BUILDing an Early Advantage: An Examination of the Role of Strategic Interventions in Developing First-Year Undergraduate Students' Science Identity

M. Kevin Eagan Jr., Ph.D.,
University of California, Los Angeles

Ana L. Romero, Ph.D.,
University of California, Los Angeles

Shujin Zhong, Ph.D.
University of Maryland

Introduction

The federal government, via agencies like the National Institutes of Health (NIH) and the National Science Foundation (NSF), has invested billions in research training programs to diversify the scientific workforce. Reports have advocated for a more diverse pool of scientists (National Academies of Sciences, Engineering, and Medicine, 2007; Sullivan Commission, 2004; U.S. Department of Education, 2006), and concerns exist about a potential shortage of qualified scientists (President's Council of Advisors for Science and Technology, 2012). Federal and non-governmental agencies fund programs that capture the imagination of young aspiring scientists and develop research skills, talent, and innovation among undergraduates in science, technology, engineering, and mathematics (STEM). Other opportunities fund graduate students and financially support early career scientists, helping them to expand professional and academic networks, protect their research time, and establish their laboratories.

Individuals have directly benefited from these programs, but it's unclear to what extent they've contributed to scalable, effective, and cost-efficient strategies for diversifying the scientific workforce. In 1980, Black, Latino, and Native American graduates earned 2.2% of doctorates in the life sciences and 1.7% of doctorates in the physical sciences (National Science Foundation [NSF], 2019). By 2000, the share of doctorates earned by these graduates had increased to 5.7% in the life sciences and 4.1% in the physical sciences. In 2019, 10.6% of all doctorates in the life sciences and 5.5% of all doctorates in the physical sciences were earned by Black, Latino, and Native American graduates (NSF,

2019). Despite these gains, persistent disparities exist in certain measures of career success (e.g., successful grant applications) (Ginther et al., 2011).

NIH and the Founding of the Diversity Program Consortium

These findings suggest that the investments made to diversify the scientific workforce may have had some limited success with respect to improving racial diversity but may be falling short in ensuring more equitable outcomes related to the career success of scientists from traditionally underrepresented groups. In the years following the publication of Ginther et al.'s (2011) study, the NIH funded initiatives focused on faculty and students. The Building Infrastructure Leading to Diversity (BUILD) initiative has similarities with other NIH-funded student-focused training programs but distinguishes itself by also providing support to faculty training programs and institutional capacity building (Hurtado et al., 2017).

At its core, the BUILD initiative captures students' attention and imagination by exposing them to hands-on research experiences. The BUILD program aims to achieve its vision by enabling students to develop a strong science identity; namely, to begin seeing themselves as scientists by providing opportunities for them to think and act like scientists. Students selected into the program are designated as BUILD Scholars and awarded funding. This article reports the results from an examination of data collected during students' first year of college to evaluate whether students involved in the BUILD program finished their first year of college with a stronger science identity than their counterparts pursuing similar majors and sharing similar pre-college and in-college experiences. Specifically, this study addresses the following research questions:

1. Do BUILD Scholars make significantly stronger gains in their development of science identity by the end of their first year of college compared to first-year students with similar demographic characteristics and academic indicators who are not BUILD Scholars?

2. Does the effect of participating in BUILD as a Scholar vary based on participants' racial/ethnic identity or gender identity?

Defining Science Identity

Individuals develop the confidence and competence to begin viewing themselves as scientists by actively thinking and acting like scientists, in addition to having their self-perception reinforced by meaningful others such as faculty, mentors, and family (Carlone & Johnson, 2007; National Academies of Sciences, Engineering, and Medicine, 2019). Carlone and Johnson (2007) posit that three distinct domains serve to cultivate an individual's science identity: competence, performance, and recognition. Science identity forms from developing competence in a particular domain through education, trainings, and hands-on experiences. Performing as a scientist includes authentic experiences that provide opportunities to perform or take on the roles and responsibilities of a scientist. Finally, recognition from meaningful others may come from receiving awards, rewards, publication acceptances, job offers, or other forms of validation for making contributions to the field (Carlone & Johnson, 2007).

In the case of undergraduate students pursuing science-related majors, science identity develops as students cultivate their competence in science. Students develop science identity by mastering concepts in their coursework and developing specific inquiry and technical skills relevant to their particular domain of interest. Science identity is also enhanced when students apply these technical skills through hands-on research experiences. Undergraduate students have opportunities to perform as scientists in their coursework and in mentored research experiences with faculty. Dissemination of research findings via conference presentations and publications also serve as another avenue for students to perform as scientists. Receiving feedback from mentors and faculty, awards and honors, and course grades recognizes students' development as a scientist and enhances their science identity. Hazari et al. (2010) propose that a fourth component of science identity includes demonstrating an interest in STEM. They argue that interest in STEM is a given component in Carlone and Johnson's (2007) discussion of science identity, but, without a cultivated interest in a scientific domain, an individual's identity as a scientist may wane (Hazari et al., 2010).

