Title: Collaborative vs. Apprenticed Undergraduate Research Experiences

Submission Type: Stage 1
Running Title: Collaborative vs. Apprenticed Undergraduate Research Experiences

Authors:
Christopher T. Goode
Department of Psychology
Georgia State University
PO Box 5010
Atlanta, GA, 30302-5010

Melissa K. Demetrikopoulos
Institute for Biomedical Philosophy
PO Box 1528
Dunedin, FL 34697

Shari L. Britner
Department of Teacher Education
Bradley University
Westlake Hall 233
Peoria, IL, 61625

Laura L. Carruth
Neuroscience Institute
Georgia State University
PO Box 5030
Atlanta, GA, 30302-5030

Brian A. Williams
Early Childhood and Elementary Education
Georgia State University
PO Box 3980
Atlanta, GA, 30302-3980

John L. Pecore
Department of Teacher Education and Educational Leadership
Building 85, Room 176
University of West Florida
Pensacola, FL 32514

Robert L. DeHaan (the late)
Division of Educational Studies
Emory University
1794 N. Decatur Road
Atlanta, GA 30322

Kyle J. Frantz
Neuroscience Institute
Georgia State University
PO Box 5030
Atlanta, GA, 30302-5030

Corresponding Author:
Christopher T. Goode
Department of Psychology
Georgia State University
PO Box 5010
Atlanta, GA, 30302-5010
cgoode@gsu.edu
Ph: 404-413-5638
Fax: 404-413-6207

## ABSTRACT

To recruit bright students from diverse backgrounds and retain them on a path toward biomedical research careers, we compared two curriculum formats for a 10-week summer undergraduate research program. Students were randomly assigned to either a collaborative learning model (CLM) in which they worked in teams of 3-4 in a dedicated classroom-based laboratory, or a traditional apprenticeship model (AM) in which they conducted mentored research in senior investigators' laboratories. We reported previously that this program enhanced internal dispositions that often predict retention in research career paths, such as Scientific Research Self-Efficacy, and Identity as a Scientist, with no significant differences in these outcomes between CLM and AM. Yet the routes to these positive outcomes may be different in the two program models.

As part of a large mixed-methods (qualitative and quantitative) approach to assessing program effects, we employed the Experience Sampling Method (ESM), which allowed us to collect data from students in real time throughout their research experiences. We explored the hypothesis that the CLM would promote research-related social interactions to the same or greater degree than the traditional AM, thereby increasing student engagement and Scientific Research Self-Efficacy. Our Experience Sampling Form (ESF) captured information about activities, social environments, and cognitive-emotional status. In support of our hypothesis, students in the CLM reported participating in social interactions more frequently and collaborating more than did those in the AM. They also reported feeling more collaborative. On the other hand, students specifically from underrepresented groups (URGs) in the AM interacted less and were alone more than those from well-represented groups in the AM, a demographic group difference not observed in the CLM.

Together with our previous results, the current evidence suggests the CLM can and should be implemented more broadly, particularly where research mentors and/or laboratories are scarce and in programs that aim to engage diverse populations in research. We recommend that scientists continue to explore such routes to positive outcomes for diverse undergraduate student populations, in order to optimize research experiences in various academic environments, and to promote efficient progress toward biomedical research careers.

# INTRODUCTION

## Diversity in science

Retention of bright students in pathways toward careers in science, technology, engineering, and mathematics (STEM) fields in the United States (U.S.) is declining, and lagging behind other nations (NCES, 2014). For example, 48% of bachelor's degree students and 69% of associate's degree students switched out of STEM majors or left STEM majors without a degree. Furthermore, rates of attrition from STEM majors are higher among women and African American students than other subpopulations (NCES, 2014); and the proportion of African American and Hispanic/Latina/-o individuals earning doctorates in STEM fields is less than half or one-third, respectively, of the proportion living in the U.S. population (DePass & Chubin, 2008; NSF, 2015). Thus, diversity among STEM professionals does not reflect the diversity of the general U.S. population. Women represent less than 25% of the population working at the highest levels of STEM organizations (NSF, 2015). These and similar subpopulations (e.g., students with disabilities) are thus known as under-represented groups (URGs) in STEM.

Probable causes of the challenges in STEM retention and diversity include poor academic performance in high school and early college, inadequate or uninspiring K-12 or college STEM teaching, and a lack of role-models from diverse demographic groups (NCES, 2014; Seymour & Hewitt, 1997). Probable effects of low retention and diversity include failure to maximize recruitment from the top talent pools, perpetuation of public health disparities, and a lag in discovery or scientific advance (NRC, 2007; Sullivan, 2004; Wenzel, 2004).

## Research experience for undergraduates

Proposed interventions to broaden participation in STEM careers emphasize hands-on, minds-on engagement of novice students in solving real-world STEM problems, perhaps best exemplified by recruiting young people into authentic research environments (Hofstein & Lunetta, 2004; NRC, 2003; 2007; Osborn & Karukstis, 2009). Quantitative and qualitative studies indicate clear benefits of research involvement by STEM students. As summarized in a large meta-analysis and extensive quantitative studies, positive outcomes following research experience include understanding of research and how to approach scientific problems; thinking like a scientist; and growth in confidence, independence, and responsibility (Lopatto, 2009; Lopatto et al., 2008; Seymour et al., 2004). Through a survey of thousands of STEM graduate students, Russel et al. (2007) also identified advanced research and communication skills, interest in STEM careers, and likelihood to take concrete steps to attain STEM careers, among students with undergraduate research experience. These outcomes are just as robust among students from URGs as those from well-represented groups (Lopatto, 2004), and undergraduates who engage in research are more likely to pursue STEM PhDs even after controlling for intended major, educational background, and parents' education (Carter et al., 2009).

A traditional approach to involving students in this beneficial STEM research is an apprenticeship of an individual student with a single academic or industry research professional. Yet this arrangement limits the number of research positions available, especially for students at institutions that are not research-intensive, which are primarily undergraduate institutions (PUIs), Historically Black Colleges and Universities (HBCUs), Hispanic-Serving Institutions (HSIs), community colleges, and other types of institutions with significant proportions of students from URGs, which thus disproportionately deprives them from research experience. An effective approach to creating research opportunities for more students at any institution is a Course-based Undergraduate Research Experience (CURE), which typically facilitates student-driven inquiry and discovery by enabling students to explore research questions in the context of a college science course (Cejda & Hensel, 2009; Perna et al., 2009; Lopatto, 2009). Like traditional apprenticeships, CUREs appear to increase understanding of science content and the nature of science, improve graduation rates, and promote persistence in STEM career paths (Beck et al., 2014; Harrison et al., 2011; Lopatto, 2009; Rodenbusch et al., 2016; Russel & Weaver, 2011; Shaffer et al., 2010; Thiry et al., 2012). Given that CUREs can be offered at institutions with low research intensity, they provide attractive opportunities to help reduce achievement and retention gaps between students from URGs and well-represented groups (WRGs; Bangera & Brownell, 2014).

We compared a variation on a CURE curriculum, which we called a Collaborative Learning Model (CLM), with a traditional Apprenticeship Model (AM) in the context of a neuroscience-based summer research program hosted across seven Atlanta-based colleges and universities (Frantz et al., 2017). For education research purposes, the most remarkable element of this program is that we used stratified random assignment to place students into the CLM or AM. This enabled us to compare outcomes across program models in an optimized research design.

**Mechanisms of growth during research experience**

The social and psychological mechanisms through which research helps students mature as scientists can be considered in the context of theoretical frameworks such as Social Cognitive Career Theory. Based on Bandura's (1977) general social cognitive theory, SCCT suggests that career-related self-efficacy is the key factor that predicts future career-related behavior, specifically whether or not one takes concrete steps toward career attainment (Betz & Hackett 1983; Chemers et al. 2011). Per Bandura (1977), self-efficacy arises from four major sources: mastery experiences, vicarious experience, verbal encouragement, and physiological states (e.g. related to mood/emotions). It is relatively easy to identify examples of each of these sources of Scientific Research Self-Efficacy (confidence in one's ability to carry out scientific research tasks) in the context of training a student in the laboratory environment. In a prior qualitative analysis of four participants in our program, we identified mastery experiences as especially important (Britner et al., 2012). In applying SCCT to a population of mostly URGs in various stages of STEM careers, Chemers and colleagues (2011) found that these types of experiences predict long-term persistence in STEM careers.

These and other mechanisms of growth toward a research career depend heavily on social interaction. In a classic line of research about social influence on STEM learning, Riesman and colleagues linked study groups to better grades, and turned the observation into an intervention that reduced achievement gaps, particularly between Asian and African American participants (Fullilove & Treisman, 1990; Treisman, 1992). Enriching academic and social support structures for students studying introductory chemistry or physics similarly eliminates achievement gaps (Hall et al., 2014; Watkins & Mazur, 2013). Subsequent research on collaborative learning has identified many benefits, such as increased engagement, effort, achievement, length of study time, knowledge retention, self-efficacy, and improved attitudes toward STEM study, with benefits extending across demographic subgroups (Bowen, 2000; Dirks & Cunningham, 2006; Fencl & Scheel, 2005; Fullilove & Fullilove, 1989; Halme et al., 2006; Johnson & Johnson, 1999; Peterson & Miller, 2004; Sadler et al., 2010; Slavin, 1995; Springer et al., 1999; Tien et al., 2002).

