Title

Investigating the relationship between participation in the building infrastructure leading to diversity (BUILD) initiative and intent to pursue a science career: A cross-sectional analysis.

Authors

Ramos, Hector
Cobian, Krystle
Srinivasan, Jayashri
et al.

Publication Date

2024-02-01

DOI

10.1016/j.evalprogplan.2023.102380

Peer reviewed

HHS Public Access

Author manuscript

Eval Program Plann. Author manuscript; available in PMC 2024 February 05.

Published in final edited form as:

Eval Program Plann. 2024 February ; 102: 102380. doi:10.1016/j.evalprogplan.2023.102380.

Investigating the relationship between participation in the building infrastructure leading to diversity (BUILD) initiative and intent to pursue a science career: A cross-sectional analysis

Hector V. Ramos[a,*], **Krystle P. Cobian**[a], **Jayashri Srinivasan**[a], **Christina A. Christie**[b], **Catherine M. Crespi**[a], **Teresa Seeman**[a]

[a] Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, USA

[b] School of Education and Information Studies, University of California, Los Angeles, USA

Abstract

This paper presents an analysis of survey data to examine the association between supervised structured mentoring and students' intent to pursue a career in science. Data were collected from students in the 10 Building Infrastructure Leading to Diversity (BUILD) research training programs, developed through grants from the National Institutes of Health. Propensity score matching and multinomial logistic regression demonstrated that exposure to BUILD programs—meaning participation in undergraduate research, receipt of mentoring from a primary mentor, and/or participation as a funded scholar and/or associate of each BUILD site's training program—was associated with increased intent to pursue a science career. These findings have implications for STEM program evaluation and practice in higher education.

Keywords

Science education; Undergraduate research; Mentoring; Underrepresented groups; STEM

The National Institutes of Health (NIH) recognizes the need to diversify the scientific workforce by encouraging and enhancing the participation of people from groups identified as underrepresented in the biomedical, clinical, behavioral, and social sciences (collectively

[*] Correspondence to: Coordination and Evaluation Center University of California, 1100 Glendon Avenue Suite 850, Los Angeles, CA 90024, USA. hvramos@ucla.edu (H.V. Ramos).
Hector V. Ramos, Ph.D., Associate Project Scientist, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles.
Krystle P. Cobian Ph.D., Associate Project Scientist, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles.
Jayashri Srinivasan Ph.D., School of Education and Information Studies, University of California
Christina A. Christie Ph.D., Dean, School of Education and Information Studies, University of California, Los Angeles.
Catherine M. Crespi Ph.D., Professor, Biostatistics, University of California, Los Angeles.
Teresa Seeman, Ph.D., Professor, Epidemiology, Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles.

CRediT authorship contribution statement

Hector V. Ramos: Conceptualization, Methodology. **Krystle P. Cobian:** Methodology. **Jayashri Srinivasan:** Methodology. **Christina A. Christie:** Conceptualization. **Catherine Crespi:** Methodology. **Teresa Seeman:** Methodology, Conceptualization.

termed "biomedical") research workforce (https://grants.nih.gov/grants/guide/rfa-files/RFA-RM-18-006.html). Recognizing that problems persist at all career stages of the biomedical career path, the NIH requested evaluation plans from all potential projects to understand what worked and to demonstrate the impacts of programming to support a diverse group of individuals across their careers (Diversity Program Consortium, n.d.). Collectively, this network of institutions awarded funding by the NIH is known as the Diversity Program Consortium (DPC). The DPC is a national collaborative research project in which the NIH works together with institutions to improve training, mentorship, biomedical research career development, and bolster institutional research and research training infrastructure (Diversity Program Consortium, n.d.).

While funding agencies endorse evaluation as crucial for determining whether, why, and for whom a program is working, it is still uncommon to construct the evaluation and coordination component into the multisite research grants. As a noteworthy element of the DPC, the CEC's evaluation serves to help better understand the relationships between the DPC program activities and outcomes, and gain insight into how to best execute a complex evaluative process and then use data to address primary evaluation questions; two features common to most large evaluations.

One challenge of ongoing, multisite evaluation is the need to produce meaningful and rigorous research that informs the field, while waiting for longitudinal data. To maximize the collected data, it is necessary to conduct analyses with currently available data; thus, we used cross-sectional survey data to investigate an important student-focused outcome of the DPC– intent to pursue a biomedical research career. The analyses presented in this paper illustrate how we untangled the complexities of survey data from the 10 BUILD sites to develop a more comprehensive understanding of BUILD program outcomes. Accordingly, we examined whether undergraduate student participation in BUILD programs was associated with increased reporting of intent to pursue a biomedical research career; and further assessed the extent to which frequency of mentoring contributed to any BUILD impact on undergraduate students' career intentions. Using data from the 10 programs, we examined participation in mentored or supervised research (one of the DPC Hallmarks of Success) with respect to undergraduate participation in BUILD programs and respondents' intent to pursue a science-related career. Central to our analyses, we used propensity scores and multinomial logistic regression, and controlled for several additional experiences and demographic information (i.e., income concerns, race/ethnicity, gender, institution, and GPA). The following research questions focused our analyses:

1. Are there differences in undergraduates' intent to pursue a science-related career between students who participated in a BUILD program compared to those students who did not participate in a BUILD program?
2. To what extent do differences in students' reported frequency of mentoring and research participation explain why BUILD exposure is associated with stronger reported intent to pursue a science career?

1. Relevant program and outcome literature

The Diversity Program Consortium (DPC) is a national project funded by the National Institutes of Health (NIH) that implements and evaluates approaches intended to improve research training, mentoring, faculty development and institutional capacity building to support diversity in biomedical research training and career pathways (Diversity Program Consortium, n.d.). One component of the DPC is the Building Infrastructure Leading to Diversity (BUILD) program. Granted to 10 higher education institutions to implement and determine effective ways to engage students from diverse backgrounds in biomedical research, the BUILD initiative's long-term goal is to develop interventions that will best prepare an increased number of students from diverse backgrounds to become potential future contributors to NIH-funded research (Hurtado et al., 2017; McCreath et al., 2017; Guerrero et al., 2022).

To fulfill the DPC's efforts to identify effective approaches concerning research training and mentoring, the DPC includes a Coordination and Evaluation Center (CEC). The CEC provides BUILD sites with coordination and evaluation support, including data coordination, data collection, and overall program evaluation of the BUILD initiative. To work reach these objectives, the CEC partners with the BUILD sites. The corresponding sites administer annual surveys to students, alumni, and faculty.

