT). The instruments used to measure these constructs are described in detail by Goode *et al.* (2012). Briefly, SRSE, LTSE, COMMIT, and SCIID were measured using surveys adapted from Chemers *et al.* (2001). SA and NA were measured with surveys adapted from Britner (2008). Electronic surveys of each of the constructs above were solicited before the program began, at the midpoint, and again at the end of the 10-week BRAIN program. Alumni tracking was conducted through membership in private social media groups and direct electronic mail, phone, and/or in-person communications, with the most recent update occurring ~4–7 years after the end of the summer program participation.

Data Analysis

Internal Reliability and Correlations in Survey Instruments.

Each of the survey instruments used to measure the internal dispositions contained multiple response items. To establish internal reliability in these instruments, we computed coefficient alpha at pre-, mid-, and postprogram, collapsing across program type and cohort. To explore how individual dispositions related to one another, we tested correlations between scores on each disposition instrument from the preprogram survey (SPSS, 2015).

Program and Time Effects. We used mixed, within- and between-subjects, factorial analyses of variance (ANOVAs) in SPSS (SPSS, 2015) to examine the effects of time (pre-, mid-, or postprogram, the within-subjects factor) and program type (CLM or AM) on the six internal attributes listed above. Because Mauchly's test for sphericity was failed for all dependent measures except neuroscience anxiety, Huynh-Feldt-adjusted degrees of freedom were used for those analyses. Where main effects or interactions were significant, post hoc paired comparisons were made using the Bonferroni adjustment.

Gender and Representation Effects and Interactions. To test for effects of student self-identified belonging to racial and/or ethnic groups underrepresented in STEM, we dummy coded our race/ethnicity variable to indicate whether the self-identified race/ethnicity was well represented or underrepresented in STEM. To test for effects of time, gender, and representation/underrepresentation in the sciences and for interactions among these factors, we used $3 \times 2 \times 2$ mixed ANOVAs. The data for all attributes failed Mauchly's test for sphericity, and Huynh-Feldt degrees of freedom are reported.

Predictors of Commitment to Science. To determine how internal attributes interacted to best predict commitment to a science career after controlling for preprogram commitment, we created a multiple regression model with preprogram COMMIT in the first block and postprogram COMMIT as the crite-

riion variable. In the second block of this regression, we entered postprogram SRSE, LTSE, SCIID, SA, and NA stepwise.

Analysis of Alumni Career Status. Alumni tracking of academic, preprofessional, or professional positions enabled identification of four general categories of current status: non-science (e.g., business consultant, software developer, lawyer, not working), clinical (e.g., medicine, dentistry, clinical psychology), clinical/research (e.g., MD/PhD, public health), and research (e.g., technician, graduate research assistant, postdoctoral fellow). Kindergarten through 12th-grade teaching and tutoring were classified as non-science, whether the teaching was science or not, in order to maintain focus on scientific research in the science categories rather than science teaching, writing, or other related endeavors. We compared these groups' survey responses at each time point to determine whether our measures of their internal attributes predicted their future career paths. We used a χ^2 test to determine whether students entered different career paths on the basis of program type, gender, or representation.

We computed change over the course of the program in internal dispositions of interest (SRSE, LTSE, SCIID, SA, NA, and COMMIT) by subtracting preprogram scores from postprogram scores. ANOVA was used to test whether measures of these dispositions at any time point and whether change in disposition over the course of the program significantly predicted later alumni career status. Analysis of covariance (ANCOVA) was used to control for preprogram scores when significant differences among career status groups in internal disposition change scores were detected.

RESULTS

Internal Reliability and Correlation in Survey Instruments

The values of coefficient alpha ranged from 0.84 to 0.96 for pre-, mid-, and postprogram measures of SRSE, LTSE, SCIID, SA, NA, and COMMIT. A correlation matrix reveals that all internal dispositions correlate significantly with one another (Figure 1). Measures of anxiety correlate negatively with other internal dispositions, whereas all other correlations are positive.

Scientific Research Self-Efficacy

The main effect of time on SRSE was significant, $F(1.75, 250.84) = 111.01$, with a large effect size, partial $\eta^2 = 0.564$ (Figure 2A). Post hoc testing revealed that SRSE increased significantly from pre- to midprogram, $p < 0.001$, and again from mid- to postprogram, $p < 0.001$. There was no main effect of program type on SRSE, $F(1, 143) = 0.412$, $p = 0.522$. The interaction was significant, $F(1.75, 250.84) = 3.43$, $p = 0.040$. Bonferroni-adjusted confidence intervals (95%) around the program means at each time point revealed that participants in the CLM reported significantly greater SRSE than those in the AM at the midprogram point only.

Leadership/Teamwork Self-Efficacy

The main effect of time on LTSE was significant, $F(1.87, 266.90) = 3.75$, $p = 0.028$, but small, partial $\eta^2 = 0.026$ (unpublished data). In fact, our post hoc paired comparisons revealed no significant differences in LTSE among the three time points. There was no main effect of program, $F(1, 143) = 0.299$,

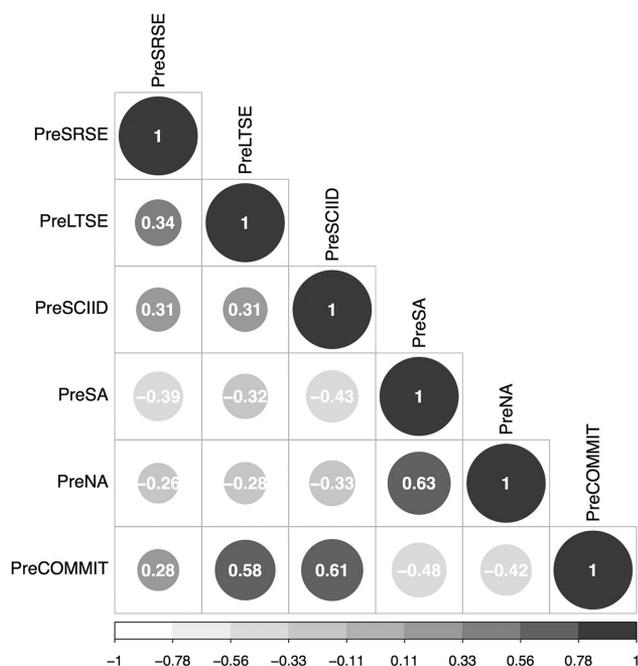


FIGURE 1. Correlation matrix of all preprogram measures of internal dispositions: scientific research self-efficacy (PreSRSE), leadership/teamwork self-efficacy (PreLTSE), science identity (PreSCIID), science anxiety (PreSA), neuroscience anxiety (PreNA), and commitment to a science career (PreCOMMIT). All correlations are significant with $p < 0.05$.