In the context of scientific research specifically, undergraduates identify clear mentor-mentee communication as another key social interaction in the research experience (Pfund et al., 2006), and national mentor training now emphasizes strategies for effective communication, aligned expectations, equity and inclusion in mentor-mentee relationships, and clear assessment of research understanding (Pfund et al., 2015). Particularly for trainees from URGs working with non-minority mentors, strategies such as "wise criticism" may be critical to retention (Cohen et al., 1999). Furthermore, for communities of practice such as scientific research teams, social interaction defines the concept, as people with shared interests and goals build relationships that enable them to learn and develop skills together (Gazley et al., 2014). Moreover, social constructivism in general, and collaborative learning in particular, underscore the important role that shared learning plays in acquisition and retention of knowledge (Smith & MacGregor, 1992; Vygotsky, 1980). Together, these influences of social interaction on STEM research recruitment and retention warrant further attention.

Our training program provides an opportunity to compare a traditional, mentored research apprenticeship with a potentially more social and more collaborative, yet authentic and mentored scientific research experience. Because the CLM and AM provide different social environments, we considered whether students in the two programs differed in their social interactions, and how they made gains over the course of the program. We also considered whether students from URGs and WRGs differed in their social interactions in the CLM and AM components of our program, given that collaborative learning has been shown to reduce the STEM achievement gap between these two populations.

Through a mix of quantitative and qualitative methods, we previously determined that the CLM confers similar benefits as the traditional AM (Britner et al., 2012; Frantz et al., 2006; 2017; Goode et al., 2012). Using pre-, mid-, and post-program electronic surveys, we measured internal dispositions reported to predict progress toward STEM careers, including Scientific Research Self-Efficacy, Leadership/Teamwork Self-Efficacy, Identity as a Scientist, Science Anxiety, Neuroscience Anxiety, and Commitment to a Science Career. We also incorporated post-program interviews and focus groups, and are still tracking long-term retention in science careers. Scientific Research Self-Efficacy and Identity as a Scientist increased over the course of the program. Similarly, Science and Neuroscience Anxiety (worries or other negative attitudes about performance and ability in science or neuroscience) decreased significantly, with no differences between CLM and AM. Further analysis of the internal dispositions revealed that Scientific Research Self-Efficacy predicted Commitment to a Science Career, and that this relationship was fully mediated by Identity as a Scientist (thinking of oneself as a scientist). Thus, experiences that promote self-confidence in research skills may solidify students' scientific identity, and retain them on pathways toward research careers. Importantly, all these short-term gains were seen in students from URGs as well as well-represented groups (WRGs). More recent surveys of program alumni show that students from each program model were equally likely to remain in research-related career paths four to seven years since their participation, with 68% of alumni remaining in research overall (Frantz et al., 2017). Although we detected no significant differences between program types in any of the short-term or long-term benefits to students, we wanted to determine whether students in the CLM and AM progressed toward these gains by the same means. Given the differences in social environments between a traditional research apprenticeship and our collaborative team-based CURE-like model, we suspected the process of growth in Scientific Research Self-Efficacy and Identity as a Scientist might differ between the two programs. To test this, we required a method to capture their social interactions and other experiences within their research environments.

**Experience sampling method**

Most of the social mechanisms for STEM career success summarized above were measured by periodic surveys, individual interviews, or focus group discussions at times or places removed from the actual research environment. In contrast, the Experience Sampling Method (ESM) is an approach that balances the ease of survey techniques with the moment-to-moment data collection provided by audio- or videotaping. Developed by Csikszentmihalyi and Larson (1987) and summarized by Hektner et al. (2007), ESM is a temporally random sampling procedure that prompts participants to fill out surveys immediately after notification on an individual device such as a beeper, watch, or smartphone. External coordinates of participant experience are usually queried (e.g. time, location, main and secondary activities, other people present), along with internal attributes of the experience (e.g., via a scale to rate cognitive-emotional status). Participation often lasts a week or two at a time, with surveys prompted numerous times per day, at times appropriate for the research questions (e.g., during the school day in education settings). Resultant data sets are incredibly rich, and they point to important influences of daily activities on current cognitive-emotional status as well as future activities and decisions. ESM in general appears also to reduce retrospective bias, providing a potentially more accurate description of the environment than instruments such as pre-post surveys (Kahneman & Krueger, 2006).

ESM is well suited for enriching our understanding of STEM education (Zirkel et al., 2014). It has been used to demonstrate that science classes more than other classes integrate intellectual challenge with skill application to solve a problem or complete a task, generating a cognitive state known as “flow” (Shernoff et al., 2003). As reviewed in Zirkel et al. (2014), high school science students also reported high “flow” when teacher discourse focused on higher order thinking and questioning. Whereas they reported less engagement and a reduced sense of relevance during lab lessons, compared with lecture, labs were nevertheless associated with more enjoyment and interest.

In terms of differential experiences across subpopulations, ESM has revealed that women and students of color are more likely than men and white students, respectively, to be concerned about appearing intelligent in class, an effect most robust in STEM classes, and unfortunately predictive of less frequently asking questions or seeking help during office hours. Moreover, losses in engagement, motivation, and confidence were tied to the experience of racially charged or gendered incidents among students of color in law school (London et al., 2007). ESM has also been used to reveal that women's self-efficacy decreases while receiving instruction in physics (Nissen & Shemwell, 2016). Anxiety may also emerge in contexts that promote stereotype threat (Steele, 1997), thereby adversely affecting performance. Although ESM has not yet been employed extensively in STEM education research, these existing forays provide insights into the social and cognitive-emotional environments in which students are exposed to and explore STEM concepts and processes.

## Current research questions

We used the ESM to compare student experiences in our CLM vs. AM curricula, while also comparing experiences in women vs. men, students from URGs vs. WRGs, and between experiences early in the program when guided instruction might dominate vs. near the end of the program when maximum independence might be exhibited. Specifically, we tested two initial research questions: (1)  whether participants in the CLM interacted with others more than students in the AM, and whether they felt more collaborative than participants in the AM, because the CLM was designed as a team-based, interactive, and explicitly collaborative route to research experience; and (2) whether participants reported feeling more engaged, collaborative, and confident when in the presence of a mentor, and whether they felt less anxious. To probe the disparate experiences in the two training models across demographic groups, we measured Engagement, Anxiety, and Confidence, along with Collaboration, as indicators of various cognitive-emotional states. In all cases, we examined whether responses varied systematically for women vs. men, or among students from URGs or WRGs.

Finally, given that we used a newly developed experience sampling form (ESF), we also carried out several tests of reliability and validity.

Our ESF requested the common external coordinates noted above, such as time of day, activities, presence of others, along with answer options tailored to our program (e.g. conducting specific research activities, presence of teammates or lab mates). These elements enabled us to describe the social environment, including frequency of interaction and which types of individuals were interacting. It also requested responses on a 16-item cognitive-emotional indicator chart intended to explore engagement, anxiety, collaboration, and confidence. Engagement is a critical part of flow, a traditional ESM element because of its documented support of knowledge acquisition. Collaboration was included as a direct test of the basic premise that CLM is more interactive than AM. Anxiety was measured because reduced or declining anxiety could be a source of self-efficacy, per Bandura's original conceptualization, and measuring in the ESF facilitated direct comparisons of the in-the-moment measures of anxiety with the pre-, mid-, and post-program electronic measures of similar dispositions. Similarly, confidence is a defining feature of self-efficacy and its measurement on the ESF facilitated direct comparison with the electronic surveys, as with the anxiety measures. We report here an initial exploration of the basic quantitative outcomes from the ESM, in comparison with prior data from electronic surveys at pre-, mid-, and post-program time points in this summer undergraduate research program.

## METHODS

### Participants

Thirty-nine undergraduate students were recruited from around the United States in Spring, 2012, to participate in our summer research experience in neuroscience. Our selection criteria favored members of demographic groups under-represented in STEM. Self-reported ethnicity was as follows: African Descent/African American (14), Asian Descent/Asian American (6), Caucasian (10), Hispanic/Latino/Latina (8), and Native American/Pacific Islander (1). Some may have had documented cognitive or social disabilities, or non-documented disabilities, but did not declare them on application to this program. Overall, at least 56% were from demographic groups under-represented in STEM. Almost 62% were women. Selection criteria also favored research novices; our sample included freshmen (5), sophomores (21), juniors (9), and seniors (4), for a ratio of approximately 67% first-years or sophomores and 33% juniors or seniors.

For stratified random assignment to program models, accepted students were categorized by race/ethnicity, sex, academic year, and GPA scores. Participants 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 would further balance the treatment groups on those additional variables. Students were then invited into their assigned program models and were not provided with an opportunity to switch to the other. In this manner, two balanced treatment groups of n=19 (CLM) and n=20 (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.

**Behavioral Research Advancements in Neuroscience, The BRAIN Program**

As described previously (Britner et al., 2012; Frantz et al., 2006; 2017; Goode et al., 2012) Behavioral Research Advancements in Neuroscience (BRAIN) was a ten-week, paid, intensive summer research program with participation contingent on a competitive application process. The overall goals of the BRAIN program were 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). Students admitted to the BRAIN program were randomly assigned to either the CLM or the AM. After a few days to move in and adjust to the local environment, the program curriculum began with one week of intensive classroom instruction in basic neuroscience shared by all participants, followed by nine weeks of neuroscience laboratory research in either the CLM or the AM group. The introductory classroom instruction addressed cellular and molecular neuroscience as well as behavioral neuroscience, using activities, lectures, and hands-on mini-experiments (approximately 9:00 a.m. – 5:00 p.m. daily for 5-6 days). During the subsequent nine weeks, all participants were expected to work 35 hours/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), as well as 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 $4,000, 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*). Eight instructors were deployed over eight weeks for the CLM (e.g., one faculty member, two post-docs, four graduate students, one senior undergraduate teaching assistant), 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 five weeks in the CLM, daily activities generally consisted of 1-2 hour introductions to new material (via lecture, demonstration, and discussion of assigned readings and 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 each week, their specific experimental questions were based on individual team interests. During the last three 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 at least once per week 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 17 different laboratories at four 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 known to provide strong research opportunities for undergraduates. Based on submission of weekly time sheets signed by mentors, participants fulfilled the expectation to conduct research activities at 35 hours/week. Daily schedules were designed individually by participants and mentors to fit for the diverse research paradigms, laboratories, and institutions that comprised the apprenticeship experiences. Research topics ranged from central pattern generators in *C. elegans*, through object recognition in *R. norvegicus*, to visual saccades in tool use and conditioned fear extinction for post-traumatic stress disorder in *H. sapiens*.