Research examining baccalaureate-level STEM education reveals that mentoring in STEM (Byars-Winston & Dahlberg, 2019) and undergraduate research experiences (Linn et al., 2015) are associated with students' retention and advancement into STEM graduate study. Associations have been identified between supervised undergraduate research with increased science identity (Atkins et al., 2020; Maton et al., 2016; Summers & Hrabowski, 2006), and with increased academic achievement and longer-term success (Aikens et al., 2017; Atkins et al., 2020; Tsui, 2007; Winterer et al., 2020). Quality of mentoring relationships in these undergraduate research experiences vary (Pfund et al., 2022).

The American Association of Colleges and Universities (AAC&U, n.d.) recognizes undergraduate research experiences (URE) as high-impact practice. UREs contribute to students' interests in STEM and progression into STEM graduate study (Russell et al., 2007), development of science identity (Atkins et al., 2020; Maton et al., 2016), and increased interest in a scientific career (Linn et al., 2015). Research mentoring strongly predicts a variety of academic and career outcomes, including science identity (see, Merolla & Serpe, 2013; Chang et al., 2014) and research self-efficacy (see, Estrada et al., 2018; Schwitzer & Thomas, 1998). In this study, we define UREs as undergraduate research experiences provided by the BUILD programs to their program participants. Course-based undergraduate research experiences developed at some BUILD sites were not included in our analyses.

Mentoring has been associated with increased academic achievement and long-term success in STEM for underrepresented students (Aikens et al., 2017; Atkins et al., 2020; Tsui, 2007; Winterer et al., 2020). STEM education research suggests that mentoring within undergraduate research experiences is essential to successfully support students with STEM career aspirations (Linn et al., 2015). Thus, a key component of the benefits obtained from undergraduate research is the mentoring provided by the primary mentor in charge of the research project and/or lab.

Primary research mentors for UREs can vary by STEM discipline and lab, ranging from principal investigators, postdoctoral researchers, graduate students, to other project/lab staff (Atkins et al., 2020). In a qualitative study of students from underrepresented groups who were involved in UREs via a Howard Hughes Medical Institute (HHMI) and university-sponsored STEM program, researchers found that students perceived research mentors to be supportive colleagues who provided opportunities to develop mentees' science identities and served as examples to model (Atkins et al., 2020). Students that felt less support from their mentors identified lower levels of self-confidence (Haeger & Fresquez, 2016). Moreover, mentoring increases academic GPA (Jones et al., 2010).

Lastly, mentoring can foster undergraduate students' STEM career development and entry into graduate study. Carpi & colleagues (2017) examined career development outcomes of participation in a STEM URE program at a large, minority-serving institution. Among the sample of current and former program participants, 68% reported developing an interest in pursuing graduate school during their program experience. Investigations included how research experiences affect graduate school expectations. Findings suggest that the duration of mentoring relationships between faculty and students via the URE program structure are what partly contribute to shifting students' career aspirations. Haeger and Fresquez (2016) also found that prolonged exposure to mentored research increases research skills when compared to students that received no mentoring or less mentoring; thus, longer mentoring duration also suggests increased academic achievement and persistence.

Considering the large investment by federal and private agencies to support STEM interventions, campuses with grant-funded initiatives and funding organizations have considerable interest in seeking lessons learned on how to extend sustainable, student-centered program innovations to support success in persisting in biomedical research education and the workforce. This study on mentorship and undergraduate research experiences contributes to this effort.

2. Methods

We aimed to determine how frequently primary mentors and students should meet to strengthen students' STEM-related career aspirations. Almost all the BUILD sites required their scholars to participate in a URE during the academic year (and sometimes summer), yet they also offered several program elements that provided additional support (e.g., cohort meetings, access to career workshops, conference funding, etc.). Thus, we tested for differences between BUILD-exposed and non-BUILD students at all 10 BUILD sites, while also examining the relationship between undergraduate students' reported frequency of mentorship with a primary mentor and self-reported intent to pursue a science-related career.

2.1. Data and sample

Data used are from the Student Annual Follow-up Survey (SAFS), developed by the Coordination and Evaluation Center (CEC). Every spring, the CEC invites undergraduate students across the 10 BUILD programs to complete the SAFS. The SAFS asks students about attitudes, perceptions, and participation in a variety of experiences during their time in college. We used SAFS data from the 2017 administration because that version of the survey asked students about the frequency with which they met with their research mentor. Data on mentoring frequency were not collected in subsequent versions of the survey. The study sample includes 4753 undergraduate students who did not participate in BUILD programming and 555 undergraduate students who participated in one or more BUILD-sponsored activity (i.e., scholar, associate, undergraduate research experience) prior to Spring 2017.

3. Variables

3.1. Outcome

Responses to the survey item, *Will you pursue a science-related research career?* were used to measure the study outcome. The response scale was a five-point Likert scale of “definitely yes”, “possibly yes”, “uncertain”, “possibly yes, and ”definitely no”, plus an option to “choose not to answer this question” (see Table A). For these analyses, we collapsed the “choose not to answer” option with missing data, because we lack information on students in both categories regarding “intent to pursue.” We also collapsed “definitely no”, “possibly no”, and “uncertain” into a category we label as “unlikely to pursue a science career” due to small numbers of respondents in each of those categories relative to the size of the other two response options (i.e., “possibly yes” and “definitely yes”), which were retained as separate categories.

3.2. BUILD participation

We operationalized BUILD participation as our key treatment variable. For these analyses, participation in BUILD was comprised of all the students who were identified as BUILD Scholars, BUILD Associates, and/or students with BUILD-sponsored undergraduate research experiences (URE),[1] and who had participated in any of those activities by September 1st, 2016. This date was selected to ensure that students classified as having participated in BUILD had at minimum 6 months for participation in their program category by the time of the administration of the SAFS 2017 follow-up survey which began as early as February 2017 for some of the BUILD sites. While exposure to different types of support and training for biomedical careers varied across institutions, most BUILD sites require undergraduate research for BUILD Scholars while some also offered such URE to additional students who were not receiving the full BUILD package. Our BUILD classification includes Scholars, Associates and students involved in URE. Mentoring was a fundamental component at all campuses.

[1]These categories denote varying levels of participation in BUILD across the sites, and we used the broadest universe of program engagement for the analysis presented here.

The main explanatory variables for the study along with a detailed description for each of the variables are tabulated in Table 1. These variables include race/ethnicity, gender, cumulative grade point average (GPA) for students from their previous academic year, financial concerns while enrolled in college, frequency of mentoring received, research experience, and the BUILD sites (see Table 1 for detailed descriptions). A 5-category frequency of mentoring variable was created based on information regarding (a) whether they have a mentor and (b) the reported frequency of mentoring for those reporting that they have a mentor. These scales are self-reported scores by students on the SAFS 2017 surveys. Missingness across the variables ranged from a minimum of 4% to maximum of 25%. The highest percentage of missing values was seen in the cumulative GPA variable, which is about 25%.