Science Identity as a Hallmark of Success

Science identity relates to several short- and long-term academic indicators. Students who more strongly identify as scientists tend to earn higher grades in their STEM-related courses and higher grade point averages (GPAs) overall (Lu, 2015; Stets et al., 2017). Success in STEM coursework, particularly in introductory STEM courses, often contributes to longer-term success in STEM, as students may internalize their early academic success as recognition from their faculty that they have the potential to become a scientist. Similarly, students who report stronger science identities tend to have higher odds of persisting in their science-related majors (Chang et al., 2011; Seymour et al., 2004; Trujillo & Tanner, 2014). Undergraduate students' science identity also contributes to the likelihood of pursuing graduate education in the sciences (Eagan et al., 2013; Estrada et al., 2011; Piatt et al., 2019) and long-term STEM career intentions (Byars-Winston & Rogers, 2019; Estrada et al., 2018; Estrada et al., 2011; Hazari et al., 2010; Hazari et al., 2013; Stets et al., 2017). Thus, evidence suggests that interventions targeted to undergraduate students with an explicit aim of developing their identity as scientists have the potential to directly or indirectly increase their likelihood of persisting in science to bachelor's degree completion. A strong sense of science identity can also lead to students continuing their education by pursuing graduate degrees in STEM-related fields and seeking careers in STEM fields.

Strengthening College Students' Science Identity

Several experiences in college contribute to building a sense of science identity among undergraduate STEM majors. Participation in authentic research experiences significantly contributes to building science identity among undergraduate students (Byars-Winston & Rogers, 2019; Carlone & Johnson, 2007; Estrada et al., 2018; Lane, 2016; Seymour et al., 2004; Stets et al., 2017). In authentic research experiences, student investigators explore novel research questions as part of the process and follow protocols provided by the instructor; in other types of research experiences, students may be presented with a research problem and must then develop their own question and protocol to address the problem (Spell et al., 2014). Building research opportunities into the curriculum addresses

the opportunity gap for students who do not have access to mentored research experiences with faculty or to structured research programs, as undergraduate research experiences collectively assist students in building their competence in a particular domain and their methodological and critical thinking skills as investigators. These types of experiences allow students to experience first-hand how science is "done" and contribute to students' development of a stronger identity as a scientist (Byars-Winston & Rogers, 2019).

Although research experiences offer a platform for students to develop competence in a discipline and later perform as scientists, effective mentorship from scientific experts is also critical to developing a science identity (Byars-Winston & Rogers, 2019; Estrada et al., 2018; Piatt et al., 2019; Robnett et al., 2018). Undergraduate students who report receiving more instrumental (skill-based) mentoring and socioemotional mentoring tend to report feeling more strongly identified with science (Robnett et al., 2018). Whether students connect with faculty in formal or informal spaces, establishing a solid rapport with professors represents a critical component to students' likelihood of success in pursuing graduate education and cultivating a professional academic network of colleagues and collaborators. Mentors have a critical role in providing the space for students to advance in their competency and skill development, facilitating opportunities to perform as scientists through research experiences or venues to disseminate research findings, and, perhaps most importantly, recognizing their mentees as scientists.

Mentors also serve as advocates and sounding boards, especially when their mentees come from underrepresented groups. Underrepresented students who have had a mentor with the ability to discuss and acknowledge underrepresentation in STEM have reported stronger science identities (Hazari et al., 2010). For example, among women in physics, self-reported science identity was stronger among students who had advisors who more regularly discussed the underrepresentation of women in physics compared to women whose mentors did not engage in such discussions (Hazari et al., 2010). The ability for mentors to speak with students about the disparities in representation create opportunities to engage in more meaningful discussion with underrepresented students. Studies have also shown that it is important to engage in a discussion on the connection between STEM fields and values (Carlone & Johnson, 2007; Hazari et al., 2010; Hazari et al., 2013). These conversations inject a human element into scientists' work and may also serve as a signal to mentees from underrepresented groups that the mentor is aware of and even sympathetic to the cultural realities of the department, discipline, or field (Hernandez et al., 2017).