**Mentors**

A diverse group of faculty members and advanced trainees (e.g. postdoctoral fellows and graduate students) served as research mentors and instructors for both program models. The CLM included a series of faculty, postdoctoral fellow, and graduate student instructors with expertise in specific methods, each participating for one or two of the five weeks of guided instruction. These individuals were in the laboratory facility with participants for approximately eight hours per day, five days per week, except during professional development workshops. The instructors were complemented by senior faculty mentors (BRAIN program leaders) who met with the participants at least weekly to mentor and facilitate the development of the teams and their individual research hypotheses, methods, analyses, and reporting. A dedicated laboratory manager served as the daily mentor with the longest duration and highest frequency of interaction with the CLM participants (over 280 contact hours). The CLM instructional team members included five women, two self-identified African American individuals, one Native Hawaiian, and one Hispanic/Latina.

In the AM, a structured mentor-matching process preceded participant arrival to the program site. Recruited mentors provided short 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. Pre-program mentor-mentee 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 daily with graduate students or postdoctoral fellows. With both senior (principal investigator) and junior mentors combined, the AM mentor population included 16 faculty members, 1 post-doc, 5 graduate students, and 2 research technicians,. This population represented 17 women, 2 African American, 1 with a documented disability, four First-Generation College-Bound individuals, and one mentor who chose not to provide race, ethnicity, or family history of college education.

**Measures and Procedure**

Modeled after classic ESM forms (Hektner et al., 2007), as well as specific sample forms from other biomedical education scientists (Arora et al., 2011), the ESF in the present study balanced a broadly validated and typical format, with specifically targeted answer options and a novel cognitive-emotional domains chart for this study. Thus, this ESF was a paper-and-pencil form that requested external coordinates such as participant codes, date, time prompted, time answered, time finished, location at the time of the prompt, thoughts at the time, primary/secondary activities, and descriptors of the social environment, a chart with internal dimensions of experience via a chart on cognitive-emotional domains, and room for any other comments (see **Appendix A**). A provided list of possible activities focused on research tasks (e.g. planning, data collection, solving problems), and an open-ended question gave the opportunity to write in any other activities. Similarly, a provided list of possible people in the social environment focused on program affiliates (e.g., CLM teammates, AM lab mates, or mentors), with an open blank allowing specification of other individuals. Given the present interest in social interactions, a blank space was provided to describe the nature of any identified social interactions (e.g. talking/listening, showing/watching).

For the affective-cognitive domains, four areas of interest were queried with four scales each, for a total of 16 items, each constructed as a bipolar, semantic, differential inquiry. The domains were chosen to help validate and extend the exploration of similar dispositional queries in separate electronic surveys implemented at the beginning, middle, and end of the summer. The first four items measured Engagement in terms of activity, interest, engagement, and focus. The second four items measured Anxiety in terms of worry, stress, tension, and emotional control. The next four items measured Collaboration or other work-related interactions in terms of collaboration, brainstorming, communication, and cooperation. The final four measured Confidence in terms of confidence, self-assuredness, self-efficacy for research, and self-efficacy for learning science concepts.

During the program orientation week, participants were introduced to the ESM using a scripted orientation, filled out a sample ESF together, were provided reassurance that responses would remain confidential, were prompted at two random times to practice the ESF, were provided reassurance that responses would remain confidential, and had the opportunity to ask questions about the procedure. Prompts were delivered via text messages to cellular phones. The program purchased phones for three participants who did not own their own cellular phones, and paid for text messages for participants without unlimited messaging plans. For data collection in the interval sampling procedure, participants were prompted four times per day, six days per week, during two separate weeks in the summer program (early week 2, late week 6), for a maximum of completed ESFs of 48 per participant. Sampling intervals were divided into business hours (10:00 am to 6:00 am) and evening hours (6:00 pm to 10:00 pm), with one prompt in each of three sampling intervals during business hours (10:00 am to 12:40 pm, 12:40 to 3:20 pm, 3:20 to

6:00 pm), and only one prompt in the evening. Because ESFs for the week were provided and collected at the weekly professional development workshops, only one prompt occurred on the first day (evening interval after the workshop) and only three prompts occurred on the last day during the business-hours interval (before or during the workshop).

In separate retrospective electronic surveys implemented at pre-, mid-, and post-program time points, internal dispositions typically predictive of retention in research careers were measured. These included Scientific Research Self-Efficacy, Leadership/Teamwork Self-Efficacy, Identity as a Scientist, Science Anxiety, Neuroscience Anxiety, and Commitment to a Science Career. Participants were also asked via online surveys adapted from Chemers et al. (2011;  see Britner et al., 2012 and Goode et al., 2012 for complete descriptions of the electronic surveys of internal dispositions), to rate their agreement with 7-10 statements indicative of the six dispositions listed above on 5-poing Likert scales. Summed ratings of statements were used as measures of each disposition. Surveys were administered before the program began, at the midpoint, and again at the end of the ten-week program. The pre-program electronic survey time point corresponds closest to the early ESF prompts, and the mid-program electronic surveys were completed closest in time to the late ESF prompts. 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 four to seven years after the end of the summer program participation. Details of the data analysis, and associated figures, can be found in **Appendix B**.

## RESULTS

Results from the use of the ESM form the core of the experiment comparing two types of summer undergraduate research experience: a collaborative, team-based, workshop-like model with elements of a course-based undergraduate research experience (CLM), and a traditional research apprenticeship (AM). Not only were we able to validate new measures of undergraduate student subjective states as cognitive-emotional status indicators “in-the-moment,” but we also were able to isolate and quantify specific social situations potentially important to students joining the scientific community. By filtering responses that indicated participants were actively engaged in research, interacting with others, and/or in the presence of senior or junior mentors, we compared program models, social influence on cognitive-emotional status, and sex and race/ethnicity differences in specific program experiences. Generally, we confirmed that the CLM evokes more collaboration, more interaction, and less “alone time” than the AM, and in some circumstances, more engagement (A, Figure 4). Notably, statistical interaction effects suggest that the CLM eliminates some sex and racial/ethnic disparities we found in the AM.

A previous study of this same program confirmed there were no significant differences between program models in short-term gains in dispositions predictive of persistence in STEM careers, including Scientific Research Self-Efficacy, Leadership/Teamwork Self-Efficacy, Identity as a Scientist, Science and Neuroscience Anxiety and Commitment to a Science Career (Frantz et al., 2017). Moreover, at four to seven years since program participation, approximately 68% of known program alumni remain in research career tracks, with no differences between CLM and AM. Given the different research settings of these programs, however, we considered whether students in each program type progressed toward the same gains via different pathways. The ESM responses collected and analyzed herein therefore provide a detailed view of the social, cognitive, and emotional status among participants in summer research experience already known to be transformative and effective.

We began our analysis by validating our new instrument. The present data suggest that the ESF scale items intended to measure the cognitive-emotional domains of Engagement, Anxiety, Collaboration and Confidence are valid and reliable. The relatively lower within-subjects variation compared with between-subjects variation suggests that our participants, like other ESM participants, express “person-level traits” that maintain some stability over time, but certainly vary during specific activities (Hektner et al., 2007). Thus, these measures have sufficient internal consistency to be used to test for between-groups differences, e.g. across program method, sex, representation, etc.

Our factor analysis and correlations among the cognitive-emotional status indicators confirm strong discriminant validity. The factor analysis extracted four factors, and each item within a specific measure, such as Engagement, loaded strongly onto the same factor, and not significantly onto any other factor. Correlations among the four factors support discriminant validity further, with Anxiety correlating negatively, and Collaboration correlating positively with Confidence for students in both the CLM and AM programs. Engagement also correlated positively with Collaboration and Confidence, and negatively with Anxiety. These relationships support our use of this instrument to describe student experiences.

With respect to internal validity, our Cronbach’s alpha values were extremely high for some items, suggesting that some scales could be improved by deleting redundant items. Although our response rates were good, this may improve future ESF collection with similar populations.

In terms of cross-validation with internal dispositions measured periodically in the same participants taking electronic surveys at three time points (pre-, mid-, and post-program), correlations among our cognitive-emotional status indicators and survey measures demonstrate some convergent and discriminant validity of our ESF measures (Figure 1). For example, Confidence on the ESF correlates positively with both Scientific Research and Leadership-Teamwork Self-Efficacy, as well as Commitment to Science. Collaboration correlates positively with Leadership-Teamwork Self-Efficacy. Yet the correlation between Confidence on the ESF and Scientific Research Self-Efficacy on the retrospective, electronic surveys was not particularly robust, pointing out slight variations in the domains queried, with ESF Confidence perhaps is broader than Scientific Research Self-Efficacy.

We also compared temporal aspects of individual development in the electronic surveys and the ESFs. For instance, CLM participants reported significantly higher Scientific Research Self-Efficacy than AM participants already by the mid-program electronic survey (Frantz et al., 2017), but this difference was not observed in the related ESF measures of overall Confidence taken around the same time in the program. Other changes over time are apparent in the correlations among our ESF and internal disposition measures. For example, the correlation between early Engagement and Anxiety is not strong in the first ESF set (B, Figure 1), but is strong and negative by the latter ESF set (C, Figure 1). Similarly, the relationships between Anxiety on the ESF and Science and Neuroscience Anxiety are weak at the early time point, and strong and positive later, when considering the AM participants alone. For the CLM participants, the relationship between Anxiety on the ESF and Science and Neuroscience Anxiety was negative early, and weakly positive later. Altogether these findings suggest that participants' retrospective ratings of their confidence and anxiety were sometimes different than their reflections in the moment (Zirkel et al., 2015).