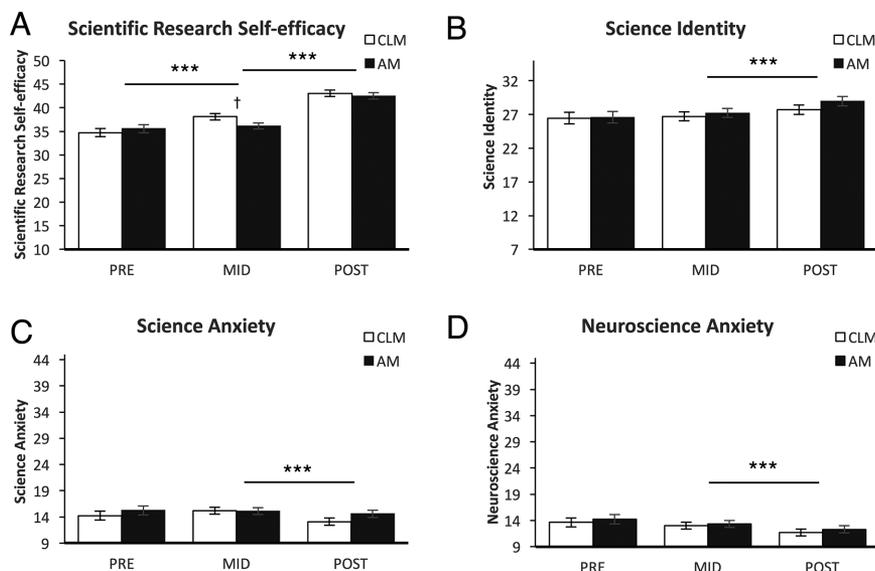


FIGURE 2. Average scores of students assigned to the collaborative learning model (CLM, open bars) and the apprenticeship model (AM, filled bars) at the beginning (PRE), middle (MID), and end (POST) of the program for (A) scientific research self-efficacy, (B) science identity, (C) science anxiety, and (D) neuroscience anxiety. The y-axis ranges reflect the range of the instrument scale. Error bars indicate 1 SE. Asterisks indicate significant differences between the indicated time points: ***, $p < 0.001$; dagger (†) indicates a significant difference between program types at midprogram.

$p = 0.585$, nor was the interaction significant, $F(1.87, 266.90) = 0.896$, $p = 0.403$. Scores collapsed across program models were $M = 43.31$, $SE = 0.41$ at preprogram; $M = 44.29$, $SE = 0.41$ at midprogram; and $M = 44.21$, $SE = 0.45$ at postprogram.

Science Identity

The main effect of time on SCIID was significant, $F(1.85, 263.90) = 10.85$, $p < 0.001$, partial $\eta^2 = 0.071$ (Figure 2B). Post hoc testing revealed that SCIID did not increase significantly from pre- to midprogram, $p = 0.071$, but increased significantly from mid- to postprogram, $p < 0.001$. There was no main effect of program, $F(1, 143) = 0.772$, $p = 0.381$, nor was the interaction significant, $F(1.85, 263.90) = 1.07$, $p = 0.339$.

Science and Neuroscience Anxiety

The main effect of time on SA was significant, $F(2, 286) = 8.10$, $p < 0.001$, partial $\eta^2 = 0.054$ (Figure 2C). Post hoc testing revealed that SA decreased significantly from mid- to postprogram, $p < 0.001$. There was no significant difference between program types, $F(1, 143) = 0.459$, $p = 0.499$, nor was the interaction significant, $F(2, 286) = 2.83$, $p = 0.061$.

The main effect of time on NA was significant, $F(1.92, 274.95) = 13.44$, $p < 0.001$, partial $\eta^2 = 0.086$ (Figure 2D). Post hoc testing revealed that NA decreased significantly from mid- to postprogram, $p = 0.001$. There was no significant difference between program types, $F(1, 143) = 0.187$, $p = 0.666$, nor was the interaction significant, $F(1.92, 274.95) = 0.078$, $p = 0.919$.

Commitment to Science Careers

There was no significant main effect of time on COMMIT, $F(1.88, 269.12) = 1.11$, $p = 0.328$; no program differences, $F(1, 143) = 0.210$, $p = 0.647$; and no significant interaction, $F(1.88, 269.12) = 0.091$, $p = 0.861$ (unpublished data). Scores collapsed across program models were $M = 30.70$, $SE = 0.35$ at preprogram; $M = 31.25$, $SE = 0.41$ at midprogram; and $M = 31.26$, $SE = 0.44$ at postprogram.

Gender and Representation Effects

For SRSE, the three-way interaction among time, gender, and representation was significant, $F(1.78, 22.97) = 3.25$, $p = 0.046$ (Figure 3). The effect size was medium, partial $\eta^2 = 0.025$. Post hoc testing revealed that men from URGs reported significantly greater SRSE than women from well-represented groups at each time point, $p < 0.05$, and this effect was somewhat large, Cohen's $d = 0.73$. At midprogram, women from URGs reported significantly lower SRSE than men generally, $p < 0.05$. This effect was medium, Cohen's $d = 0.48$.