Finally, the recognition component of Carlone and Johnson's (2007) framework for the manifestation and development of science identity includes viewing oneself as a science person, so to identify as a scientist means that an individual needs to develop a level of confidence in skills and knowledge within a particular field of study. Carlone and Johnson (2007) found that repeated recognition from meaningful others, such as members of the scientific community, contributed to having a stronger science identity for women of color. Rodriguez et al. (2019) found that, for Latina undergraduate STEM majors, recognition from peers, faculty, and family positively contributed to participants' view of themselves as a STEM person. Similarly, researchers have determined that encouragement from teachers and family contribute to having a stronger physics identity (Hazari et al., 2010). Science

self-efficacy is the belief in one's ability to carry out scientific research and communicate findings. Building competence in research skills leads students to believe they know how to *do* science and is positively associated with doing well in science courses (White et al., 2019). Science self-efficacy has strong correlations with an individual's identity as a scientist. As individuals report feeling more confident in their ability to design an answerable research question or conduct an experiment, they also tend to report feeling more identified with science (Byars-Winston & Rogers, 2019; Estrada et al., 2011). Self-efficacy is also an important predictor of the likelihood of pursuing a scientific career (Byars-Winston & Rogers, 2019; Estrada et al., 2011) and contributes to building a science identity.

Social identities also influence the ability for students to develop a science identity (Carlone & Johnson, 2007). The strength of individuals' science identity varies significantly across categories of race and gender. For example, white females tended to view themselves as biologists more than white males; by contrast, males generally had a greater proclivity to identify as physicists compared to females (Hazari et al., 2013). Additionally, Hispanic females were either below or on par with other women and both white men and men of color with regards to their science identity (Hazari et al., 2013). These differential experiences along race and gender suggest that the development of a science identity does not occur uniformly across demographic groups; therefore, we should not expect experiences associated with facilitating science identity development to affect participants in identical or even similar ways.

Enhancing BUILD Scholars' Science Identity

The current analyses seek to extend prior research on the development of science identity among college students and provide an initial evaluation of the efficacy of targeted interventions aimed at promoting diversity in science majors and the scientific workforce. We focus explicitly on whether an intervention that incorporates multiple strategies to expose students to research, connects students with mentors, and provides financial support to offset college expenses can effectively promote the development of science identity during the first year of college. Prior studies have more often relied on samples of juniors or seniors to understand how a concept like science identity changes throughout students' college career; however, this study considers the experiences of first-year students to determine whether administrators, practitioners, and faculty can begin shifting the extent to which students conceive of themselves as scientists much earlier in college. Given that much of the attrition in STEM majors occurs during the first year, understanding effective strategies that enable and encourage students to begin thinking of themselves as aspiring scientists can equip policymakers and administrators with critical information to make more data-informed decisions regarding STEM retention efforts.

Distinct from previous research (e.g., Eagan et al., 2013), this study also examines a specific intervention program that incorporates several different initiatives designed to enhance the success of underrepresented students in their STEM majors. As we describe in the methods section, the data analyzed in this study originated from four sites that provide opportunities for first-year students to engage in research, connect with faculty mentors, and receive a

stipend designed to enable students to spend more time focused on their academics by worrying less about finding a job or working to pay for college. Each of the four sites has implemented each of these elements in ways that speak to the local needs and concerns of the student population, but the broad strategies at each site come from the same set of requirements. In the following section, we describe our data source and sample, the measures of interest, and our analytic approach.

Methods

Research Design Overview

Following a quasi-experimental design, this study analyzes longitudinal data collected from first-year students at four institutions to determine whether participants in a federally funded initiative (BUILD) have significantly stronger science identities compared to their peers in a control group that did not have access to the full slate of activities provided as part of the initiative. Prior studies that have examined the efficacy of undergraduate research interventions have risked mis-estimating the effects of receiving treatment (participating in the intervention) by not accounting for the nonrandom assignment of individuals to the treatment and control conditions (Lopatto, 2004; Maton & Hrabowski, 2004; Laursen et al., 2010). We overcome this challenge in this study by first estimating students' propensity of participating in the suite of intervention activities. We then derive inverse probability weights from these propensity scores and apply them to the sample before analyzing the data using multiple linear regression to examine the association between participation as a BUILD Scholar and first-year science identity, controlling for background characteristics, pre-college preparation, and a targeted set of first-year college experiences. The following sections provide relevant details regarding our methodological approach.

Data Source and Sample

This study draws from longitudinal survey data collected between the fall of 2016 and spring of 2020 from four sites funded by the NIH as part of its BUILD program. The Higher Education Research Institute's (HERI) Freshman Survey collected baseline data on incoming first-time, first-year students as they started college in the fall terms of 2016, 2017, 2018, and 2019. Students completed the Freshman Survey prior to their exposure to the BUILD program, which allowed for estimation of students' baseline or pre-BUILD science identity. The survey collects data on students' demographic characteristics, pre-college academic preparation and extracurricular experiences, lifelong educational and career goals, personal views and values, considerations related to their choice of college, and expectations for college life. TFS response rates for these four campuses varied for each institution and ranged from 40% to 60%.