## CONCLUDING THOUGHTS

One of the motivations for employing ESM was to investigate the social component of the research experience during which internal dispositions such as self-efficacy and science identity are known to develop. Thus, one of the most basic and important validations to emerge from the present ESM study is confirmation that the CLM we designed indeed produced more collaboration. CLM students reported being more Collaborative than AM students (A, Figure 3). While the present CLM produced subjective ratings that were more collaborative as intended, it is important to note that not all research experiences need to look this way. Not only are both collaboration and independence important factors in student progression in research careers (Lopatto, 2009), but also self-ratings of independence predict retention in science careers (McGee & Keller 2007). We do not consider collaboration to be more important than independence. In our pilot study (Frantz et al., 2006), we did record some participant views that the CLM was perceived as better for early career, novice scientists, whereas the AM might be better later in the college career after scientific research self-efficacy was well-developed.

When we conducted deeper analysis of social contact, statistical interactions among program method, sex, and representation generally supported the conclusion that the CLM eliminates subpopulation disparities we observed in the AM. This general pattern is seen for time spent interacting and alone, as well as for cognitive-emotional status indicators in the presence of a junior mentor. When engaged in research, students from WRG reported interacting more than URG, but only in the AM (A, Figure 2). Similarly, WRG women reported being alone in more of their research responses than URG women and WRG men, but only in the AM (B, Figure 2).

Students from WRG reported more Engagement and Collaboration (A, B, Figure 4) when in the presence of their junior mentors, but again this effect is seen only in the AM, not in the CLM. Although we temper the present results by acknowledging the low sample size (ranging from one to nine reports within a subgroup, e.g. only one report in the presence of a junior mentor from under-represented men in the AM), such benefits of junior mentors may relate to a peer leader or a near-peer role played by the junior mentors, welcoming the present participants into the laboratory workforce and the scientific community (e.g. Tinto, 1993; Nadga et al., 1998; Fullilove & Treisman, 1990). Future qualitative analysis of the present ESM data may provide information about the nature of discourse between participants and mentors (Cohen et al., 1999), perhaps even revealing roles in the relationship (e.g. learner, leader, facilitator; Stamovlasis et al., 2006).

Beyond the present study, it would also be interesting to determine whether measures of Engagement and Collaboration are related to senses of belonging, motivation, and security, which predict feelings of personal compatibility with STEM majors and perceived social support, at least among women (London et al., 2011). Given that good mentoring is key in training new scientists (Hunter et al., 2007, Nagda et al., 1998), intentional facilitation of effective mentor-mentee communication is likely to enhance research experiences for all population subgroups. Moreover, students who receive more support from faculty are more likely to report plans for a graduate STEM degree (Eagan et al., 2013).

To explore mechanisms through which students mature in internal dispositions predictive of research careers, we again integrated analyses of the ESM data with electronic survey data collected from the same participants (Frantz et al., 2017). For example, we reported earlier that although post-program Scientific Research Self-Efficacy predicts post-program Commitment to Science, the effect is fully mediated by Science Identity (Frantz et al., in press). We report a similar finding here: late (week 6) ESF reports of Confidence while engaging in research activities significantly predict post-program Commitment to Science on the electronic surveys, after controlling for pre-program Commitment (A, Table 2). Yet this effect disappears when Science Identity is added to the first block of the model, suggesting a mediation effect of Science Identity (B, Table 2). This relationship has been well described by Chemers et al. (2011), and is also evident in the work of Estrada et al. (2011), who found that self-efficacy is related to science identity and is predictive of commitment to science careers.

As for limitations, although 1,511 surveys were collected and analyzed,(an 81% response rate), this still addresses only 39 participants and very few ESFs collected under certain conditions of interest, e.g., from specific demographic subgroups in the presence of a mentor, as noted above. Another drawback of the ESM in general is the invasiveness and time required to respond to frequent surveys (London et al. 2011). Four participants mentioned in their surveys that they were so frustrated with the frequency of the interruptions that we ought not to trust their responses as reliable indicators of their cognitive-emotional status. Finally, the differential response rates between CLM and AM students may have contributed to the program differences we reported in Collaboration, and Engagement in the presence of a junior mentor.

Looking to the future, studies increasing the number of students surveyed would allow for nested statistical analysis, and consideration of more individual responses instead of relying on averages for certain circumstances, e.g. in the presence of mentors, etc. Future analysis can also determine whether responses from the reportedly disgruntled respondents were outliers. Future modifications to the method, such as electronic ESFs rather than the paper/pencil strategy employed here, are also likely to minimize intrusions while maintaining rich data collection. Combining the ESF with wearable physiological monitors would enable correlation between subjective ratings of anxiety with physiological measures of stress.

Several analyses could emerge even within the present data set. A qualitative analysis of participant text entries would enable us to provide more detail regarding the nature of the interactions described by participants, and correlate them with the cognitive-emotional domains in the same moment. The potential for rich, qualitative indicators of the mechanisms through which these research experiences transform student career preparation is vast and largely untapped. Finally, an extensive alumni survey under way at the time of submission of this report will enable us to link ESM results with longer-term outcomes, including decisions to stay in STEM majors, seek additional research opportunities, or pursue graduate degrees.

Although the goal of this study was to investigate student experiences in a single summer research program, preliminary broader implications for program directors and research mentors should be considered. Most importantly, the CLM is an effective approach to offering undergraduates a real research experience. It is at least as good as the AM at raising Scientific Research Self-Efficacy, reducing Science Anxiety, and raising Science Identity in the short-term. At four to seven years since the experience, the same proportion of participants remain in research career paths. Faculty at various types of institutions should choose which approach to research experience for undergraduates best fits their infrastructure. Furthermore, the CLM provides more socialization in research environments, suggesting that it is an effective route to enculturation of novice scientists into the scientific community (Gazley et al., 2014). Finally, our ESM data suggest that the AM provides a more variable social experience for undergraduates, thereby perhaps requiring more attention to mentor preparation and both formal and informal social interactions than is often provided, especially when efforts to retain students from underrepresented groups are emphasized.

In conclusion, this study capitalizes on a true experimental design with stratified random assignment of participants to research conditions. In combination with the ESM, our approach provides an especially powerful exploration of student experiences in undergraduate research. We show that a novel, collaborative model for undergraduate research training can eliminate, and even reverse in some cases, disparities among racial, ethnic, and gender subpopulations in collaboration, engagement, and interaction with others. Ultimately, in-depth and careful STEM education research using tools such as ESM will help to diversify and enhance the biomedical research workforce.

## ACKNOWLEDGMENTS

The authors would like to thank Elizabeth Weaver, Laurie Murrah-Hanson, and Emily Hardy for their BRAIN Program Coordination, as well as Rob Poh and Tom McKlin for their input to data management and survey design, and Jacob Shreckengost for help with the manuscript. Tiffany Oliver and Michael Black were excellent CLM laboratory mentors. Numerous graduate students and postdoctoral research associates were excellent CLM instructors. Numerous faculty members, graduate students and postdoctoral research associates were excellent AM mentors. The late Robert L. DeHaan, co-author, was an outstanding mentor and collaborator for all involved in this program and research study, participants and co-authors alike. This program and the research on it were funded by NIH (1R25GM097636; 1RO1GM085391), NSF (IBN-9876754), Georgia State University, Emory University, Georgia Institute of Technology, Clark Atlanta University, Morehouse College, and Spelman College.

## REFERENCES

Arora V, Schneider B, Thal R, Meltzer R (2011). Design of an intervention to promote entry of minority youth into clinical research careers by aligning ambition: The TEACH (Training Early Achievers for Careers in Health) research program. J Soc Issues 67, 580-598.

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.

Bowen C (2000). A quantitative literature review of cooperative learning effects on high school and college chemistry achievement. Chem Educ Res 77(1), 116-119.

Britner SL, Williams BA, Pecore JL, Gagne P, Demetrikopoulos M K, 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 Min Sci Eng, 18(3), 273-293.

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 An 31, 441-462.

Cejda BD, Hensel N (2009). Undergraduate Research at Community Colleges, Washington, DC: Council on Undergraduate Research.

Chemers MM, Zurbriggen EL, Syed M, Goza BK, Bearman S (2011). The role of efficacy and identity in science career commitment among under-represented minority students. J Soc Issues 67, 469-491.

Cohen A, Steele CM, Ross LD (1999). The mentor's dilemma: providing critical feedback across the racial divide. J Pers Soc Psych 25(10), 1302-1318.

Csikszentmihalyi, M, Larson R (1987). Validity and reliability of the experience-sampling method. J Nerv Ment Dis 175(9), 526-536.

DePass AL, Chubin DE (Eds.) (2008). Understanding Interventions that Encourage Minorities to Pursue Research Careers, Bethesda, MD: American Society for Cell Biology. Retrieved from: http://understanding-interventions.org/wp-content/uploads/2015/02/08Understanding_Interventions-1.pdf

Dirks C, Cunningham M (2006). Enhancing diversity in science: is teaching science process skills the answer. Cell Biol Educ 5, 218-226.

Eagan MK, Hurtado S, Chang MJ, Garcia GA, Herrera FA, Garibay JC (2013). Making a difference in science education: the impact of undergraduate research programs. Am Ed Res Jour, 50(4), 683-713.

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.