4. Analyses

4.1. Descriptive analyses

We first conducted descriptive analyses comparing the BUILD and non-BUILD students on various demographic and background variables. We calculated the frequencies and percentages for categorical variables, and the mean and standard deviation for the cumulative GPA, which is a continuous variable. The differences between BUILD and non-BUILD students were tested using Chi-square tests and two-sample t-tests.

As noted above, intent to pursue a science-related career is a 3-level Likert-scale item. The 3 categories are: unlikely, possibly yes and definitely yes, with “unlikely” as the reference category. We used a multinomial logistic regression model to model this outcome. First, an unweighted multinomial model was fit to the data, and after propensity score estimation, a weighted multinomial model was fit using the propensity score weights. For both models, covariates were added in a stepwise fashion.

4.2. Propensity score modeling

Since this is an observational study with non-random assignment into treatment groups, we made use of propensity score methods (Hong & Raudenbush, 2008; Rubin, 2001). Data collection via surveys can be expensive and difficult, therefore, we aimed to make use of the entire sample available. Our total sample size was 4753, of whom 555 were BUILD scholars. Because of the large control group, we chose to use the inverse propensity weighting approach rather than matching approaches, which involves dropping unmatched observations from the study. The propensity score method allowed us as to adjust for differences between BUILD scholars and non-BUILD scholars by controlling for a set of confounding variables.

The propensity score is the conditional probability of being assigned to treatment given observed values of covariates (Rosenbaum & Rubin, 1983). Propensity score modeling was used to estimate these conditional probabilities for each individual. The outcome variable for the propensity score model was a binary indicator variable for BUILD vs non-BUILD. Predictors in the propensity score model were gender, race/ethnicity, financial worry, cumulative college GPA prior to 2016, institution, and major. Propensity score estimation and weighting were conducted using the twang package (Ridgeway, 2021) in R (R Core team, 2016). This package implements gradient boosted modeling to fit the propensity score model. Gradient boosted modeling is a tree-based method that allows for nonlinear effects and interactions among the model predictors. The twang package handles missing data by attempting to balance the levels of missingness on each variable through the propensity score weighting. After convergence, we assessed balance between the two groups after weighting by examining standardized mean differences and compared the distributions of propensity scores in the two groups to assess whether there was adequate overlap.

The weights obtained from the propensity score estimation were used to estimate the BUILD effect in a weighted multinomial model that contained a BUILD treatment indicator. We fit a weighted multinomial model using the survey package (Lumley, 2020) in R, which allowed us to properly account for the weighting when estimating standard errors of the parameters in the model. To study treatment effects, we estimated the average treatment effect (ATE) and the average treatment effect on the treated (ATT). The ATE is the overall average effect of BUILD were it to be applied to the entire population. To estimate the ATE, the propensity score weight for each individual is the inverse of the probability of exposure to the condition that the individual was exposed to. The ATT is the average effect of BUILD among individuals who received the treatment. To estimate the ATT, the propensity score weight equals 1 for treated individuals and the odds of treatment for untreated individuals (Stuart, 2010). In this study, we focussed on the ATT.

5. Results

5.1. Descriptive analyses

In Table 2 we present the descriptive statistics for the BUILD and non-BUILD students across the various background characteristics in the study. Comparison of the BUILD students to their non-BUILD counterparts indicated that BUILD students were significantly more likely to report stronger intent to pursue a science career. That is, 31% of BUILD students responded that they will “possibly yes” pursue a science related career, while 27% of non-BUILD students responded to “possibly yes” category. Furthermore, 55% of BUILD students indicated they would “definitely” pursue a science related career, whereas only 21% of non-BUILD students responded that they would “definitely” pursue a science related career. Next, BUILD students were also more likely to report having a mentor (85% BUILD vs 15% non-BUILD), and, among those students with a mentor, reported greater frequency of contact with their mentor. For example, 65% of BUILD students meet with their mentors weekly, while only 14% of non-BUILD students meet with their mentors weekly.

BUILD students were also more likely to report a “biomedical major” and had higher GPA’s than the non-BUILD students. In terms of demographic characteristics, BUILD students were more likely to be women than men and less likely to be Asian or White than Latinx (we have chosen to use the term Latinx as it is the current terminology used in many higher education journals at this time. We understand there are significant questions related to the use of the term); there were no significant differences between BUILD and non-BUILD students who reported being African American.

5.2. Propensity score estimation

Before fitting the outcome model to draw inferences regarding the impact of frequency of mentoring on the intent to pursue a science related career for BUILD and non-BUILD students, we estimated propensity scores in order to use propensity score weighting. Propensity score weighting allowed us to balance key pretreatment covariates between BUILD and non-BUILD students.

First, we conducted a diagnostic check to assess the convergence of the gradient boost algorithm. We examined the absolute standardized effect size and the Kolmogorov-Smirnov (KS) statistics as the two stopping rules and number of iterations were sufficient to minimize the two statistics. Table A in the Appendix presents the pretreatment covariates before and after weighting, which helps understand the covariate balance between the BUILD and non-BUILD students after propensity score weighting. Small values of standardized effect sizes between groups after ATT propensity score weighting indicate that the control group is appropriately weighted on the pretreatment characteristics.

Fig. 1 illustrates the effect of the weights on the magnitude of standardized mean differences between BUILD and non-BUILD students on each pretreatment covariate. We see substantial reductions in effect sizes post weighting for most of the pretreatment covariates, indicating that the differences between the BUILD and non-BUILD students were reduced after weighting. Table 3 provides the relative importance of the covariates included in the propensity score model. Relative importance is a variable importance measure from gradient boosted models that is based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Elith et al., 2008). BUILD institution, cumulative GPA and major (biomedical vs. not) were the three most influential covariates for predicting an individual’s propensity score.

The propensity score estimation and ATT weighting is an important step for us to make comparisons between the BUILD and non-BUILD students with respect to the outcome variable. This technique allowed us to compare similar groups of BUILD vs non-BUILD students in our outcome model.