Scientific Research Self-efficacy by Gender, Representation, and Time

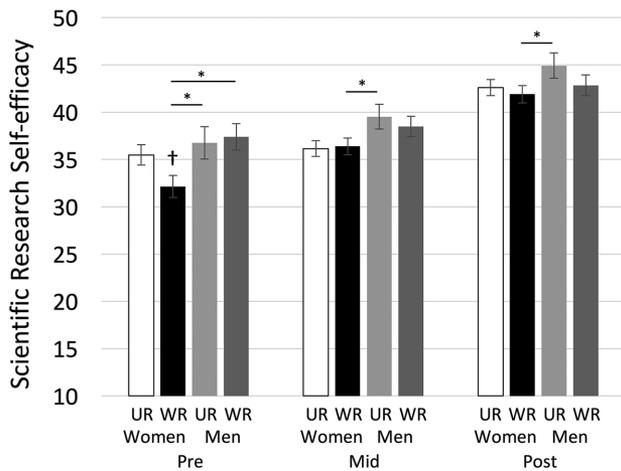


FIGURE 3. Average SRSE scores at pre-, mid-, and postprogram for women from URGs (open bars), women from well-represented groups (filled bars), men from URGs (light gray bars), and men from well-represented groups (dark gray bars). Error bars indicate 1 SE. Asterisk indicates significant differences between the indicated groups: $p < 0.05$; dagger (†) indicates a significant difference between well-represented women and both under- and well-represented men: $p < 0.05$. UR, underrepresented; WR, well represented.

There was a main effect of gender on SCIID, $F(1, 20.5) = 4.313, p = 0.040$, with men reporting significantly greater SCIID ($M = 28.42, SE = 0.69$) than women ($M = 26.66, SE = 0.50$) regardless of time point. The effect was medium, Cohen's $d = 0.40$ (unpublished data).

There was a significant three-way interaction among time, gender, and representation for COMMIT, $F(1.90, 237.57), p = 0.047$. The effect size was medium, partial $\eta^2 = 0.025$. Post hoc testing revealed that women from well-represented groups reported significantly less COMMIT ($M = 29.69, SE = 0.82$) than women from URGs ($M = 32.09, SE = 0.76$) and less COMMIT than men from both well-represented groups ($M = 31.82, SE = 0.98$) and URGs ($M = 32.50, SE = 1.20$) at the end of the program only, $p < 0.05$. These effect sizes were small to medium, Cohen's $d = 0.33, 0.30$, and 0.36 , respectively (unpublished data).

Predictors of Commitment to Science

The final stepwise regression model retained, in order, postprogram SCIID, SA, LTSE, and NA, $F(4, 146) = 28.23, p < 0.001, R^2 = 0.436$. The results for each predictor in the model appear in Table 1. We found it curious that SRSE was excluded from this model, as it was shown previously to predict commitment to science careers (Chemers *et al.*, 2011). We regressed SRSE directly on COMMIT and found the relationship to be significant, $F(1, 149) = 25.08, p < 0.001, R^2 = 0.144$. However, mediation analysis showed that this relationship was mediated entirely by SCIID, as shown in Figure 4.

The second, stepwise block of our multiple regression retained only postprogram SCIID and LTSE as significant predictors of postprogram COMMIT, after controlling for preprogram COMMIT in the first block, $F(3, 143) = 39.83, p < 0.001, R^2 = 0.455$. Semipartial correlation coefficients revealed that SCIID made the greatest unique contribution to the overall model, $r = 0.381$ (Table 1).

Alumni Career Status

Among the 155 participants who completed the BRAIN program, 116 (80%) have maintained contact and shared current career status. Of those, 106 (91%) remain in research and/or science-related careers. Research careers (e.g., PhD programs, postdoctoral fellowships, research MS programs, research technicians) are pursued or engaged by 55 alumni (47%); research/clinical careers (e.g., MD/PhD, MPH) are pursued or engaged by 14 alumni (12%); and clinical careers (e.g., medical doctorate, dental doctorate, clinical psychology PhD) are pursued or engaged by 37 alumni (32%).

Preprogram COMMIT differed among alumni career status groups, $F(3, 119) = 3.33, p = 0.022$. Post hoc testing revealed that students who ended up in a research/clinical career had reported significantly greater preprogram COMMIT ($M = 31.24, SD = 4.61$) than students in non-science careers ($M = 27.17, SD = 6.53$), $p < 0.05$ (Figure 5). This was a large effect, Cohen's $d = 0.77$. Students who ended up in non-science careers reported significantly greater SA at the end of the program ($M = 15.60, SD = 5.03$) than those in clinical, research/clinical, or research careers ($M = 13.21, SD = 4.2; M = 12.03, SD = 2.59; M = 13.12, SD = 4.83$; Figure 5). These were all large effects, Cohen's $d = 0.72, 1.15$, and 0.71 , respectively. Students who

TABLE 1. Multiple regression analysis preprogram COMMIT was regressed on postprogram COMMIT in the first block, with five postprogram predictors of postprogram COMMIT then entered stepwise in the second block, with SCIID and LTSE being the only predictors retained in the final model^a

Model		Unstandardized coefficient		Standardized coefficient	t	p	Correlation		
		B	SE	β			Zero-order	Partial	Semipartial
1	(Intercept)	13.211	2.941		4.492	<0.001			
	Preprogram COMMIT	0.585	0.095	0.456	6.166	<0.001	0.456	0.456	0.456
2	(Intercept)	6.473	2.658		2.435	0.016			
	Preprogram COMMIT	0.351	0.086	0.273	4.054	<0.001	0.456	0.320	0.255
	Postprogram SCIID	0.492	0.065	0.506	7.510	<0.001	0.605	0.530	0.472
3	(Intercept)	1.183	3.340		0.354	0.724			
	Preprogram COMMIT	0.337	0.085	0.263	3.967	<0.001	0.456	0.315	0.245
	Postprogram SCIID	0.427	0.069	0.440	6.175	<0.001	0.605	0.459	0.381
	Postprogram LTSE	0.170	0.067	0.172	2.536	0.012	0.404	0.207	0.157

^aCriterion variable: postprogram COMMIT.