Fencl H, Scheel K (2005). Engaging Students: An Examination of the Effects of Teaching Strategies on Self-Efficacy and Course Climate in a Nonmajors Physics Course. J Coll Sci Teach 35(1), 20-26.

Frantz KJ, DeHaan RL, Demetrikopoulos MK, Carruth LL (2006). Routes to research for novice undergraduate neuroscientists. CBE-Life Sci Educ 5, 175-187.

Frantz KJ, Demetrikopoulos MK, Britner SL, Carruth LL, Williams BA, Pecore JL, DeHaan RL, Goode CT (2017). A comparison of internal dispositions and career trajectories after collaborative vs. apprenticed research experiences for undergraduates. CBE Life Sci Educ 16(1), ii.

Fullilove MT, Fullilove RE (1989). UCSF Medical Scholars Program Quantitative Analysis, 1986-1987; San Francisco: University of California. p 54.

Fullilove RE, Treisman PU (1990). Mathematics achievement among african american undergraduates at the University of California, Berkeley: an evaluation of the mathematics workshop program. J Negro Educ 59(3), 463-478.

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. Hauppage, NY: Nova.

Hall DM, Curtin-Soydan AJ, Canelas DA (2014). The science advancement through group engagement program: Leveling the playing field and increasing retention in science. J. Chem. Educ. 91, 37–47.

Halme EG, Khodor J, Mitchell R, Walker GC (2006). A small-scale concept-based laboratory component: the best of both worlds. Cell Biol Educ 5, 41-51.

Harrison M, Dunbar D, Ratmansky 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.

Hektner, JM, Schmidt JA, Csikszentmihalyi M (2007). Experience Sampling Method: Measuring the Quality of Everyday Life. Sage Publications, Thousand Oaks, CA.

Hofstein A, Lunetta VN (2004). The laboratory in science education: foundations for the twenty-first century. Sci Educ 88, 28-54.

Hunter AB, Laursen SL, Seymour E (2007). Becoming a scientist: the role of undergraduate research in students' cognitive, personal, and professional development. Sci Ed, 91, 36-74.

Johnson DW, Johnson RT (1999). Making cooperative learning work. Theory Into Practice 38(2), 67.

Kahneman D, Krueger AB (2006). Developments in the measurement of subjective well-being. J Econ Persp 20(1), 3-24.

Lawson AE, Banks DL, Logvin M (2007). Self-efficacy, reasoning ability, and achievement in college biology. J Res Sci Teach 44, 706-724.

London, B, Rosenthal, L Gonzalez, A (2011). Assessing the role of gender rejection sensitivity, identity, and support on the academic engagement of women in nontraditional fields using experience sampling methods. J. Soc. Issues 67, 510–530.

London B, Rosenthal L, Levy SR, Lobel M (2011). Influences of perceived identity compatibility and social support on women in nontraditional fields during the college transition. Applied Social Psychology, 33, 303-321.

London B, Downey G, Mace S (2007). Psychological theories of educational engagement: A multi-method approach to studying individual engagement and institutional change. Vandebuilt Law Rev. 60, 455–481.

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. Research Corporation for Science Advancement, Tucson, AZ.

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 Sciences Education 6, 316-331.

Nagda BA, Gregerman SR, Jonides J, von Hippel W, Lerner JS (1998). Undergraduate student-faculty research partnerships affect student retention. The Review of Higher Education 22(1), 55-72.

National Science Foundation (2015). Women, Minorities, and Persons with Disabilities in Science and Engineering. http://www.nsf.gov/statistics/2015/nsf15311/start.cfm (accessed 17 May, 2016).

NCES (2014). STEM Attrition: college students' paths into and out of STEM fields. National Center for Education Statistics, U.S. Department of Education.

Nissen JM, Shemwell JT (2016). Gender, experience, and self-efficacy in introductory physics. Phys. Rev. Phys.

Educ. Res. 12, 20105.

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.

Osborn, JM and Karukstis, KK (2009). The benefits of undergraduate research, scholarship, and creative activity. In: Boyd, M and Wesemann, J (Eds.), Pp 41-53, Broadening Participation in Undergraduate Research: Fostering Excellence and Enhancing the Impact. Council on Undergraduate Research, Washington, DC.

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.

Peterson SE, Miller JA (2004). Comparing the Quality of Students' Experiences During Cooperative Learning and Large-Group Instruction. J. Educ. Res. 97, 123–134.

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 Sciences Education 14, 1-12.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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, 1-10.

Rosenthal L, London B, Levy SR, Lobel M (2011). The Roles of Perceived Identity Compatibility and Social Support for Women in a Single-Sex STEM Program at a Co-educational University. Sex Roles 65, 725–736.

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, Hewitt L (1997). Talking About Leaving: Why Undergraduates Leave the Sciences. Boulder, CO: Westview Press.

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, Shandrasekeran 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.

Shernoff D, Csikszentmihalyi M, Schneider B, Shernoff E (2003). Student engagement in high school classrooms from the perspective of flow theory. School Psychology Quarterly 18, 158–176.

Slavin R (1995). Cooperative Learning: Theory, research and practice. Boston: Allyn & Bacon.

Smith BL, MacGregor JT (1992). What is collaborative learning? National Center on Postsecondary Teaching, Learning, and Assessment, Penn State University.

SPSS, Inc. (2015). IBM SPSS Statistics 23, Armonk, NY: IBM Corp.

Springer L, Stanne ME, Donovan SS (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. Rev Educ Res 69(1), 21-51.

Stamovlasis D, Dimos A, Tsaparlis G (2006). A study of group interaction processes in learning lower secondary physics. J Res Sci Teach 43(6), 556-576.

Steele CM (1997). A threat in the air: How stereotypes shape intellectual identity and performance. Am Psychol, 52(6), 613-629.

Sullivan LW (2004). Missing persons: minorities in the health professions, a report of the Sullivan commission on diversity in the healthcare workforce. Sullivan Commission.

Tabachnick BG, Fidell LS (2013). Using multivariate statistics. (6th Ed.) Pearson Education, New York.

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.

Tien LT, Roth V, Kampmeier JA (2002). Implementation of peer-led team learning instructional approach in an undergraduate organic chemistry course. J Res Sci Teach 39(7), 606-632.

Tinto V (1993). Leaving college: Rethinking the causes and cures of student attrition (2d ed.) Chicago, IL: University of Chicago Press.

Treisman PU (1992). Studying students studying calculus: a look at the lives of minority mathematics students in college. College Math J 23(5), 362-372.

Vygotsky LS (1980). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Watkins J, Mazur E (2013). Retaining students in science, technology, engineering, and mathematics (STEM) majors. J Col Sci Teach 42(5), 36-41.

Wenzel TJ (2004). Meeting Report: Outcomes from the Undergraduate Research Summit. CBE-Life Sciences Education, 3, 150-151.

Zirkel S, Garcia JA, Murphy MC (2015). Experience-sampling research methods and their potential for education research. Educ. Res. 20, 1–10.

# APPENDIX A

**Experience Sampling Form.** When prompted, participants were instructed to fill out the form completely, indicating the time they were prompted as well as the time the form was completed.

Appendix A

| MY CODE: | | | | | |
|---|---|---|---|---|---|
| DATE: | | TIME BEEPED: | □ AM □ PM | TIME ANSWERED: | □ AM □ PM |

*As you were beeped...*
**Where** were you? ____________________
PLEASE BE SPECIFIC

What was **on your mind**? ____________________

What were you were **doing**? (Choose all that apply from the list below and/or write in more specific activities.)

| ACTIVITY | √ | ACTIVITY | √ |
|---|---|---|---|
| Research planning | | Consulting with someone on research | |
| Data collection | | Research team meeting | |
| Data analysis | | BRAIN workshop/seminar | |
| Reading for research | | Making discoveries | |
| Writing for research | | Solving problems | |

Other items not listed above? ____________________
PLEASE BE SPECIFIC

What **else** were you doing? ____________________
PLEASE BE SPECIFIC

Were you **alone**? □ **yes** □ **no**

Were you **interacting with anyone**? □ **yes** □ **no**

If you were interacting with anyone, who was it? (Indicate how many men and women, or boys and girls, fall into each category. Do not include yourself.)

| PEOPLE | # Men | # Women |
|---|---|---|
| Alone | 0 | 0 |
| BRAIN Senior Mentor | | |
| BRAIN Daily Mentor(s) | | |
| BRAIN Teammate or Labmate | | |
| Other BRAIN Fellows | | |
| Friends outside BRAIN | | |
| Family Members | | |
| Other. Please specify on lines below: | | |
| | | |

If you were interacting with someone as you were beeped, please write a sentence that **describes *how* you were interacting** (e.g. what were you doing and what was your role in the interaction - talking/listening, showing/watching, teaching/learning, everyone reading and discussing together, anything else…)?