5.3. Outcome modeling

Table 4 presents results of a series of 3 multinomial unweighted models examining the relationship between BUILD exposure and the likelihood that a student will report “possibly” or “definitely” planning to pursue a career in science. In all cases, BUILD exposure is associated with a greater likelihood of either “probably” or “definitely” pursuing a science career. As shown, when comparing the unadjusted effect of BUILD (Model 1 – no covariates) to the sequence of models controlling for “frequency of mentoring” (Model 2), and then adding “participation in research” (Model 3), and finally all additional demographic covariates (Model 4), BUILD exposure remains a significant predictor, although the size of the effect is reduced with adjustment for frequency of mentoring and research participation.

The estimates reported in Table 4 can be interpreted as odds ratios that compare the odds for "possibly" or "definitely" pursue versus "unlikely" to pursue for a one unit increase in the predictor (Rabe-Hesketh & Skrondal, 2012). For example, in Model 1, we see that the relative odds of a student responding that they definitely intend to pursue a science related career versus that they are unlikely to do so is 10.51 for BUILD students vs. non-BUILD students. The relative odds of responding that they are probably going to pursue a science related career versus unlikely to do so is 4.65 for BUILD students vs. non-BUILD students.

As shown in Model 3, when we further control for frequency of mentoring and research participation, the odds ratio for the BUILD effect is reduced from 10.51 to 3.78 for responding that one is "definitely intending to purse a science career" and from 4.65 to 2.89 for responding that one is "probably going to pursue a science career." These reductions indicate that mentoring and participation in research help to explain a large portion of the BUILD effect. As also shown, when comparing frequency of mentoring against "not having a mentor" (reference group), only weekly or more frequent mentoring is associated with significantly greater likelihood of intending to pursue a science career with an odds ratio of 2.3 for those reporting "probably intending to pursue such a career" and an odds ratio of 2.0 for "definitely intending to pursue a science career". Students who report participating in research are also significantly more likely to report that they "probably" or "definitely" intend to pursue career in science (see Models 3–4).

Finally, independent of BUILD exposure, frequency of mentoring, and research participation, women are less likely than men to report intentions to pursue a science career and Black and Asian students are more likely than White students to report "probably" or "definitely" intending to pursue a science career. /Latinx students are also more likely than White students to report "definitely" intending to pursue such a career.

Table 5 presents a parallel set of weighted multinomial models. Results for BUILD exposure follow a similar trend as the unweighted findings reported above with BULD exposure associated with significantly greater likelihood of reporting "probably pursue" or "definitely pursue" a science related career after adjusting for frequency of mentoring, research participation and major demographic characteristics. Likewise, weekly mentoring and research participation again remain significant even after adjustments for each other and demographic characteristics. The major difference between these weighted analyses and the unweighted models is seen for demographic characteristics which are largely not significant in these weight analyses. The same is true for the "financial worries" measure.

These results help us better understand the importance of mentoring as a significant component of the BUILD experience. After controlling for frequency of mentoring, BUILD students were approximately twice as likely as non-BUILD students to "probably" pursue a science career.

6. Discussion

In this study we examined whether exposure to the NIH-funded BUILD program (being a BUILD Scholar, BUILD Associate, or participation in an URE) was associated with a stronger intent to pursue a science-related research career. We also examined the extent to which frequency of mentoring and participation in research activities were contributors to any BUILD difference. Our findings are consistent with prior research but have the added advantage of using a large national study sample and looking across a multi-site, localized and tailored implementation of program activities.

In the case of our national evaluation of the 10 BUILD programs, the most feasible study design was observational, without random assignment. Thus, it is important to employ statistical methods that best help address the limitations and challenges posed by the study design and survey data collected. We used a propensity score estimation approach to create equivalent groups of students who were exposed to the BUILD program versus those who were not exposed to the BUILD programs and created propensity scores for the students based on the key background characteristics to control for any differences between the BUILD and non-BUILD students. Using inverse propensity weighting, our results indicate that the BUILD and non-BUILD students were similar on the key covariates, which allowed us to make inferences regarding the students exposed to the BUILD program along with its relationship to frequency of mentoring and research participation. This was critical to offering a robust analysis that contributes to our understanding of the program outcomes. It also provides an illustration of how using more advanced post-hoc methodological approaches may address some of the study design and implementation challenges.

6.1. Understandings about the program

Our results from the multinomial regression analyses suggest that several activities/ experiences that are part of the BUILD programs, such as higher frequency of mentoring and research participation are both highly associated with intent to pursue a science related career. First, and most generally, we found that participation in the BUILD program had a strong relationship to a student's intent to pursue a science related career even after controlling for covariates in the study (see Tables 4 and 5). More specifically, frequency of mentoring was found to be a powerful predictor of intent to pursue a science career. Students who met more frequently, that is, weekly or more with a research mentor, were more likely to report interest in pursuing a science-related career; those who received weekly mentoring were found to be 3 times as likely to state they *definitely* intended to pursue a science related career.

We also found that research participation had a significant relationship to the intent to pursue a science related career, after controlling for background characteristics. This study confirmed prior findings (Hurtado et al., 2010) showing that research participation increases intent to pursue among other academic and interpersonal outcomes. When controlling for all other variables, women were about one-third less likely (odds ratio=0.64) to definitely pursue a science career than men. Latinx students were approximately two and a half times more likely (odds ratio=2.5) to definitely pursue a science career compared to White students.

7. Implications

The complexity of the multisite evaluation of the BUILD program provides several lessons related to using statistical models to evaluate program goals. Because the aim of the BUILD initiative is to implement and study biomedical training activities to better understand how to best promote the desired program outcomes beyond those enrolled in the BUILD programs, we are mindful of evaluation use and influence (Alkin & Taut, 2003), and the extent to which our findings might inform programs outside of BUILD. The type of analysis presented in this paper is intended to help educators select program activities for students with a relatively high degree of specificity. For example, institutional leaders can actively implement research participation and mentoring for undergraduates at their campuses. Consequently, since this study is an evaluation of the BUILD program, it specifically identifies mentoring and research participation as two practices that increase intent to pursue. Furthermore, the study provides a strong statistical framework for future studies through its use of propensity scores to create similar groups, thus contributing to the current landscape of evaluation in STEM higher education.

Also important for similar evaluation studies is to build longitudinal, time-varying statistical models. Future analyses might potentially identify dosage of mentoring, type of mentoring, and level of mentor, and examine which of these components are strongest or how they interact on the various academic and interpersonal outcomes. More in-depth mentoring data would be necessary to identify further types of mentoring (e.g., faculty, peer) and specific types of mentoring (e.g., research, academic, career). This could be effective in increasing actual graduate school and/or career outcomes, such as enrollment into graduate study and/or working in a STEM-related occupation.