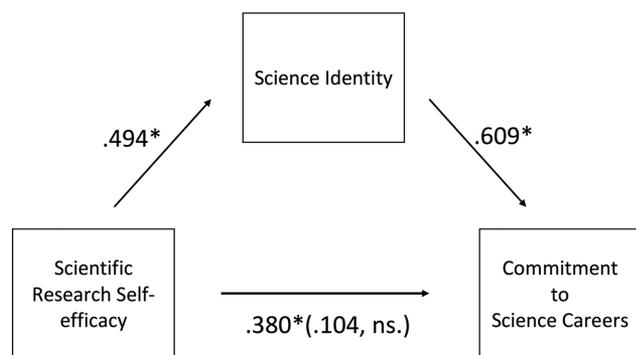


FIGURE 4. Mediation analysis. The relationship between scientific research self-efficacy and commitment to a science career is fully mediated by science identity. Standardized β coefficients are shown; asterisk indicates a significant difference from 0: $p < 0.001$.

ended up in non-science careers reported significantly greater NA at the end of the program ($M = 15.00$, $SD = 5.37$) than those in clinical, research/clinical, or research careers ($M = 11.32$, $SD = 3.73$; $M = 9.53$, $SD = 2.90$; $M = 11.51$, $SD = 5.34$; Figure 5). These were all large effects, Cohen's $d = 0.80$, 1.27 , and 1.02 , respectively.

Change in SRSE over the course of the program (post – pre) differed among alumni career status groups, but this effect only approached significance, $F(3, 116) = 2.65$, $p = 0.052$. Nevertheless, we followed this finding with an analysis of covariance, controlling for preprogram SRSE, which we suspected might be driving this effect. After controlling for preprogram SRSE, the effect on SRSE change was no longer close to significance, $F(3, 115) = 0.754$, $p = 0.522$.

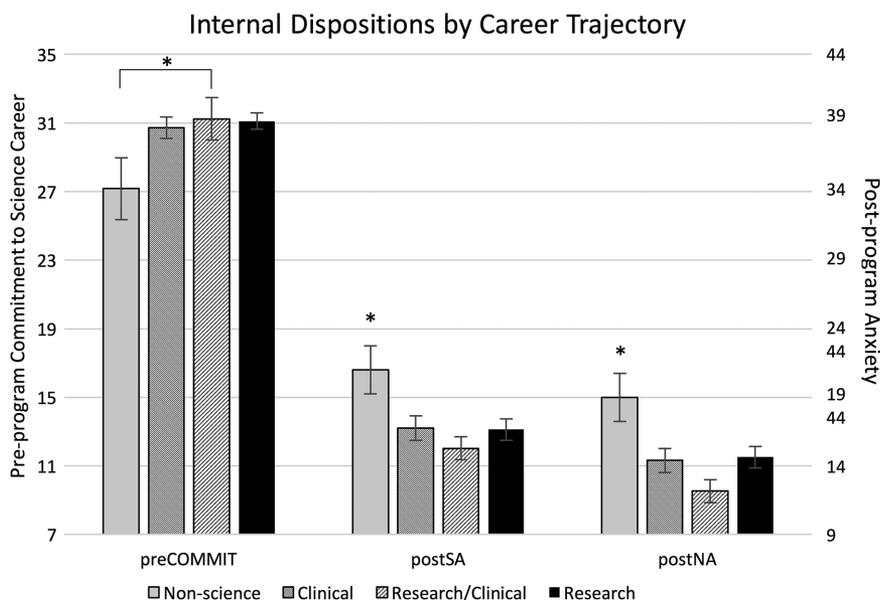


FIGURE 5. Average preprogram commitment to a science career (preCOMMIT), postprogram science anxiety (postSA), and postprogram neuroscience anxiety (postNA) for students categorized in non-science, clinical, research/clinical, or research career types. Error bars are 1 SE. Asterisk indicates significant differences between nonscience and research/clinical for preCOMMIT and between nonscience and all other career types for the anxiety measures, all with $p < 0.05$.

Chi-square analysis revealed that there was no significant effect of program type, gender, or representation on alumni career status; that is, our observed frequencies within each career category did not deviate significantly from expected frequencies, regardless of program type, $\chi^2 = 0.978$, $p = 0.807$; gender, $\chi^2 = 0.799$, $p = 0.850$; or representation, $\chi^2 = 0.731$, $p = 0.866$ (see Figure 6). Because of the group differences in postprogram COMMIT reported above, we compared alumni career status specifically among underrepresented women, well-represented women, underrepresented men, and well-represented men. There were no significant differences among these groups.

DISCUSSION

Using a stratified random assignment of program participants to either a collaborative, team-based model (CLM) or a traditional apprenticeship model (AM) for summer undergraduate research programming, we have shown that both program models produce similar short-term positive outcomes for students, in the form of enhancing internal dispositions predictive of retention in research. In addition, more than 90% of known program alumni remain in science career paths 4–7 years after program participation, with 47% in research tracks specifically, and again no differences across program models. These results fully support our hypothesis that the CLM produces benefits at least as positive as the traditional AM. Although the CLM was not technically a CURE, because the CLM occurred as a competitive-admission, paid summer research internship alongside the traditional AM, the present study achieves its goal of filling a gap in current knowledge on the efficacy of CURE curricula by comparing outcomes after random assignment of participants to a CURE-like curriculum versus a traditional research apprenticeship. The results are encouraging, in that the present CLM outcomes lend strong support to the recommendation that CUREs should be integrated into the undergraduate STEM curriculum nationwide.