In the spring of students' first year of college, the Coordination and Evaluation Center invited students at the 10 BUILD institutions to complete the Student Annual Follow-up Survey (SAFS). This instrument explores the frequency and nature of any research opportunities, academic and extracurricular experiences, and interactions with mentors that students may have had during college. Importantly, the survey asks many of the same items related to students' personal views, goals, and values that appear on TFS, which

provides opportunities to track longitudinal change. Students who consent to participating in the broader evaluation by completing the Freshman Survey or by participating in a BUILD-sponsored activity while in college receive invitations to complete the SAFS annually, allowing researchers to document students' development across many relevant domains. Our analyses described in this article focus exclusively on the experiences of first-year students at the four campuses that targeted BUILD activities to freshmen. The analytic sample includes cases that had data at both timepoints of college entry and the end of the first year. In addition to the two surveys, our analyses incorporate administrative data provided by each institution indicating whether participants in the sample were designated as BUILD Scholars by the site.

The initial sample included longitudinal responses from 2,281 first-year students enrolled across the four sites. The final analytic sample included 139 BUILD Scholars and 1,737 subjects in the control condition (non-Scholars) for a total of 1,876 cases. Most of the reduction from the initial to the final analytic sample occurred after enforcing the tenet of common support, which assumes overlap between the treatment and control conditions in the distribution of propensity scores (Guo & Fraser, 2010).

Variables

Our analyses examine two sets of variables. The first set of variables informed the logistic regression model predicting students' propensity to be designated as a BUILD Scholar. The second set of variables was included in the multiple regression analysis that aimed to explain variation in first-year science identity. All measures included in the BUILD Scholar propensity model were subsequently included in the science identity multiple regression model.

Dependent variable.—We operationalize students' science identity at the end of the first year of college using four agreement items appearing on the SAFS: *I have a strong sense of belonging to a community of scientists; I derive great personal satisfaction from working on a team that is doing important research; I think of myself as a scientist*; and *I feel like I belong in the field of science* (Estrada et al., 2011). These items were on a five-point scale ranging from "strongly disagree" to "strongly agree." HERI established the validity of this latent construct through the application of Item Response Theory (for details about the application of IRT to latent measures from surveys, see Sharkness, DeAngelo, & Pryor, 2010). HERI used the national sample of incoming first-year students from the 2016 Freshman Survey to establish population parameters associated with the science identity construct. These parameters were then applied to subsequent cohorts (i.e., TFS 2017–2019) and held constant for estimation of students' science identity scores at the end of their freshman year; therefore, any changes between the pre-test and post-test scores are due to actual changes in how students responded to the survey items. We also used confirmatory factor analysis (CFA) to further validate the latent properties of science identity and used reliability estimates to confirm the internal consistency of the items. Appendix A provides both the IRT-estimated parameters, which we used to create a construct score, and the factor loadings from CFA. The Cronbach's alphas for the items were 0.82 at baseline and 0.84 at

the end of the first year of college, suggesting strong internal consistency. The pre-test and outcomes were standardized with a mean of 50 and standard deviation of 10.

Key measure of interest/treatment variable.—We created a dummy variable to operationalize our key measure of interest pertaining to students' participation as BUILD Scholars at their institution during the first year of college. This indicator served as the outcome measure in the logistic regression predicting BUILD Scholar participation and as the key predictor of interest in the multiple linear regression model predicting science identity. The four sites analyzed in this study had explicit entry points into their BUILD projects for first-year students. Each site conducted its own recruitment and selection process to form each local cohort of BUILD Scholars; however, the sites were fairly consistent with respect to the minimum criteria students needed to meet in order to participate as a BUILD Scholar. These criteria included enrolling in college full-time, majoring in a biomedical or STEM-related field, having previously earned a minimum GPA of 2.75 or 3.0 in high school, aspirations to pursue graduate work in a biomedical or closely related field, and demonstrated interest in conducting biomedical research. BUILD Scholars received coverage for tuition and fees, a monthly stipend of roughly $1,100, and reimbursements for expenses related to travel to disseminate research findings. Each project required Scholars (1) to participate in peer learning communities, (2) enroll in newly developed courses that emphasized innovation, (3) engage in mentored research experiences, and (4) participate in various career advancement and development workshops (Hurtado et al., 2017). Although members of the control group may have had access to some of these activities, BUILD Scholars represent a distinct set of individuals who received all of these services, financial supports, and opportunities.