____________________

Indicate your **status** as you were beeped. (For every pair of 'opposites,' put an 'x' over just one mark on the table below.)

| | Very | Quite | Some | Neither | Some | Quite | Very | |
|---|---|---|---|---|---|---|---|---|
| Active | □ | □ | □ | · | □ | □ | □ | Passive |
| Interested | □ | □ | □ | · | □ | □ | □ | Bored |
| Engaged | □ | □ | □ | · | □ | □ | □ | Disengaged |
| Focused | □ | □ | □ | · | □ | □ | □ | Distracted |
| Worried | □ | □ | □ | · | □ | □ | □ | Unworried |
| Stressed Out | □ | □ | □ | · | □ | □ | □ | Calm |
| Tense | □ | □ | □ | · | □ | □ | □ | Relaxed |
| Frenzied | □ | □ | □ | · | □ | □ | □ | Composed |
| Working collaboratively | □ | □ | □ | · | □ | □ | □ | Working independently |
| Brainstorming with others | □ | □ | □ | · | □ | □ | □ | Limited to own ideas |
| Communicating | □ | □ | □ | · | □ | □ | □ | Not communicating |
| Cooperative | □ | □ | □ | · | □ | □ | □ | Competitive |
| Confident | □ | □ | □ | · | □ | □ | □ | Insecure |
| Sure I can succeed in research | □ | □ | □ | · | □ | □ | □ | Unsure I can succeed in research |
| Good at research | □ | □ | □ | · | □ | □ | □ | Not good at research |
| Good at learning science concepts | □ | □ | □ | · | □ | □ | □ | Not good at learning science concepts |

**Any other comments?** ____________________

**Time Finished:** __________

# APPENDIX B

## Data Analysis Notes

Data Management: ESF responses were entered into a database by a single individual blind to treatment conditions and program goals. Within the database, responses were linked by participant code to the program application and electronic survey data for comprehensive analysis. Responses to each cognitive-emotional item (Appendix A) were coded into numerical scores ranging from -3 to +3 for analysis. On the basis of strong internal reliability and validity scores (see below), responses on the four scales in each cognitive-emotional domain were summed within an individual ESF to create a single composite score for engagement, anxiety, collaboration, and confidence. Anxiety and Collaboration score means are presented with inverse sign, such that a higher number means less anxiety and more collaboration to align with the other positive indicators for Engagement and Confidence. Responses to open-ended questions were entered verbatim into the database. All statistical analyses were conducted with SPSS v.23 (SPSS, Inc., 2015) supplemented with R (R Core Team, 2016). Unless otherwise indicated, alpha was set at .05 for the purposes of null hypothesis significance testing.

Internal Reliability, Validity, and Correlations in Survey Instruments: To analyze internal reliability of the newly developed cognitive-emotional instrument, Cronbach's alpha was calculated, with all participants and all ESFs except the two training prompts included, and data collapsed across participant subgroups. As an additional test of reliability, we analyzed whether variability in cognitive-emotional status was greater between participants than within participants. The average standard deviations of each participant's composite scores for Engagement, Anxiety, Collaboration, and Confidence, were compared with the standard deviation of all participants' composite scores.

Despite a relatively limited number of participants, we used factor analysis of all items on the cognitive-emotional scale to explore validity of the elements within this cognitive-emotional instrument, again with all participants and all test ESFs contributing. Because we knew the measures to correlate with one another to some extent, we used oblique rotation as per Tabachnick and Fidell (2013).

To further evaluate convergent validity among different instruments within our broader study, we tested correlations between the cognitive-emotional status indicators on the ESFs during the first week of ESM data collection and the internal dispositions recorded on the pre-program electronic survey about ten days prior. We also computed correlations between ESM status indicators during the second week of ESM and internal dispositions recorded on the mid-program electronic survey about ten days prior.

Program, Sex, and Representation Effects: We used between-subjects factorial ANOVAs in SPSS to examine the effects of program type (CLM or AM), self-identified sex, and self-identified racial and/or ethnic group on the four cognitive-emotional composite scores listed above. In testing for effects of race/ethnicity, we dummy coded the racial and/or ethnic group variable to indicate whether the self-identified race/ethnicity was under-represented or well-represented in STEM. Where main effects or interactions were significant, post-hoc, paired comparisons of the estimated marginal means were made using the Sidak adjustment for multiple comparisons.

Similarly, we tested for program type, sex, and/or representation differences in the social environment in which research was conducted. Thus, we calculated the proportion of prompts when any of the research activity options was checked and the "yes" option was checked in answer to the question "Were you alone?" On the other hand, we also calculated the proportion of prompts when any of the research activity options was checked and the "yes" option was checked in answer to the question "Were you interacting with anyone?" On these subsets of ESFs, cognitive-emotional indicators were also analyzed as composite scores for each of the domains. The proportions and composite scores were compared across program type, sex, and representation subgroups, and analyzed separately for each ESM week.

Finally, we further explored the specific impact of junior or senior mentors on cognitive-emotional status by selecting only those ESFs that indicated the presence of one or the other type of mentors and calculating composite scores for each of the cognitive-emotional domains. Scores were compared across program type, sex, and representation subgroups, and analyzed separately for each ESM week.

Response Rates

Of the 1,872 total ESFs requested over all sampling times (four times per day, six days per week, for two one-week intervals from thirty-nine participants), we collected 1,511, for an 80.72% overall response rate. More CLM response (86%) than AM response (76%) was observed ($X^2$=6.33, *p*=.01), with non-significant differences between women and men (82%, 79%) as well as between students from URGs and WRGs (83%, 77%). Compared with typical ESM studies, especially among undergraduate students, these response rates are excellent.

**Validation of Survey Instruments**

<u>Intra-Individual Reliability:</u> For each of the four cognitive-emotional composite scores (Table 1), average variability within individuals was lower than that between individuals. Averaged standard deviations within participants for Engagement (*SE*=5.13), Anxiety (*SE*=4.68), Collaboration (*SE*=5.01), and Confidence (*SE*=3.06) were lower than standard deviations between participants for the same measures (*SE*=5.60; *SE*=5.91; *SE*=5.58; and *SE*=4.71).

<u>Internal Reliability:</u> Coefficient alpha values were high for all four cognitive-emotional status indicators (Engagement, .925; Anxiety, .966; Collaboration, .878; and Confidence, .971). For all four measures, high alpha values and high r values on the inter-item correlation matrix indicated strong inter-correlation for some items within each scale, and suggests some redundancy of indicator items.

<u>Discriminant and Convergent Validity</u>: Factor analysis indicated adequate sampling (*KMO*=.761), Bartlett's test of sphericity was significant, $X^2(120)$=807.94, *p*<.001, and communalities ranged from .549 to .980. Four factors were extracted, with Engagement, Anxiety, Collaboration, and Confidence items loading separately and strongly onto each factor. Loadings for each item were mostly high and ranged from .598 to .978.

Correlation matrices plotting the relationships among cognitive-emotional status indicators averaged across all responses and our electronic survey measures of internal dispositions averaged across pre-program and mid-program time points (A, Figure 1) showed that ESF measures correlated positively with similar measures of internal dispositions, e.g. Confidence correlated positively with both Leadership/Teamwork (LTSE) and Scientific Research Self Efficacy (SRSE). Similarly, Anxiety and Science Anxiety correlated positively. Within the cognitive-emotional status indicators, Engagement correlated positively with Collaboration and Confidence, and Anxiety correlated negatively with Confidence. Separate correlation matrices for ESF responses between second week ESF measures of cognitive-emotional status indicators and pre-program electronic surveys of internal dispositions, and between sixth-week ESF and mid-program electronic surveys, show a similar trend, but generally weaker relationships earlier (B, Figure 1) than later in the program (C, Figure 1).

Social Interactions during Research Activities: A three-way ANOVA revealed a significant program model by representation interaction for the proportion of all responses indicating research activity in which participants reported they were also interacting with others $F(1,31)=5.025$, $p=.032$, partial$\eta^2=.139$. Students from URG and WRG reported a similar frequency of interactions while doing research in the CLM, but those form WRG reported more interactions ($M$=58.88%, $SE$=13.94) than those from URG ($M$=41.65%, $SE$=8.16; $p=.022$) in the AM (A, Figure 2).

Similarly, another three-way ANOVA revealed a main effect of program model on the proportion of all responses indicating research activity in which participants also reported being alone, $F(1,31)=11.257$, $p=.002$, partial$\eta^2=.266$, with participants in CLM indicating they were alone a lower proportion of the time (22%) than those in the AM (54%) . This effect must be considered in light of a significant three-way interaction among program model, sex, and representation, $F(1,31)=9.412$, $p=.004$, partial $\eta^2=.233$, with a sex x representation interaction evident in the AM, but not in the CLM. In the AM only, WRG women reported significantly more time alone while doing research ($M$=82.48%, $SE$=46.10) than URG women ($M$=29.22%, $SE$=7.66, $p=.007$; B, Figure 2). Well-represented women in the AM spent more time alone while doing research than well-represented men in the AM as well ($M$=32.14%, $SE$=5.92, $p=.014$; B, Figure 2).

General Cognitive-Emotional Composite Scores: ANOVAs were used to test for main effects and interactions of program, sex, and representation on cognitive-emotional status indicators. There was a significant effect of program on Collaboration, $F(1, 31)=9.905$, $p=.004$, partial $\eta^2=.242$ with CLM reporting more collaboration on average ($M=1.952$, $SE=0.446$) than AM ($M=-0.484$, $SE=0.633$; A, Figure 3). There was also a significant interaction between program and representation, $F(1,31)=17.935$, $p=.034$, partial $\eta^2=.138$ but there were no significant differences between specific pairs of means after adjusting for multiple comparisons.

For Confidence, there were main effects of sex, $F(1,31)=4.37$, p=.045, partial$\eta^2=.124$ and representation, $F=4.678$, $p=.038$, partial $\eta^2=.131$, with women reporting significantly lower Confidence ($M=3.688$, $SE=0.797$) than men ($M=6.477$, $SE=1.070$), and URG reporting significantly greater Confidence ($M=6.525$, $SE=0.986$) than WRG ($M=3.64$, $SE=0.899$, $p=.038$, B, Figure 3, inset).

There was a significant interaction between sex and representation for Anxiety, $F(1,31)=4.296$, $p=.047$, partial $\eta^2=.122$, but there were no significant differences between specific pairs of means after adjusting for multiple comparisons (data not shown).