Specific short-term program gains included significant gains in internal dispositions previously predictive of retention in science careers, including scientific research self-efficacy and science identity. Moreover, both science and neuroscience anxiety decreased significantly. None of these overall benefits was dependent on assignment into the CLM or AM program model.

In terms of effect size, increased scientific research self-efficacy was the most robust student outcome. Detailed analysis of self-efficacy trajectories revealed that participants in the CLM reported dispositions that matched those in the AM at the start of the program. Yet gains among the CLM participants actually outpaced gains among AM participants, revealed by higher scientific research self-efficacy scores among CLM participants at the mid-program time point, perhaps attributable

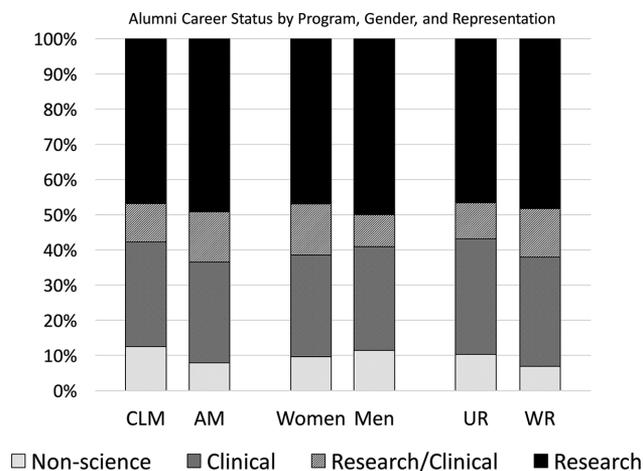


FIGURE 6. Alumni career status by program, gender, and representation. Percentage of participants with known career status ($n = 116$) in non-science, clinical, research/clinical, and research careers by program type (CLM, collaborative learning model; AM, apprenticeship model), gender, and representation (UR, underrepresented; WR, well represented). No significant differences between program types, genders, or representation groups were found.

to elements in the CLM curriculum that bolstered research confidence earlier in the summer than experiences in traditional apprenticeships. Nonetheless, both CLM and AM participants closed the program with similarly heightened scientific research self-efficacy. Bandura (1977) described four sources of self-efficacy: mastery experiences, vicarious experiences, social persuasion, and physiological/emotional states. On the basis of an in-depth qualitative analysis of four participants, students in our program attributed gains in self-efficacy to mastery experiences, such as engagement in tasks, interpretation of results, and developing beliefs in their capacity to perform tasks in accordance with the results, for example: “[Having been] through the process once ... next time I’ll know what I’m doing and what to look for in a project and how to design one” (Britner *et al.*, 2012, p. 283). Although this analysis was based on pre/postprogram interviews, without a midprogram interview to provide higher temporal resolution, it is possible that influential mastery experiences occurred earlier in the summer for the CLM participants, perhaps in the form of acquiring independent research skills in the first few weeks of the crayfish lab and/or generating hypotheses and designing experiments by the midprogram time point. This time course confirms a pilot data set using a slightly different assessment measure (Frantz *et al.*, 2006).

In a regression model, postprogram scientific research self-efficacy correlated significantly with postprogram commitment to science careers. Mediation analysis, however, revealed that this relationship was entirely mediated by postprogram science identity. These results mirror the relationships among research experiences, self-efficacy, and science identity described by Chemers *et al.* (2011), who proposed that it is the psychological reaction or personal application of the research experience that builds the relationship between self-efficacy and commitment to a science career. On the other hand, Estrada *et al.* (2011) found that science self-efficacy did

not directly predict commitment to a science career and that science identity was the stronger predictor. Although they did not directly test the mediation model described by Chemers *et al.* (2011), their results are consistent with that model; self-efficacy was related to science identity, which was predictive of commitment to science careers. In the present analysis, we report that science identity fully mediates the relationship between science self-efficacy and commitment to science careers. Another notable difference between the present outcomes and those of Estrada *et al.* (2011) lies in the role of leadership/teamwork self-efficacy. It played an important role in mediating commitment to science among their participants but did not change over time in our program and did not mediate any of our outcome measures. A ceiling effect of very high leadership/teamwork self-efficacy among the present participants may explain this apparent contradiction. Moreover, Chemers *et al.* (2011) found that leadership/teamwork self-efficacy was a significant mediator only for graduate trainees and not for undergraduates at academic levels similar to our participants.

Although the internal dispositions measured here did not differ robustly by program model, some gender and racial/ethnic group effects warrant consideration. First, this research experience closed a gender gap. Women from well-represented groups came into the program with low levels of scientific research self-efficacy on preprogram surveys, compared with men from well-represented groups. Also women from URGs reported lower self-efficacy than all men. By the mid- and/or postprogram surveys, however, these gender differences were eliminated, as all groups increased self-efficacy and reached similar higher levels. These results are encouraging in terms of the ability of a summer experience to ameliorate gender differences in self-efficacy. A second gender difference, however, provides warning that women still need focused retention efforts to encourage identification as scientists. All women reported lower science identity than all men, a difference that remained significant throughout the program. A third gender difference also provides warning. Women from well-represented groups reported lower levels of commitment to science by the end of the program compared with women from URGs and all men. These alarming short-term outcomes suggest that women may still be at higher risk for defection from science compared with men. Interventions based on these results alone might consist of continued aggressive recruitment of women and deeper attention to the factors that do increase identity and commitment in science among women in particular. As alarming as these lower levels of science identity and commitment to science among women are, however, none of these gender differences on short-term internal dispositions resulted in any long-term differences in alumni career status among the present participants. In other words, alumnae and alumni were equally likely to remain in specific science career paths, as discussed in the next section.

The last gender difference of note is that the present subset of men from URGs reported higher levels of scientific research self-efficacy than women from well-represented groups at all three time points. Given high confidence already at the start of the program, these results may relate to the program’s effective recruitment strategy rather than to the program effects per se. In other words, it could be the case that, in order for males from

URGs to apply or meet admissions standards for this or similar programs, they must already possess very high confidence in their abilities to do the tasks associated with the program. African-American males comprise a population group at very high risk for dropping out of school, especially out of science-related studies (Quality Education for Minorities Network, 2010; Bidwell, 2015). Thus, programs such as BRAIN may aid recruitment and retention efforts by providing strong growth opportunities specifically for these highly confident and high academically performing men.