Demographic characteristics.—Various background characteristics based on participant responses in the TFS were entered into the model including race, gender, major, first-generation college student status, and financial aid received. BUILD projects collectively aim to identify, test, and implement effective strategies designed to contribute diversity to the scientific workforce, and the literature suggests there are differential experiences along race and gender identities with underrepresented populations facing barriers towards developing their science identity (Eagan et al., 2013). Most characteristics are entered as dummy variables; however, we used effect coding (see Mayhew & Simonoff, 2015) to account for differences by race/ethnicity.

To account for possible socioeconomic differences in the selection of students and the constraints financial concerns may place on some students' ability to focus on their science identity development, we included in the BUILD participation logistic regression an ordinal measure related to students' level of concern about their ability to finance their college education. This variable appeared in both the selection model and the science identity prediction model. Also, for both models, we used a dummy variable to represent whether students were first-generation to attend college, and we considered a student to be first-generation if neither parent had attended college. Our science identity regression also included an indicator as to whether students were Pell Grant recipients.

Likewise, in both the BUILD selection and science identity prediction models, we controlled for intended major at college entry using two dummy variables to represent three classifications: natural sciences (including traditional STEM fields), social sciences (e.g., anthropology, sociology, social work, psychology), and non-science fields (e.g., education, business, arts, humanities).

Competence and confidence.—To account for the possibility that sites might prefer to offer slots as BUILD Scholars to higher achieving students and the connection between academic achievement and developing a stronger science identity (Lu, 2015; Stets et al, 2017), the propensity score model and the model predicting science identity include high school GPA as a covariate. Likewise, as mentioned in the section describing the dependent variable, our models included a direct pre-test of the outcome. Finally, we included a measure of students' self-confidence in their ability to perform various science-related tasks, including developing an answerable research question and conducting an experiment. The full construct has 10 items, and we analyzed the items using IRT to create a unidimensional latent measure. We only included the self-efficacy measure to the science identity model, as this measure did not significantly differentiate BUILD Scholars from their peers in the control group in predicting whether students were BUILD Scholars during the first year.

Goals and aspirations.—One primary critique of quasi-experimental design relates to the role of personal ambition and self-selection bias in an individual's decision to pursue a science career opportunity. To overcome this liability, our BUILD selection and science identity models accounted for students' degree aspirations and career plans. Specifically, TFS respondents reported the highest degree they intended to earn in life, and we collapsed their responses into three dummy variables representing four categories: aspirations to earn a bachelor's degree or less (reference group), master's degree, medical doctorate or equivalent, and Ph.D. or other professional doctorate (e.g., Ed.D.). Additionally, in both models we included a TFS item that asked students to rate the likelihood that they would pursue a biomedical research career. We also included fixed effects terms to represent students' institutional affiliation and the year they started college to account for possible institutional and cohort differences in the likelihood of being selected as a BUILD Scholar and end-of-first-year science identity.

Additionally, several related measures were included only in the linear regression analyzing variation in science identity. In an attempt to avoid over-stating the effect that participating as a BUILD Scholar in the first year of college has on students' development of science identity, we added dummy variables representing whether students reported on the SAFS that they had conducted hands-on research, received mentorship from one or more mentors, and whether students reported participating in summer workshops designed to improve their skills as a mentor or mentee. These measures are included in nested regression models to account for the fact that students in the control condition, in particular, may have discovered research opportunities or identified mentors on their own despite not receiving the broad services and support provided to Scholars by the BUILD projects. By examining how the introduction of these terms into the model affect the estimated parameter associated with

being a BUILD Scholar, we aimed to better understand the specific components of the BUILD Scholar experience that contribute to science identity development.

Analyses

We incorporated three primary analytic tools to understand the characteristics of our sample, determine the parameters associated with assignment into the treatment or control conditions, and predict end-of-first-year science identity. First, we analyzed the dataset using descriptive statistics including crosstabulations and measures of central tendency. The crosstabulations in particular provided initial context for how students in each of our two conditions (BUILD Scholars and those who were not Scholars) differed on the covariates described in the previous section.

Next, we followed a two-step analytic approach to account for possible selection bias between BUILD Scholars and their peers in the control group given the nonrandom assignment of each student to one of the two conditions. Given the benefits of becoming a BUILD Scholar, the requirements of Scholars to participate in a variety of activities, and the minimum selection criteria, we expected to find qualitative differences between Scholars and the control group. Any causal inferences regarding the effect of becoming a BUILD Scholar on end-of-first-year science identity would be more credible by accounting for the endogeneity of the data (Desjardins et al., 2002).