It would also be ideal to examine the impact of specific research mentorship strategies such as peer mentorship vs. faculty mentorship on a wider variety of psychosocial outcomes that are identified in the literature as important mediators of future STEM career persistence, such as science identity, research self-efficacy, and academic self-efficacy. Science identity, for example, has been identified as an essential quality for success in STEM higher education across a host of outcomes, including intent to pursue, baccalaureate attainment, graduate school enrollment, among others (Eagan et al., 2022).

This information could help program leaders design programs with even greater specificity and increase the impact of the evaluation beyond the DPC. The challenge for evaluators, of course, is to plan in advance of data collection, for all the potential data that might be needed to develop a refined understanding of program activities and outcomes for the program under study and for future program design. Data collection and study participation burden is always weighed against the potential power of the evaluation to inform current and future program decision-making. There is the “ideal” and the “reality” of what can be accomplished in any given evaluation study. Taken in total, findings from analysis of the DPC initiatives have the potential to contribute to enhancing the diversity of the NIH-funded workforce.

Appendix

Table A

Standardized Mean Differences in BUILD (Treatment) and Non-BUILD (Control) Groups Prior to and After Propensity Score Weighting

	Unweighted					ATT Weight				
	tx.mn	**tx.sd**	**ct.mn**	**ct.sd**	**std.eff.sz**	**tx.mn**	**tx.sd**	**ct.mn**	**ct.sd**	**std.eff.sz**
Gender										
Female	0.50	0.50	0.65	0.48	−0.31	0.50	0.50	0.52	0.50	−0.04
Male	0.31	0.46	0.27	0.44	0.09	0.31	0.46	0.31	0.46	−0.01
Gender: Other category	0.01	0.09	0.00	0.04	0.08	0.01	0.09	0.01	0.08	0.03
Gender: NA	0.18	0.39	0.08	0.27	0.27	0.18	0.39	0.16	0.37	0.06
Race/ ethnicity										
Latinx	0.46	0.50	0.30	0.46	0.33	0.46	0.50	0.46	0.50	0.00
Latinx:NA	0.03	0.17	0.04	0.19	−0.05	0.03	0.17	0.02	0.15	0.03
Black/ African American	0.17	0.38	0.16	0.37	0.03	0.17	0.38	0.16	0.37	0.02
Black/ African American:NA	0.03	0.17	0.04	0.19	−0.05	0.03	0.17	0.02	0.15	0.03
Asian	0.14	0.34	0.17	0.38	−0.10	0.14	0.34	0.13	0.33	0.03
Asian:NA	0.03	0.17	0.04	0.19	−0.05	0.03	0.17	0.02	0.15	0.03
Other Race Categories	0.09	0.28	0.10	0.29	−0.03	0.09	0.28	0.08	0.27	0.02
Other Race Categories:NA	0.03	0.17	0.04	0.19	−0.05	0.03	0.17	0.02	0.15	0.03
Financial Worry										
Choose not to answer	0.03	0.17	0.03	0.17	−0.01	0.03	0.17	0.03	0.17	0.00
Major (not sure I will have enough funds to complete college	0.19	0.39	0.20	0.40	−0.02	0.19	0.39	0.19	0.40	−0.01
None (I am confident that I will have sufficient funds)	0.22	0.42	0.19	0.39	0.08	0.22	0.42	0.22	0.42	0.00
Some (but I probably will have enough funds)	0.52	0.50	0.49	0.50	0.07	0.52	0.50	0.52	0.50	0.01
Financial Worry: NA	0.03	0.18	0.09	0.29	−0.32	0.03	0.18	0.04	0.19	−0.02
Cumulative GPA	3.42	0.43	3.21	0.61	0.48	3.42	0.43	3.40	0.45	0.04
Cumulative GPA:NA	0.36	0.48	0.24	0.43	0.25	0.36	0.48	0.38	0.48	−0.04
Sites										
Site A	0.16	0.37	0.14	0.35	0.06	0.16	0.37	0.16	0.36	0.02
Site B	0.23	0.42	0.08	0.27	0.36	0.23	0.42	0.23	0.42	0.00
Site C	0.05	0.22	0.07	0.26	−0.10	0.05	0.22	0.05	0.22	0.00
Site D	0.12	0.33	0.09	0.29	0.08	0.12	0.33	0.12	0.33	−0.01
Site E	0.03	0.18	0.17	0.38	−0.78	0.03	0.18	0.04	0.19	−0.03
Site F	0.09	0.28	0.09	0.28	0.00	0.09	0.28	0.09	0.29	−0.01
Site G	0.09	0.28	0.16	0.37	−0.28	0.09	0.28	0.09	0.29	−0.02
Site H	0.16	0.37	0.11	0.31	0.14	0.16	0.37	0.15	0.36	0.03

	Unweighted					ATT Weight				
	tx.mn	tx.sd	ct.mn	ct.sd	std.eff.sz	tx.mn	tx.sd	ct.mn	ct.sd	std.eff.sz
Site I	0.07	0.25	0.08	0.27	−0.05	0.07	0.25	0.07	0.26	−0.01
Major	0.66	0.47	0.44	0.50	0.46	0.66	0.47	0.65	0.48	0.02
Biomedical Natural Science	0.18	0.38	0.12	0.32	0.16	0.18	0.38	0.18	0.39	−0.01
Biomedical Social Science	0.07	0.25	0.35	0.48	−1.12	0.07	0.25	0.08	0.27	−0.04
Non-Biomedical										
Major:NA	0.10	0.30	0.10	0.30	0.00	0.10	0.30	0.09	0.29	0.02

References

Aikens ML, Robertson MM, Sadselia S, Watkins K, Evans M, Runyon CR, Eby LT, & Dolan EL (2017). Race and gender differences in undergraduate research mentoring structures and research outcomes. CBE life Sciences Education, 16 (2), ar34. 10.1187/cbe.16-07-0211 [PubMed: 28550078]

Alkin M, & Taut S (2003). Unbundling evaluation use. Studies in Educational Evaluation, 29, 1–12.

American Association of Colleges & Universities (AAC&U). (n.d.) High-impact practices. AAC&U. https://www.aacu.org/trending-topics/high-impact.

Atkins K, Dougan BM, Dromgold-Sermen MS, Potter H, Sathy V, & Panter AT (2020). “Looking at Myself in the Future”: how mentoring shapes scientific identity for STEM students from underrepresented groups. International Journal of STEM Education, 7(1), 1–15.

Byars-Winston A, & Dahlberg ML (2019). The science of effective mentorship in STEMM. Consensus study report (p. 20001). 500 Fifth Street NW, Washington, DC: National Academies Press.