Long-Term Career Status

Given the overall increases in scientific research self-efficacy and science identity that were observed in this participant population, SCCT suggests we should anticipate persistence in the science career domain (Lent *et al.*, 1994). This is in spite of the fact that we did not detect an increase in commitment to science careers over the course of the program (which we attribute to the high preprogram commitment to science scores near the top of our scale). In fact, an impressive 91% of the program alumni whose current status is known (or 68% of all alumni) remain in science-related career paths ~4–7 years after the summer program's end. Even more important for the program's goals to sustain interest and success in biomedical research, 47% of these alumni remain in science research *per se*. When we tested for characteristics in the summer program that aligned with alumni career status, we discovered that those students with higher preprogram commitment to science careers were more likely to end up in research/clinical positions later (i.e., MD/PhD or MPH programs). We also discovered that students reporting higher postprogram science and neuroscience anxiety were more likely to end up in the non-science category 4–7 years later. While these results are limited to our student population, they may suggest maximum return on investment among students who report high commitment to science on program entry. They conversely suggest a need for targeted interventions with students who maintain relatively higher anxiety in science contexts.

Study Limitations

Although comprehensive and especially rich when considered together with our preliminary quantitative report (Goode *et al.*, 2012) and our qualitative case study (Britner *et al.*, 2012), this report has some limitations. Perhaps of greatest concern is the observation that increased scientific research self-efficacy during the program did not predict retention in a research career path. This finding is counterintuitive, given some strong studies to the contrary, as cited earlier. The fact of the matter is that this correlation just missed significance at $p = 0.052$ and thus does not provide strong evidence that the predictive relationship is non-existent in the present population. Moreover, deep exploration of the data revealed that the smallest change in research self-efficacy from pre- to postprogram was recorded among those participants who were retained in research career paths, which was also counterintuitive, until we recognized that this subgroup actually entered the program with the highest research self-efficacy score. Although that score was not significantly higher than others when tested in the initial ANOVA (time by program type), it did severely weaken the association between change in research self-efficacy and alumni career status when it was used as a cofactor in an ANCOVA analysis (change in self-efficacy by

alumni career status). Overall, the present results lend marginal support to the ability of change in scientific research self-efficacy to predict long-term career progress.

Another limitation to this study is that 155 participants is a relatively small population for a quantitative education research study. Consequently, we could not conduct some desirable analyses, such as nesting the self-identified student research teams in the CLM population into a multilevel model of student outcomes. Our career status reporting is even further compromised, as we can confirm status for only 116 of the 155 participants (75%) at the time of submission. For a full alumni survey study, currently in progress, continued attempts to re-establish contact with more of our alumni will likely increase our contact pool, and it will also include participants from other program cohorts not in this present analysis (e.g., participants from 2005, 2013, and 2014, $n = 65$). A final limitation to consider here is that participant activities after completion of the BRAIN program were not taken into account in the correlations between short-term internal dispositions and current career status. BRAIN may have been just one of the research experiences in the academic careers of these participants, and certainly is just one influential factor. The full alumni study, currently in progress, will address this issue, even providing an opportunity to test for correlations between the number or quality of research experiences and longer-term career outcomes.

Implications and Future Directions

In this study, we explored multiple critical factors in pathways toward biomedical research careers. We and others should continue investigating other influential factors as well, such as student views on mentoring (Pfund *et al.*, 2006, 2015); goal-setting (Hernandez *et al.*, 2013); students' critical thinking, creativity, and scientific reasoning (Lawson *et al.*, 2007; DeHaan, 2009); and grit or other nontraditional predictors of retention in research (McGee and Keller, 2007; Duckworth and Gross, 2014). Similarly, while we used SCCT as a unifying theory, other overarching models should also be explored in the explanation of research experience outcomes, such as identity, cultural capital, and communities of practice (Gazley *et al.*, 2014).

The present surveys were administered and analyzed in an anonymous, coded procedure for research purposes, but we propose that they could be implemented in the future to facilitate targeted advising practices or tests of career readiness. If they were administered openly throughout a program or as part of a guided self-assessment procedure, then specific scores or patterns could be identified to trigger alerts for academic advisers or program administrators to launch interventions in the form of course work advising, targeted skill development, or socioemotional support. Furthermore, the present results suggest that early undergraduate males from URGs who reveal high scientific research self-efficacy could be targeted for recruitment into strong research opportunities, whereas specific retention interventions could be designed for any undergraduates whose anxiety and/or commitment to science scores remain relatively low by the end of a CURE or a summer program.

Generally, the present results provide strong support for implementation of the CLM and/or CURE-type research experiences for undergraduates nationwide. These data are therefore relevant to faculty members and administrators at both research and teaching institutions. The CLM facilitates gains for students,

while limiting the number of well-established professional research scientists required to provide authentic research experiences, thereby addressing a current problem in providing undergraduates with research experiences. Challenges related to scaling up the CLM are surmountable and include identification of sufficient teaching laboratory space, purchase of equipment and consumables, and recruitment of several well-informed and research-experienced laboratory instructors and mentors. With specific regard to financial investments in the present CLM versus AM, direct costs in the program budget were slightly higher for offering the CLM than the AM, due to payment of the laboratory manager and purchase of lab materials. Yet the budget for a summer program like this certainly cannot account for the institutional costs to maintain the 20 research laboratories in which the AM participants were placed. For consideration, each AM research group had many salaried researchers from principal investigator to undergraduate research assistants, in addition to well-equipped research facilities, research grants, overhead costs, and other research-related costs. Thus, the institutional capacity to host large numbers of novice researchers in an AM may be lacking at many institutions that emphasize undergraduate education rather than research productivity. Unfortunately for diversity in the STEM workforce, this is especially likely to be the case at institutions with significant portions of students from groups underrepresented in STEM, such as community colleges, historically Black colleges and universities, Hispanic-serving institutions, and so on. A primary take-home message from this report, therefore, is that faculty members at all institutions should be at liberty to choose which program model best fits their academic environment. Ultimately, diversification of the types of undergraduate research programs available to students around the nation will help to diversify the biomedical research workforce.