Further analyses were applied to consider behavior in the presence of a mentor, with senior and junior mentors considered separately. A three-way ANOVA revealed a significant interaction between program model and representation for Engagement in the presence of a senior mentor, $F(1,10)=5.34$, $p=.043$, partial $\eta^2=.348$, but there were no significant differences between specific pairs of means after adjusting for multiple comparisons (data not shown). There was a significant interaction between sex and representation, $F(1,10)=12.67$, $p=.005$, partial $\eta^2=.559$, with WRG women reporting more Engagement in the presence of a senior mentor (11.00, $n=1$) than URG women ($M=4.42$, $SE=1.25$), and URG men reporting more Engagement ($M=10.3$, $SE=1.04$) than WRG men ($M=5.32$, $SE=1.19$; data not shown). A similar interaction was found for Collaboration in the presence of a senior mentor, $F(1,10)=5.3$, $p=.044$, partial $\eta^2=.346$, but there were no significant differences between specific pairs of means after adjusting for multiple comparisons (data not shown).

A main effect of program model was found for Engagement in the presence of a junior mentor, $F(1,25)=8.75$, $p=.007$, partial $\eta^2=.259$, but this must be considered in the context of a significant program x representation interaction, $F(1,25)=8.429$, $p=.008$, partial $\eta^2=.252$ (A, Figure 4). In the AM, WRG students reported greater Engagement ($M=6.98$, $SE=1.70$) than URG ($M=2.39$, $SE=1.90$, $p=.021$), but in the CLM there were no significant differences between WRG and URG. A similar pattern emerged for Collaboration in the presence of a junior mentor, with a significant effect of program model, $F(1,25)=7.114$, $p=.013$, partial $\eta^2=.222$, and a significant program x representation interaction, $F(1,25)=4.473$, $p=.045$, partial $\eta^2=.152$. WRG in the AM reported greater Collaboration in the presence of a junior mentor ($M=4.70$, $SE=1.58$) than did URG ($M=-1.87$, $SE=2.08$, $p=.022$), but there were no significant differences in the CLM (B, Figure 4).

For Confidence in the presence of a junior mentor, we found a significant effect of representation, $F(1,25)=11.69$, $p=.002$, partial $\eta^2=.319$, with URG reporting greater Confidence ($M=8.78$, $SE=0.78$) than did WRG ($M=4.64$, $SE=1.29$), regardless of program (data not shown).

## Figures

Figure 1

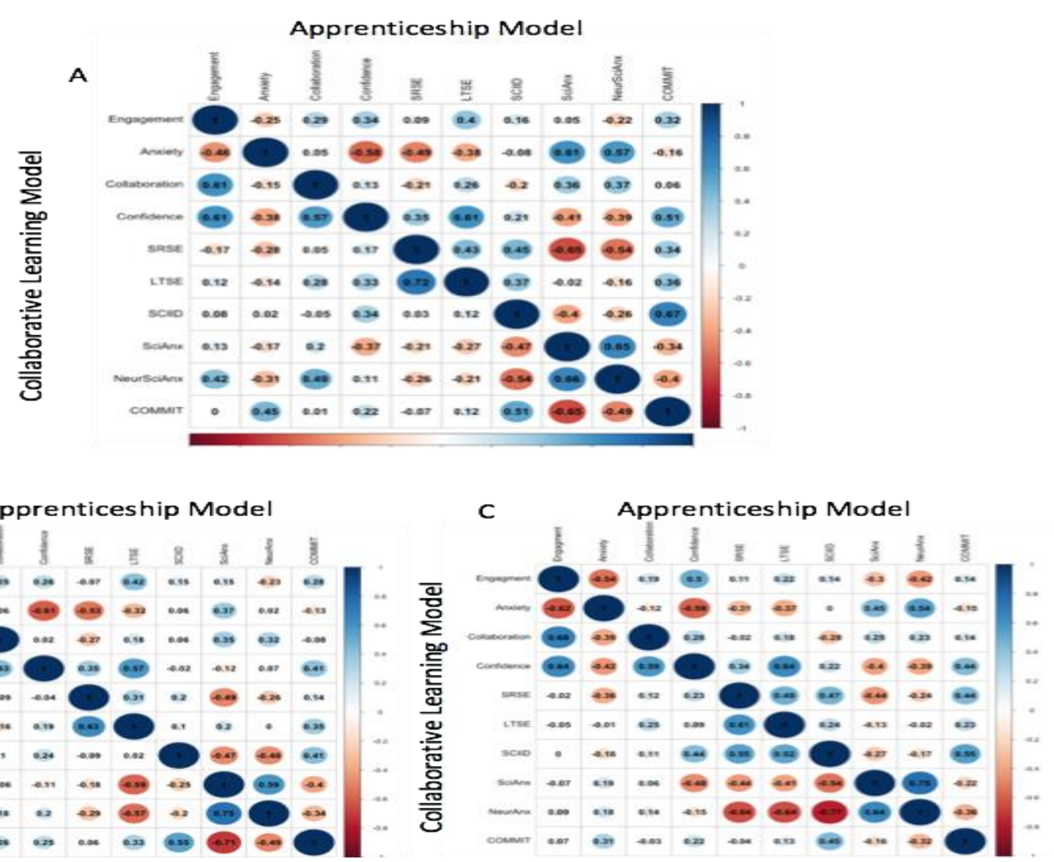


Figure 1. Correlation Matrices of Cognitive-Emotional Status Indicators and Internal Dispositions. All matrices show correlations between pairs of measures from participants in the Apprenticeship Model above the diagonals, and from participants in the Collaborative Learning Model below the diagonals. A) Correlation matrix including ESF cognitive-emotional status indicators averages and averages of pre-program and mid-program internal

dispositions from all time points. B) Correlation matrix including early (2 weeks into the program) ESF cognitive-emotional status indicators responses and pre-program measures of internal dispositions. C) Correlation matrix including late (6 weeks into the program) ESF cognitive-emotional status indicators responses and mid-program measures of internal dispositions.

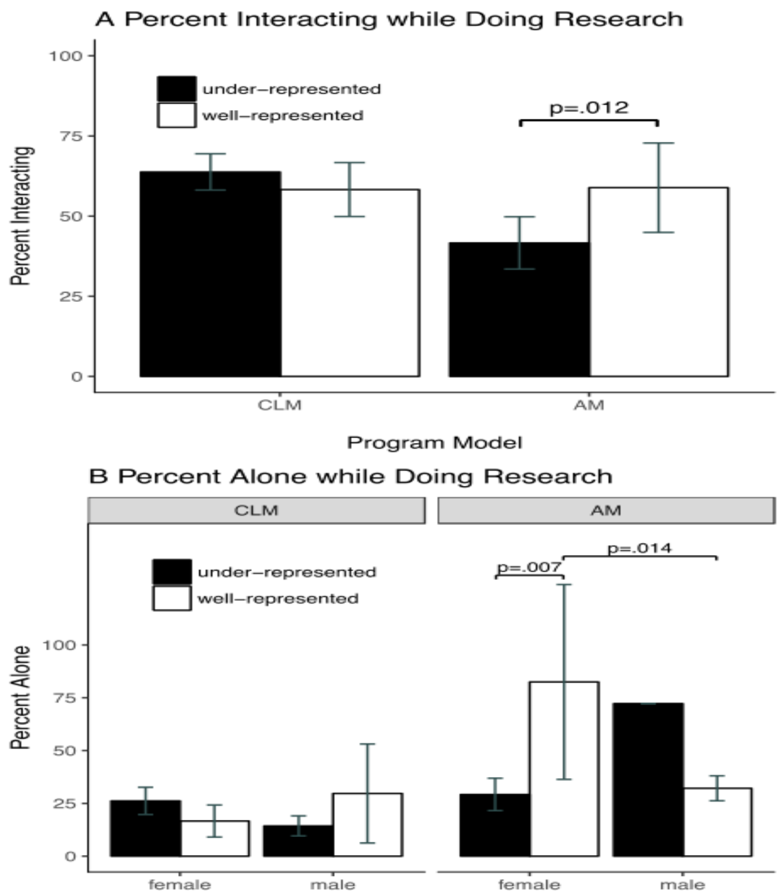


Figure 2. A, Percent Interacting while Doing Research. Bars represent the average percentage of responses during which participants reported being alone while also involved in some research activity. CLM means are plotted on the left and AM on the right. B, Percent Alone while Doing Research. Bars represent the average percentage of responses during which participants reported being alone while also involved in some research activity. CLM means are plotted in the first frame and AM means in the second, with means of female students on the left and male on the right in each frame. For both plots, filled bars indicate means of students from under-represented groups, and empty bars indicate means of students from well-represented groups. Error bars represent +/- one standard error, and *p*-

values indicate significant differences between pairs of means after Sidak correction for multiple comparisons: URG females vs WRG females, and WRG females vs WRG males.