Carpi A, Ronan DM, Falconer HM, & Lents NH (2017). Cultivating minority scientists: Undergraduate research increases self-efficacy and career ambitions for underrepresented students in STEM. Journal of Research in Science Teaching, 54(2), 169–194.

Chang MJ, Sharkness J, Hurtado S, & Newman CB (2014). What matters in college for retaining aspiring scientists and. engineers from underrepresented racial groups. Journal of Research in Science Teaching, 51, 555–580. 10.1002/tea.21146

Diversity Program Consortium. (n.d.) We, the DPC. https://diversityprogramconsortium.org/dpc/our-story.

Estrada M, Hernandez PR, & Schultz PW (2018). A Longitudinal Study of How Quality Mentorship and Research. Experience Integrate Underrepresented Minorities into STEM Careers. CBE life Sciences Education, 17(1), ar9. 10.1187/cbe.17-04-0066 [PubMed: 29351912]

Guerrero LR, Seeman T, McCreath H, Maccalla NM, & Norris KC (2022). Understanding the context and appreciating the complexity of evaluating the Diversity Program Consortium. New Directions for Evaluation, 2022(174), 11–20. [PubMed: 37538950]

Haeger H, & Fresquez C (2016). Mentoring for Inclusion: The Impact of Mentoring on Undergraduate Researchers in the. Sciences. CBE life Sciences Education, 15(3), ar36. 10.1187/cbe.16-01-0016 [PubMed: 27543635]

Hong G, & Raudenbush SW (2008). Causal Inference for Time-Varying Instructional Treatments. Journal of Educational and Behavioral Statistics, 33(3), 333–362. 10.3102/1076998607307355

Hurtado S, White-Lewis D, Norris K, McCreath HE, Norris KC, Calderón NE, & Short A (2017). The Diversity. Program Consortium: Innovating educational practice and evaluation along the biomedical research pathways. In BMC Proc (Vol. 11)(No. Supplement 12).

Linn MC, Palmer E, Baranger A, Gerard E, & Stone E (2015). Undergraduate research experiences: Impacts and opportunities. Science, 347(6222), 1261757. [PubMed: 25657254]

Lumley T (2020). survey: analysis of complex survey samples.

Maton KI, Beason TS, Godsay S, Sto Domingo MR, Bailey TC, Sun S, & Hrabowski FA 3rd (2016). Outcomes and Processes in the Meyerhoff Scholars Program: STEM PhD Completion, Sense of Community, Perceived Program Benefit, Science Identity, and Research Self-Efficacy. CBE life Sciences Education, 15(3), ar48. 10.1187/cbe.16-01-0062 [PubMed: 27587857]

McCreath HE, Norris KC, Calderón NE, Purnell DL, Maccalla NM, & Seeman TE (2017, December). Evaluating efforts to diversify the biomedical workforce: the role and function of the Coordination and Evaluation Center of the Diversity Program Consortium. In BMC proceedings (Vol. 11, No. 12, pp. 15–26). BioMed Central. [PubMed: 29375658]

Merolla DM, & Serpe RT (2013). STEM enrichment programs and graduate school matriculation: The role of science identity salience. Social Psychology of Education: An International Journal, 16(4), 575–597. 10.1007/s11218-013-9233-7 [PubMed: 24578606]

Pfund C, Sancheznieto F, Byars-Winston A, Zárate S, Black S, Birren B, Rogers J, & Asai DJ (2022). Evaluation of a Culturally Responsive Mentorship Education Program for the Advisers of Howard Hughes Medical Institute Gilliam Program Graduate Students. CBE life Sciences Education, 21(3), ar50. 10.1187/cbe.21-11-0321 [PubMed: 35862583]

R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from. https://www.R-project.org/.

Rabe-Hesketh S, & Skrondal A (2012). Multilevel and Longitudinal Modeling using Stata Volume II: Categorical Responses. Count and Survival. #rd edition. College Station, TX: Stata Press,.

Ridgeway G, et al. Toolkit for weighting and analysis of nonequivalent groups: a guide to the twang package. https://cran.r-project.org/package=twang 2021. pp. 1–31.

Rosenbaum PR, & Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55.

Rubin DB (2001). Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services & Outcomes Research Methodology, 2, 169–188. 10.1023/A:1020363010465

Russell SH, Hancock MP, & McCullough J (2007). Benefits of undergraduate research experiences. Science, 316(5824), 548–549. [PubMed: 17463273]

Schwitzer AM, & Thomas C (1998). Implementation, utilization, and outcomes of a minority freshman peer-mentor program at a predominately White university. Journal of The Freshman Year Experience & Students in Transition, 10(1), 31–50.

Stuart EA (2010). Matching methods for causal inference: A review and a look forward. Statistical science: a Review Journal of the Institute of Mathematical Statistics, 25(1), 1. [PubMed: 20871802]

Summers MF, & Hrabowski FA 3rd (2006). Diversity. Preparing minority scientists and engineers. Science, 311(5769), 1870–1871. 10.1126/science.1125257 [PubMed: 16574853]

Tsui L (2007). Effective strategies to increase diversity in STEM fields: A review of the research literature. The Journal of Negro Education, 76(4), 555–581. ⟨http://www.jstor.org/stable/40037228⟩.

Winterer ER, Froyd JE, Borrego M, Martin JP, & Foster M (2020). Factors influencing the academic success of Latinx students matriculating at 2-year and transferring to 4-year US institutions—implications for STEM majors: A systematic review of the literature. International Journal of STEM Education, 7(1). 10.1186/s40594-020-00215-6