ACKNOWLEDGMENTS

We thank Elizabeth Weaver, Laurie Murrah-Hanson, and Emily Hardy for their BRAIN program coordination and Rob Poh and Phillip Gagne for their input to data management and experimental design. Tiffany Oliver and Michael Black were excellent CLM laboratory mentors. Donald Edwards contributed his deep knowledge and appreciation for the crayfish animal model. Many graduate students and postdoctoral research associates were excellent CLM instructors, and many faculty members, graduate students, and postdoctoral research associates were excellent AM mentors. The late Robert L. DeHaan, coauthor, was an outstanding mentor and collaborator for all involved in this program and research study, participants and coauthors alike. This program and the research on it were funded by the National Institutes of Health (1R25GM097636; 1R01GM085391), the NSF (IBN-9876754), Georgia State University, Emory University, the Georgia Institute of Technology, Clark Atlanta University, Morehouse College, and Spelman College.

REFERENCES

- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Final Report, Washington, DC.
- Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Lawrie G, McLinn CM, Pelaez N, Rowland S, et al. (2014). Assessment of

- course-based undergraduate research experiences: a meeting report. *CBE Life Sci Educ* 13, 29–40.
- Bakken LL, Byars-Winston A, Wang MF (2006). Viewing clinical research career development through the lens of social cognitive career theory. *Adv Health Sci Educ* 11, 91–110.
- Bandura A (1977). *Social Learning Theory*, New York: General Learning Press.
- Bangera G, Brownell S (2014). Course-based undergraduate research experiences can make scientific research more inclusive. *CBE Life Sci Educ* 13, 602–606.
- Beck C, Butler A, da Silva KB (2014). Promoting inquiry-based teaching in laboratory courses: are we meeting the grade? *CBE Life Sci Educ* 13, 444–452.
- Betz NE, Hackett G (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *J Vocat Behav* 23, 329–345.
- Bidwell A (2015). African-American men: the other STEM minority. *US News and World Report*. www.usnews.com/news/stem-solutions/articles/2015/05/07/african-american-men-the-other-stem-minority.
- Britner SL (2008). Motivation in high school science students: a comparison of gender differences in life, physical, and earth science classes. *J Res Sci Teach* 45, 955–970.
- Britner SL, Williams BA, Pecore JL, Gagne P, Demetrikopoulos MK, Poh R, Carruth LL, Goode CT, DeHaan RL, Frantz KJ (2012). Portraits of science self-efficacy: four undergraduate women in a summer research experience. *J Women Minor Sci Eng* 18, 273–293.
- Brownell SE, Kloser MJ, Fukami T, Shavelson R (2012). Undergraduate biology lab courses: comparing the impact of traditionally based “cookbook” and authentic research-based courses on student lab experiences. *J Coll Sci Teach* 41, 36–45.
- Carter FD, Mandell M, Maton KI (2009). The influence of on-campus, academic year undergraduate research on STEM Ph.D. outcomes: evidence from the Meyerhoff scholarship program. *Educ Eval Policy Anal* 31, 441–462.
- Cejda BD, Hensel N (2009). *Undergraduate Research at Community Colleges*. Washington, DC: Council on Undergraduate Research.
- Chemers MM, Hu LT, Garcia BF (2001). Academic self-efficacy and first-year college student performance and adjustment. *J Educ Psychol* 93, 55–64.
- Chemers MM, Zurbriggen EL, Syed M, Goza BK, Bearman S (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *J Soc Issues* 67, 469–491.
- Chubin DE, DePass AL, Blockus L (Eds.) (2010). *Understanding Interventions That Encourage Minorities to Pursue Research Careers*, Bethesda, MD: UnderstandingInterventions.org. Retrieved from www.nigms.nih.gov/training/reports/Documents/InterventionsReport2009.pdf.
- Corwin LA, Graham MJ, Dolan EL (2015). Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. *CBE Life Sci Educ* 14, es1.
- CUREnet (n.d.). Course-Based Undergraduate Research Experience. <https://curenet.cns.utexas.edu/content/about-curenet> (accessed 17 May 2016).
- DeHaan RL (2009). Teaching creativity and inventive problem solving in science. *CBE Life Sci Educ* 8, 172–181.
- Duckworth A, Gross JJ (2014). Self-control and grit: related but separable determinants of success. *Curr Dir Psychol Sci* 23, 319–325.
- Espinosa LL (2011). Pipelines and pathways: women of color in undergraduate STEM majors and the college experiences that contribute to persistence. *Harvard Educ Rev* 81, 209–240.
- Estrada M, Woodcock A, Hernandez PR, Schultz PW (2011). Toward a model of social influence that explains minority student integration into the scientific community. *J Educ Psychol* 103, 206–222.
- Frantz KJ, DeHaan RL, Demetrikopoulos MK, Carruth LL (2006). Routes to research for novice undergraduate neuroscientists. *Cell Biol Educ* 5, 175–187.
- Gainor KA, Lent RW (1998). Social cognitive expectations and racial identity attitudes in predicting the math choice intentions of Black college students. *J Couns Psychol* 45, 403–413.
- Gazley JL, Remich R, Naffziger-Hirsch ME, Keller J, Campbell PB, McGee R (2014). Beyond preparation: identity, cultural capital, and readiness for graduate school in the biomedical sciences. *J Res Sci Teach* 51, 1021–1048.
- Goode CT, Britner SL, Gagné P, Pecore JL, Demetrikopoulos MK, Williams BA, Carruth LL, DeHaan RL, Frantz KJ (2012). Scientific research self-efficacy