Due to the observational, ex post facto nature of the data, we followed the counterfactual framework described by Rosenbaum and Rubin (1983, 1984, 1985). In this case, we attempted to estimate the "potential outcome, or the state of affairs that would have happened in the absence of the cause" (Guo & Fraser, 2010, p. 24). In other words, we attempted to understand the likely outcome for BUILD Scholars had they not become BUILD Scholars while also attempting to estimate how science identities among members of the control group may have been different had they had the opportunity to become a BUILD Scholar in their first year of college. The application of this framework to this study requires the estimation of a propensity score that corresponds to subjects' propensity of participating as a BUILD Scholar in the first year of college, using these propensity scores to develop inverse probability weights, and applying those weights to the sample in subsequent multivariate analyses so that the observed covariates between subjects in the treated and control conditions are equivalent and balanced (Schneider et al., 2007).

We provide the parameter estimates from the logistic regression predicting participation in BUILD as a Scholar in Appendix B. Propensity scores generated from the logistic regression model ranged from 0.008% to 73.881% for students who were not designated as BUILD Scholars, and propensity scores for BUILD Scholars ranged from 0.080% to 69.020%. Following the tenet of common support, we removed just over 300 cases from the control group that had propensity scores below 0.025% to account for the insufficient overlap between the two samples in their probability of becoming a BUILD Scholar. Next, we developed inverse probability weights using the propensity scores (Eagan et al., 2013; Hirano & Umbens, 2001; Nichols, 2007, 2008). We followed the approach described by Eagan et al. (2013) and Guo and Fraser (2010) to develop a weight representing the average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average

treatment effect on the untreated (ATU). ATE can be interpreted as the overall effect of the BUILD Scholar experience on a sample of eligible first-year college students. By contrast, the ATT suggests the effect of becoming a BUILD Scholar among research subjects with high probabilities of receiving the treatment. The ATU estimates the treatment effect among BUILD Scholars who have a lower probability of seeking out or accessing the initiative.

Having calculated the weights described above, we proceeded to analyze the weighted dataset by running a series of nested multiple linear regression models to examine the stability of the parameter estimate associated with BUILD participation. We staged the model by organizing variables into conceptually related, temporally aligned blocks and entered the blocks one at a time. With the weights applied and with controls for alternative explanations included in the model, we have greater confidence that any significant relationship between BUILD Scholar participation and science identity suggests a likely causal relationship between the two measures. As a final check, we examined the sensitivity of our model adjusted for students' propensity to participate in BUILD against a more straightforward (unadjusted) regression model with the same robust set of covariates, as some studies have suggested such strategies may be equally as effective at reducing selection bias for ex post facto treatment studies (Eagan et al., 2013; Shadish et al., 2008).

Limitations

Before presenting the results of our analyses, we wanted to acknowledge several limitations of the study. First, HERI's Freshman Survey offered the research team with a comprehensive instrument to collect baseline data; however, the survey was not originally designed with this study's purpose in mind. As a result of analyzing secondary data, we were limited with respect to the pre-college covariates that may have predicted students' decision to pursue and subsequently accept an offer to become a BUILD Scholar. For example, an indicator about any pre-college research experience may have produced slightly different estimates predicting students' participation as BUILD Scholars in their first year of college.

Additionally, quasi-experimental designs are only as good as the set of variables used to reduce bias. Although we felt we sufficiently accounted for much of the bias between the treatment and control conditions, there could be unobserved or unmeasured influences that could have been determinative in students' decision about participating as a BUILD Scholar.

Our analyses consider the treatment of being a BUILD Scholar to be uniform across all sites. With currently available data, we were unable to estimate dosage of the overall treatment or consider the intensity of various elements that likely varied by site. This limitation is not unique to quasi-experimental studies analyzing survey data or multi-site interventions, and future research may consider applying a similar approach to the one used in this study to each site as more data become available. Likewise, our matching approach incorporated site-level fixed effects rather than estimating propensity scores stratified by site. Our analyses suggested better matches could be achieved by pooling the data across sites. Additionally, the goal of the study is to understand the broader effectiveness of BUILD as an intervention rather than to examine specific site-level differences.

Finally, we understand that the lack of a randomized controlled trial constrains our ability to establish direct causal effects of participating as a BUILD Scholar. We believe our careful approach significantly reduced observed biases between treated cases and cases in the control group. Future studies may consider single-site investigations that randomly assign subjects into the treatment/intervention or the control group.