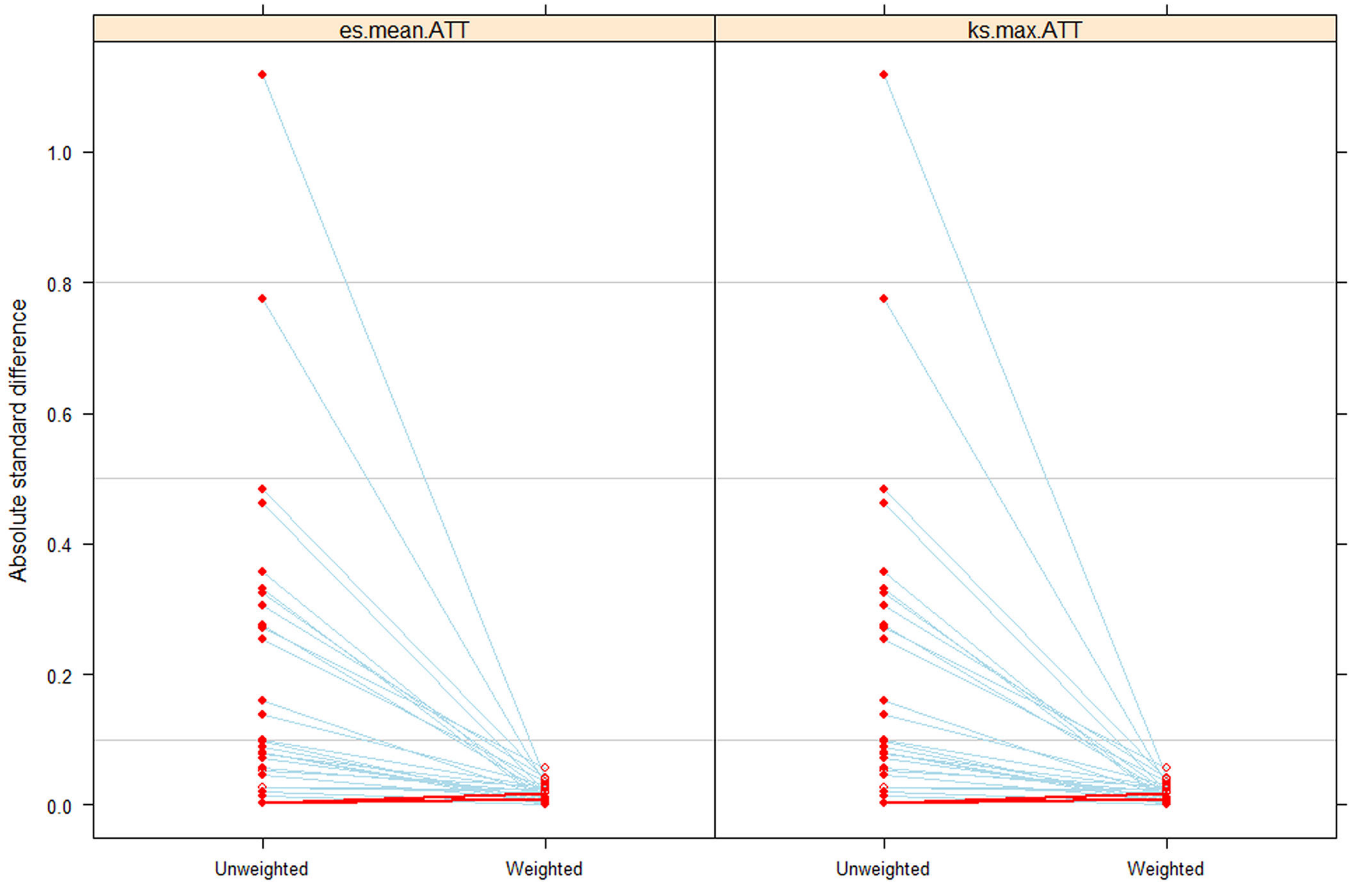

Fig. 1. Plots Comparing Difference in Covariate Distributions Between BUILD Exposed and Not Exposed students before (unweighted) and after (weighted) Applying Propensity Scores.

Fig. 1: Plots comparing difference in covariate distributions between BUILD exposed and not exposed students before (unweighted) and after (weighted) applying propensity scores. Closed red circles indicate a statistically significant difference. Each line represents a covariate. Left: absolute standardized mean differences. Right: Kolmogorov-Smirnov statistics.

Table 1

Description of study variables.

Variable	Description
Intent to pursue (dependent variable)	Will you pursue a science-related research career? Rating scale: (1) Definitely no, possibly no or uncertain (2) Possibly yes (3) Definitely yes
BUILD Student (focal treatment variable)	0 = no, 1 = yes
Race/ethnicity	White (Reference group) /Latinx Black/ African American Asian Other Race Category
Gender	Male (Reference group) Female Others
Site	Institution
Cumulative GPA	Students' cumulative GPA for the prior year (2016)
Major	Major declared in the SAFS 2017 survey with three categories: Non-Biomedical Field Biomedical Social Science Field Biomedical Natural Science Field (Reference group)
Financial concern	Do you have any concern about your ability to finance your college education? Rating Scale: (1) None (I am confident that I will have sufficient funds) (2) Some (but I probably will have enough funds) (3) Major (not sure I will have enough funds to complete college) (4) I choose not to answer
Frequency of mentoring	How often do you usually communicate with your primary mentor about your research? Rating Scale: (1) Weekly or more often (2) Monthly (3) Several time a year (4) Annually or less (5) (5) No Mentor (Reference Group)
Research participation	Conduct/participation in scientific research: 0 = no, 1 = yes

Table 2

Descriptive statistics of study variables by BUILD exposure.

	Non-BUILD students (n = 4198)		BUILD students (n = 555)		Pearson's Chi-squared test
Variable	**n**	**%**	**n**	**%**	
Outcome: Intent to Pursue					
Cat 1: (Sum of Categories 1,2,&3)	1666	52.69	69	13.37	
Cat 2: Possibly Yes	841	26.6	162	31.4	
Cat 3: Definitely Yes	655	20.71	285	55.23	Chi-square = 355.92.22, *df* (2); $p < .001$
Missing (NA)	1036	24.68	39	7.03	
Financial Worry					
I choose not to answer	131	3.44	16	2.99	
Major (not sure I will have enough funds to complete college	834	21.91	106	19.78	
None (I am confident that I will have sufficient funds)	798	20.96	124	23.13	
Some (but I probably will have enough funds)	2044	53.69	290	54.1	Chi-square = 2.33, *df* (3); $p = 0.506$
Missing	391	9.31	19	3.42	
Major					
Biomedical Natural Science Field	1842	48.76	365	73	
Biomedical Social Science Field	485	12.84	98	19.6	
Non-Biomedical Field	1451	38.41	37	7.4	Chi-square = 187.18, *df* (2); $p < 0.001$
Missing	420	10	55	9.91	
Gender					
Female	2746	70.99	278	61.37	Chi-square = 17.8, *df* (1); $p < 0.0001$
Male	1115	28.83	170	37.53	Chi-square = 14.69, *df* (1); $p < 0.001$
Other	7	0.18	5	1.1	
Missing	330	7.86	102	18.38	
Race/ ethnicity					
Asian	684	16.931	73	13.54	
Black/African American	650	16.089	92	17.07	
Latinx	1206	29.851	250	46.38	
White	1116	27.624	77	14.29	
Other Race Categories	384	9.506	47	8.72	Chi-square = 7.23, *df* (4); $p < 0.001$
Missing	158	3.764	16	2.88	
	n	**Mean (SD)**	**n**	**Mean (SD)**	**t-test**
Cumulative GPA for prior year from 2016	3202	3.21 (0.61)	356	3.42 (0.43)	$t = -8.227$ ($df = 527.28$) $p < 0.001$

Table 3

Relative influence scores for covariates in the propensity score model.