- among undergraduates: current contexts and approaches for measurement. In: *The Psychology of Self-Efficacy*, ed. SL Britner, Hauppauge, NY: Nova, 21–52.
- Hanauer DI, Frederick J, Fotinakes B, Strobel SA (2012). Linguistic analysis of project ownership for undergraduate research experiences. *CBE Life Sci Educ* 11, 378–385.
- Harrison M, Dunbar D, Ratmanský L, Boyd K, Lopatto D (2011). Classroom-based science research at the introductory level: changes in career choices and attitude. *CBE Life Sci Educ* 10, 279–286.
- Hatfull GF, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet* 2, 835–847.
- Hernandez PR, Schultz PW, Estrada M, Woodcock A, Chance R (2013). Sustaining optimal motivation: a longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *J Educ Psychol* 105, 89–107.
- Hofstein A, Lunetta VN (2004). The laboratory in science education: foundations for the twenty-first century. *Sci Educ* 88, 28–54.
- Lawson AE, Banks DL, Logvin M (2007). Self-efficacy, reasoning ability, and achievement in college biology. *J Res Sci Teach* 44, 706–724.
- Lent RW, Brown SD, Hackett G (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *J Vocat Behav* 45, 79–122.
- Linn MC, Palmer E, Baranger A, Gerard E, Stone E (2015). Undergraduate research experiences: impacts and opportunities. *Science* 347, 1261757.
- Lopatto D (2004). Survey of Undergraduate Research Experiences (SURE): first findings. *Cell Biol Educ* 3, 270–277.
- Lopatto D (2009). *Science in Solution: The Impact of Undergraduate Research on Student Learning*, Tucson, AZ: Research Corporation for Science Advancement.
- Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung HM, Du C, Eckdahl T, Goodman AL, Hauser C, Jones CJ, *et al.* (2008). Undergraduate research: Genomics Education Partnership. *Science* 322, 684–685.
- McGee R, Keller JL (2007). Identifying future scientists: predicting persistence into research training. *CBE Life Sci Educ* 6, 316–331.
- National Research Council (NRC) (2003). *BIO2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academies Press.
- NRC (2007). *Rising above the Gathering Storm, Energizing and Employing America for a Brighter Economic Future*, Washington, DC: National Academies Press.
- National Science Foundation (2015). *Women, Minorities, and Persons with Disabilities in Science and Engineering*. www.nsf.gov/statistics/2015/nsf15311/start.cfm (accessed 17 May 2016).
- Osborn JM, Karukstis KK (2009). The benefits of undergraduate research, scholarship, and creative activity. In: *Broadening Participation in Undergraduate Research: Fostering Excellence and Enhancing the Impact*, ed. M Boyd and J Wesemann, Washington, DC: Council on Undergraduate Research, 41–53.
- Perna L, Lundy-Wagner V, Drezner ND, Gasman M, Yoon S, Bose E, Gary S (2009). The contribution of HBCUS to the preparation of African American women for STEM careers: a case study. *Res High Educ* 50, 1–23.
- Pfund C, Pribbenow CM, Branchaw J, Lauffer SM, Handelsman J (2006). The merits of training mentors. *Science* 311, 473–474.
- Pfund C, Spencer KC, Asquith P, House SM, Sorkness CA (2015). Building national capacity for research mentor training: an evidence-based approach to training the trainers. *CBE Life Sci Educ* 14, ar24.
- President's Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*, Washington, DC: U.S. Government Office of Science and Technology.
- Quality Education for Minorities Network (2010). *Spring 2010 workshop on the recruitment and retention of African American male students in science, technology, engineering and mathematics (STEM)*, Washington, DC.
- Rodenbusch SE, Hernandez PR, Simmons SL, Dolan EL (2016). Early engagement in course-based research increases graduation rates and completion of science, engineering, and mathematics degrees. *CBE Life Sci Educ* 15, ar20.
- Russell CB, Weaver GC (2011). A comparative study of traditional, inquiry-based, and research-based laboratory curricula: impacts on understanding of the nature of science. *Chem Educ Res Pract* 12, 57–67.
- Russell SH, Hancock MP, McCullough J (2007). Benefits of undergraduate research experiences. *Science* 316, 548–549.
- Sadler TD, Burgin S, McKinney L, Ponjuan L (2010). Learning science through research apprenticeships: a critical review of the literature. *J Res Sci Teach* 47, 235–256.
- Seymour E, Hunter AB, Laursen SL, Deantoni T (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. *Sci Educ* 88, 493–534.
- Shaffer CD, Alvarez C, Bailey C, Barnard D, Bhalla S, Chandrasekaran C, Shandrasekaran V, Chung H, Doreer DR, Du C, *et al.* (2010). The Genomics Education Partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE Life Sci Educ* 9, 55–69.
- Shapiro C, Moberg-Parker J, Toma S, Ayon C, Zimmerman H, Roth-Johnson EA, Hancock SP, Levis-Fitzgerald M, Sanders ER (2015). Comparing the impact of course-based and apprentice-based research experiences in a life science laboratory curriculum. *J Microbiol Biol Educ* 16, 186–197.
- Simmons S (2014). One institution's approach: how the University of Texas at Austin merges research and teaching through the Freshman Research Initiative. *FASEB J* 28 (suppl 93.1).
- SPSS (2015). *IBM SPSS Statistics 23*, Armonk, NY: IBM.
- Thiry H, Weston TJ, Laursen SL, Hunter AB (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE Life Sci Educ* 11, 260–272.
- Weaver GC, Wink DJ, Varma-Nelson P, Lytle FE, Morris R, Fornes WL, Russell C, Boone WJ (2007). Developing a new model to provide first and second-year undergraduates with chemistry research experience: early findings of the Center for Authentic Science Practice in Education (CASPIE). *Chem Educator* 11, 125–129.