Results

Table 1 demonstrates the extent of bias reduction in the control group pre- and post-weighting and the extent that sample balance was achieved after applying the propensity score-based weight to the data. After applying the weights to the sample, we noticed substantial reductions in the bias between the treatment and control groups on a number of covariates. Most importantly, with the weights applied, the treatment and control groups have nearly identical values or distributions on the science identity pre-test, average high school grades, racial compositions, gender representation, degree aspirations, intended major, and financial concerns. The data in Table 1 suggest that women outnumber men in the sample by a 2-to-1 ratio. Scholars were more likely to describe their race as Black (29.5%), and one in five Scholars described their racial background as multiracial. On average, Scholars earned high school GPAs in the B+ to A- range with nearly 75% reporting having earned an A- or better GPA in high school. Nearly half of all BUILD Scholars aspired to earn a medical doctorate with more than one-third planning to pursue a Ph.D. or professional doctorate. About 10% of BUILD Scholars reported having major concerns about their ability to finance their college education while more than half indicated having some concerns. More than a third of BUILD Scholars reported having no concerns about financing college. Because some students may have been offered a slot as a BUILD Scholar prior to taking the Freshman Survey, it is possible that some respondents were already aware of their status as a BUILD Scholar prior to taking the Freshman Survey, which may have affected their response to this item regarding concerns about financing college.

Prior to applying weights, we found substantial differences in both the pre-test and post-test for science identity between BUILD Scholars and their counterparts who did not have the designation of being a BUILD Scholar during the first year of college. In Table 1 we note that BUILD Scholars (mean science identity = 61.33, SD=6.43) entered college scoring more than a full standard deviation above the mean on the science identity trait (note: all latent constructs were scored with a theoretical mean of 50 and a standard deviation of 10). BUILD Scholars started college scoring roughly one-half of a standard deviation higher than the mean calculated for the unweighted control group (mean science identity = 56.29, SD=7.96). Once we applied the weight, the treatment and control groups no longer significantly differed on incoming science identity. When we examined results of t-tests for science identity at the end of the first year of college, BUILD Scholars scored significantly higher (mean=60.68, SD=6.37) than the control group without the weight applied (mean=55.42, SD=7.07, t=13.73***) and even with the weight applied (mean=58.17, SD=6.97, t=6.73***).

We also found that BUILD Scholars started college with stronger intentions of pursuing a biomedical research career than their peers in the control group, as roughly 70% of BUILD

Scholars indicated they "probably" or "definitely" would pursue such a career compared to 51.0% of the unweighted responses in the control group. Likewise, BUILD Scholars started college with higher educational goals with 83.7% of Scholars expressing aspirations to earn either a medical degree (45.5%) or a Ph.D. or professional doctorate (37.4%). Comparatively, just over half (51.0%) of the respondents in the unweighted control group shared these aspirations.

Table 2 shows various stages of the multiple linear regression model with the ATE weight applied. The table includes unstandardized (b) and standardized (β) regression coefficients. Standardized regression coefficients can be interpreted similar to effect sizes, as they represent the expected change in the outcome (in standard deviation units) associated with a one-standard deviation change in the input variable. With the ATE weight applied, we interpret the unstandardized coefficient associated with being a BUILD Scholar as the average effect on first-year science identity for first-year students meeting the minimum eligibility to be selected into BUILD, controlling for other measures in the model at that stage. The unstandardized coefficient for BUILD Scholars was 2.67 when the indicator was first introduced in the model with the science identity pre-test as the only predictors in the model (Model 1); this parameter suggests that BUILD Scholars developed significantly stronger science identities in the first year of college – by about a quarter of a standard deviation – relative to their counterparts in the weighted control group.

The benefits derived from being a BUILD Scholar remained consistent after adding a number of covariates in Model 2; however, the parameter estimate associated with the BUILD Scholar flag experienced a substantial erosion after accounting for three types of experiences students may have encountered during their first year. The advantage associated with BUILD Scholars relative to science identity was essentially halved (b=1.38, p<0.001), about an eighth of a standard deviation, once we accounted for whether students in the sample had conducted hands-on research, participated in a summer training workshop for mentors and mentees, and worked with a mentor during their first year. Upon further examination, we discovered that controlling for whether respondents had worked with a mentor during their first year of college accounted for most of the attenuation in the effect of being a BUILD Scholar on science identity. Another way of explaining this change is that part of the reason BUILD Scholars developed stronger science identities during their first year of college could be attributed to the fact that many BUILD Scholars worked directly with one or more mentors, and mentorship positively contributed to students' science identity development.

In addition to the persistent benefits associated with becoming a BUILD Scholar, several other covariates emerged as significant predictors of first-year students' science identity. The pre-test for science identity had the strongest predictive power (β=0.41, p<0.001) of all of the variables in the final model, as students with stronger science identities at the start of college tended to also have stronger science identities by the end of the first year. Asian students tended to report significantly higher scores on science identity compared to the average student in the sample whereas Native American and multiracial students had significantly lower science identity scores compared to the average student in the sample.