Variable	Relative Influence (%)
BUILD institutions/ Sites	46.55
Cumulative GPA	19.45
Major	16.64
Gender	8.44
Latinx	4.25
Financial Worry	2.81
Black/ African American	1.45
Asian	0.22
Other Race Categories	0.18

Table 4

Multinomial Models for the Outcome of Intent to Pursue a Science Career: Unweighted Results.

Variables in model	Odds Ratio [95% Confidence Intervals]							
	Model 1		Model 2		Model 3		Model 4	
	Possibly Pursue	Definitely Pursue	Possibly Pursue	Definitely Pursue	Possibly Pursue	Definitely Pursue	Possibly Pursue	Definitely Pursue
Intercept	0.5 *** [0.46,0.55]	0.39 *** [0.36,0.43]	0.46 *** [0.42,0.51]	0.3 *** [0.27,0.34]	0.43 *** [0.39,0.48]	0.26 *** [0.23,0.3]	0.36 *** [0.28,0.47]	0.21 *** [0.16,0.28]
BUILD	4.65 *** [3.45,6.28]	10.51 *** [7.91,13.95]	3.79 *** [2.73,5.25]	6.08 *** [4.44,8.31]	2.89 *** [2.06,4.06]	3.78 *** [2.72,5.25]	3.08 *** [2.09,4.53]	3.97 *** [2.72,5.78]
Frequency of mentoring								
Annually or less			0.79 [0.47,1.33]	0.98 [0.56,1.7]	0.81 [0.48,1.36]	0.98 [0.55,1.72]	0.73 [0.42,1.28]	0.8 [0.42,1.51]
Monthly			1.2 [0.88,1.64]	2.08 *** [1.52,2.83]	1.12 [0.81,1.53]	1.76 *** [1.28,2.42]	1.15 [0.82,1.6]	1.6 *** [1.13,2.27]
Several times a year			1.12 [0.77,1.62]	1.46 * [0.99,2.15]	1.1 [0.76,1.6]	1.36 [0.91,2.03]	1.18 [0.8,1.73]	1.47 [0.96,2.23]
Weekly or More			1.67 *** [1.31,2.13]	3.32 *** [2.61,4.21]	1.36 * [1.06,1.75]	2.31 *** [1.79,2.98]	1.43 ** [1.09,1.86]	2.25 *** [1.72,2.95]
Research Participation					2 *** [1.57,2.54]	3.24 *** [2.56,4.1]	2.02 *** [1.56,2.61]	3.16 *** [2.44,4.08]
Gender								
Female							0.8 * [0.66,0.98]	0.67 *** [0.54,0.82]
Others							3.44 [0.27,44.48]	3.54 [0.31,40.41]
Race/ ethnicity								
Latinx							0.93 [0.73,1.18]	1.54 *** [1.2,1.99]
Asian							1.35 * [1.04,1.77]	1.41 * [1.04,1.93]
Black							1.02 *** [0.76,1.38]	1.55 ** [1.12,2.14]
Other categories							0.96 *** [0.68,1.35]	1.29 [0.88,1.9]
Financial worry								
Some Worries							1.41 *** [1.13,1.77]	1.26 * [0.99,1.6]
Major Worries							1.44 ** [1.09,1.91]	1.58 *** [1.18,2.13]

Note: Significance level:

* $p < .05$

** $p < .01$

*** $p < 0.001$

Table 5

Multinomial Models for the Outcome of Intent to Pursue a Science Career: Propensity Score-weighted Results.

Variables in model	Odds Ratio [95% Confidence Intervals]							
	Model 1		Model 2		Model 3		Model 4	
	Possibly Pursue	Definitely Pursue	Possibly Pursue	Definitely Pursue	Possibly Pursue	Definitely Pursue	Possibly Pursue	Definitely Pursue
Intercept	0.74 *** [0.62,0.9]	0.67 *** [0.55,0.81]	0.46 *** [0.42,0.51]	0.30 *** [0.27,0.34]	0.63 *** [0.5,0.79]	0.46 *** [0.35,0.59]	0.40 *** [0.24,0.69]	0.36 *** [0.2,0.63]
BUILD	3.16 *** [2.24,4.46]	6.19 *** [4.43,8.66]	3.79 *** [2.73,5.25]	6.08 *** [4.44,8.31]	1.60 * [1.03,2.48]	2.08 *** [1.35,3.2]	1.91 *** [1.21,3.03]	2.62 *** [1.64,4.19]
Frequency of mentoring								
Annually or less			0.79 [0.47,1.33]	0.98 [0.56,1.72]	0.68 [0.27,1.7]	1.02 [0.33,3.15]	0.60 [0.2,1.81]	0.46 [.15,1.45]
Monthly			1.20 [0.88,1.64]	2.08 *** [1.52,2.83]	1.41 [0.8,2.48]	1.74 * [1,3.03]	1.39 [0.75,2.58]	1.52 [0.81,2.85]
Several times a year			1.12 [0.77,1.62]	1.46 * [0.99,2.15]	0.96 [0.45,2.06]	1.31 [0.53,3.24]	0.97 [0.43,2.18]	1.35 [0.55,3.29]
Weekly or more often			1.67 *** [1.31,2.13]	3.32 *** [2.61,4.21]	2.50 *** [1.53,4.09]	3.63 *** [2.25,5.87]	2.50 *** [1.48,4.24]	3.47 *** [2.05,5.88]
Research participation					1.63 * [1.06,2.5]	2.45 *** [1.6,3.77]	1.58 * [0.99,2.52]	2.23 *** [1.41,3.55]
Gender								
Female							0.79 [0.54,1.17]	0.65 * [0.44,0.96]
Others							Not estimable	Not estimable
Race/ ethnicity								
Latinx							1.44 [0.91,2.3]	2.50 *** [1.56,4.02]
Asian							1.32 [0.76,2.29]	1.40 [0.77,2.55]
Black							1.12 [0.64,1.94]	1.12 [0.62,2.03]
Other categories							1.08 [0.57,2.05]	0.86 [0.43,1.75]
Financial worry								
Some Worries							1.56 * [1,2.43]	1.11 [0.7,1.77]
Major Worries							1.49 [0.86,2.59]	1.23 [0.7,2.16]

Note: Significance level:

* $p < .05$

** $p < .01$

*** $p < 0.001$