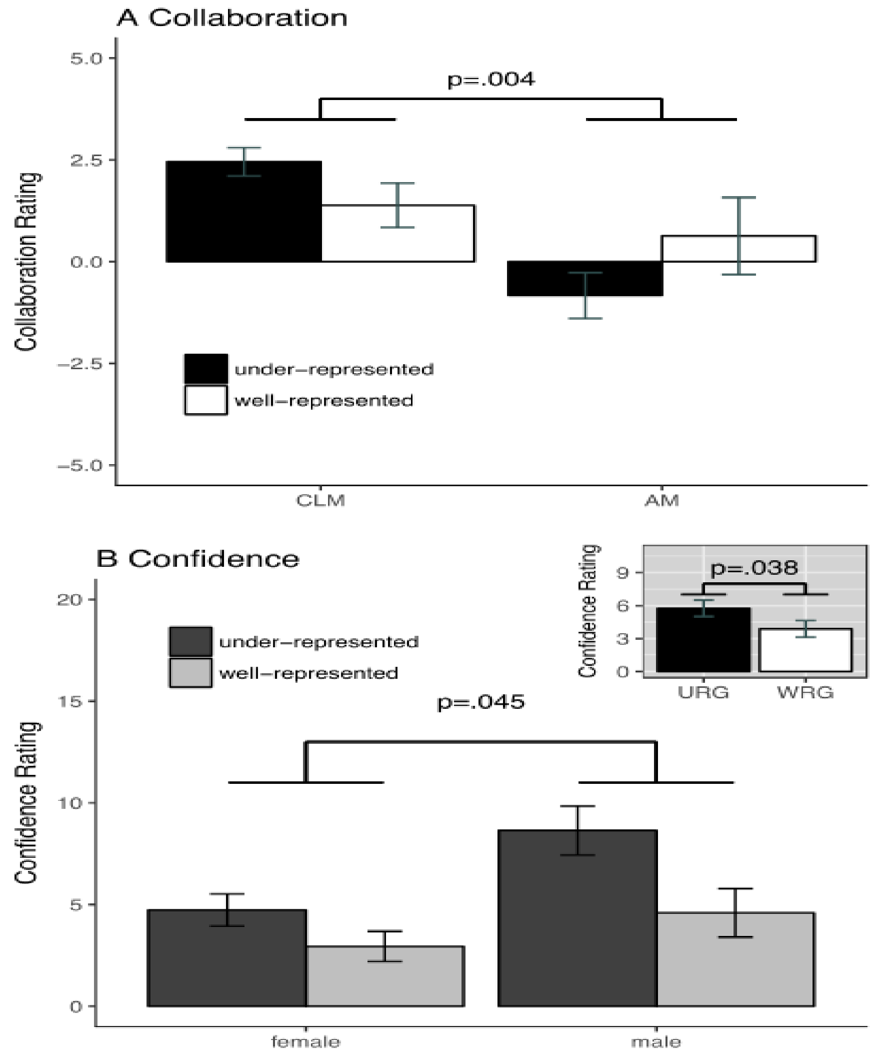


Figure 3. Average cognitive-emotional status indicators. A, Program Model by Representation. CLM means are plotted on the left, and AM on the right. Filled bars indicate means of students from under-represented groups, and empty bars indicate means of students from well-represented groups. B, Sex by Representation. Female means are plotted on the left and male on the right. Dark gray bars indicate means from URG and light gray bars represent means from WRG. For all plots, error bars represent +/- one standard error, and *p*-values indicate significant main effects of program (A) and sex (B) from the three-way ANOVAs. Inset in B shows average confidence rating of

under-represented vs. well-represented groups collapsed across program type and sex; *p*-value indicates the significant main effect of representation from the three-way ANOVA.

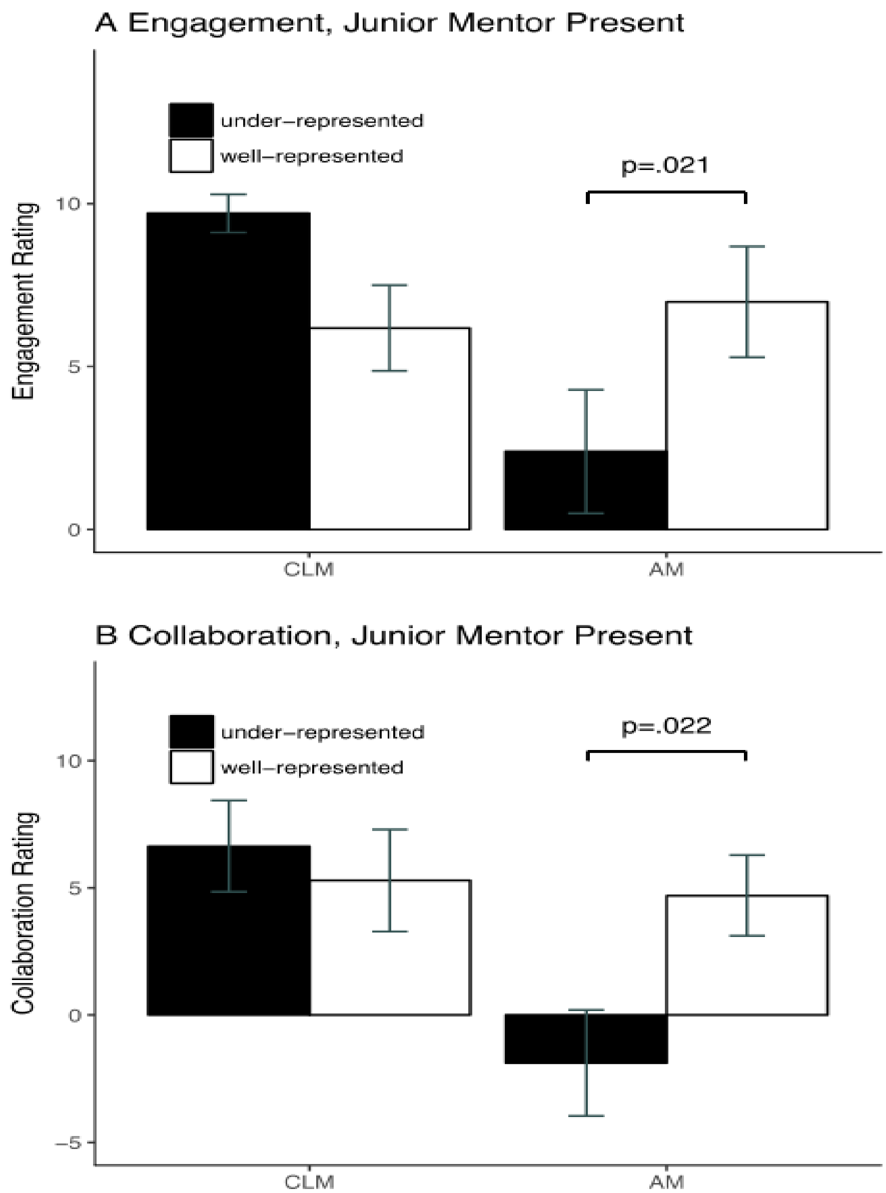


Figure 4. Average cognitive-emotional status indicators reported in the presence of a junior mentor by Program Model and Representation. Mean cognitive-emotional status indicators for A, Engagement, and B, Collaboration, from ESFs indicating the presence of a junior mentor. Filled bars indicate means of students from under-represented groups, and empty bars indicate means of students from well-represented groups. Error bars

represent +/- one standard error, and *p*-values indicate significant differences between pairs of means (URG vs WRG in the AM only) after Sidak correction for multiple comparisons.

Table 1

Program Method Sex Representation

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Total Engagement | Mean | 4.507 | 4.427 | 5.546 | 3.72 | 2.223 | 4.88 | 4.548 |
| Week 1 | Mean | 3.81 | 4.403 | 5.282 | 4.518 | 2.517 | 4.354 | 5.347 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Std. | 5.276 | 3.111 | 5.69 | 3.31 | 1.828 | 5.214 | 3.947 |
| Total Anxiety | Mean | -1.889 | -3.918 | -6.379 | -2.043 | -3.545 | -7.029 | -5.5 |
| Week 1 | Mean | -1.88 | -5.045 | -6.201 | -6.066 | -3.918 | -7.013 | -5.044 |
| Week 2 | Mean | -2.133 | -2.752 | -5.97 | -1.762 | -2.748 | -7.143 | -5.368 |
| Total Collaboration | Mean | 2.134 | 1.383 | 2.904 | 1.385 | -0.657 | 0.746 | -2.619 |
| Week 1 | Mean | 1.72 | 1.45 | 2.631 | 2.19 | -0.3 | 0.802 | -2.962 |
| Error Week 2 Mean | Std. | 2.365 | 0.9 | 3.046 | 2.357 | -0.589 | 0.69 | -1.737 |
| Total Confidence | Mean | 4.348 | 3.276 | 8.773 | 4.534 | 5.004 | 2.124 | 7.976 |
| Week 1 | Mean | 4.536 | 4.217 | 8.181 | 7.234 | 5.129 | 2.542 | 8.701 |
| Week 2 | Mean | 4.408 | 1.719 | 8.932 | 6.69 | 4.625 | 2 | 6.316 |

Table 1. Means and standard errors for all Cognitive-Emotional Composite Scores (CECS): Engagement, Anxiety, Collaboration, and Confidence by Program Method, Sex, and Representation. Total means were averaged across all responses. Week 1 and Week 2 responses are listed separately.

Table 2

A

| Model | | Unstandardized Coefficients B | Std. Error | Standardized Coefficients Beta | t | Sig. |
|---|---|---|---|---|---|---|
| 1 | (Intercept) | | | | | |
| | COMM | .84 | .21 | .611 | 3.94 | .00 |
| 2 | (Intercept) | | | | | |
| | (during research) | .86 | .19 | .621 | 4.37 | .00 |
| | | - | .27 | .349 | 2.45 | .02 |

a. Criterion Variable: post-program COMMIT

B

| Model | | Unstandardized Coefficients B | Std. Error | Standardized Coefficients Beta | t | Sig. |
|---|---|---|---|---|---|---|
| 1 | (Int | .608 | .210 | .438 | 2.898 | .008 |
| | | .379 | .135 | .424 | 2.803 | .010 |
| 2 | (Inte | .682 | .215 | .491 | 3.165 | .004 |
| | | .276 | .157 | .308 | 1.761 | .091 |
| | | -.392 | .310 | .202 | 1.265 | .218 |

a. Criterion Variable: post-program COMMIT

Table 2. A) Results from a hierarchical regression of Confidence during research activity on post-program commitment to a science career, controlling for pre-program commitment in the first block.

B) The same model, but controlling for post-program SCIID along with pre-program COMMIT in the first block.