Women reported significantly weaker science identities compared to men. This gender gap, however, was relatively modest compared to other variables in the model.

Rounding out demographics of the sample, the model included several terms related to students' socioeconomic status. First-generation students showed no significant difference in science identity relative to their peers who had at least one parent who had attended at least some college. Similarly, Pell Grant recipients were not any different from those who did not receive a Pell Grant in terms of their science identity. By contrast, students who expressed more concerns about their ability to finance their college education at the start of their freshman year had significantly weaker science identities by the end of the first year ($\beta=-0.05$, $p<0.01$).

Considerations related to education and career objectives demonstrated significant differences in science identity by academic major, degree aspirations, and intentions to pursue biomedical research careers. Perhaps not surprisingly, natural and social science majors tended to finish their first year with significantly stronger science identities compared to non-science majors. Degree aspirations provided more mixed results. Students aspiring to earn a master's degree as their highest degree in life did not finish the first year of college with a science identity that differed significantly from their counterparts with plans for a bachelor's degree or less. By contrast, students expecting to earn either a medical doctoral ($\beta=0.15$) or Ph.D. or other professional doctorate ($\beta=0.14$) had significantly stronger science identities by the end of the first year of college compared to their peers anticipating that the bachelor's degree would be their highest educational degree. Degree aspirations were measured as students began college, and the indicators associated with earning a Ph.D. or an M.D. suggested larger gaps in science identity between those with and those without these aspirations, compared to the gap suggested by the BUILD Scholar parameter (β-med=0.15, β-phd=0.14, β-build=0.10).

Beyond goals and aspirations, measures of competence and confidence were significantly related to science identity. Although high school grades were not determinative in predicting BUILD Scholar participation, our models suggested that high school grades significantly and positively correlated with science identity at the end of the first year ($\beta=0.07$). Likewise, students who started college with stronger science self-efficacy tended to finish the first year feeling significantly more connected to science compared to their peers with lower levels of science self-efficacy.

Finally, only one of the three first-year experiences significantly correlated with students' end-of-first-year science identity. As mentioned above, students who reported having worked closely with a mentor during their first year tended to develop significantly stronger science identities by the end of the year - about a quarter of a standard deviation higher than their peers who did not have the opportunity to work with mentors. There were no significant differences in science identity based on whether students reported having conducted hands-on research or participated in workshop trainings for mentors and mentees.

The overall model accounted for 49.5% of the variance in science identity at the end of the first year of college. The BUILD Scholar indicator and the science identity pretest

accounted for 7.4% and 27.2% of the variance, respectively, in end-of-first-year science identity. The introduction of measures related to race/ethnicity explained another 3% of the variance with the fixed effects of cohort and site adding another 1.8% of explained variance. Adding students' majors and socioeconomic measures explained another 1.1%, with degree aspirations, pre-college competence and confidence, and career plans adding another 4.9%. Finally, the three first-year experiences added 4.5% of explained variance to the model.

We provide the results of the final regression model without the propensity-score-adjusted weight in Appendix C. This model accounted for 45.3% of the variance in first-year science identity. As shown in that table, estimates for the effect of participating in BUILD as a Scholar on end-of-first-year science identity are quite similar to the estimates in the weighed models.

As described in the methods section, in addition to the ATE weight, we also calculated weights for the average treatment effect on the treated (ATT) and the average treatment effect on the untreated (ATU). When we applied the ATT weight to the data and reran the linear regression model, we observed similar coefficients for many of the covariates. The BUILD Scholar indicator measure entered in the first model with an unstandardized coefficient of 2.43 (p<0.001), suggesting nearly one-quarter of a standard deviation gap between BUILD Scholars and the control group for science identity. This coefficient remained remarkably stable in each subsequent model; however, the effect of being a BUILD Scholar lost its significance in the final model when we introduced whether students worked closely with a mentor during their first year of college (b=1.27, SE=0.71, p<0.07), suggesting that the BUILD Scholar experience may not significantly contribute to science identity among students who are most likely to be selected by sites for this opportunity. With the data weighted by average treatment effect on the untreated (ATU) weight applied to the data, the initial gap between BUILD scholars and the control had an unstandardized coefficient of 3.90 (p<0.001), or roughly four-tenths of a standard deviation difference. In the final model, the ATU for BUILD Scholars was 1.62 (p<0.001), suggesting the program provided significant advantages for treated participants' science identity at the end of the first year of college relative to the control group. We discuss the implications of these findings in the next section.

Finally, our tests for interaction terms between BUILD Scholar participation and students' racial and gender identities did not find any significant conditional effects. Specifically, the parameters associated with interaction terms between race/ethnicity and BUILD participation and between gender identity and BUILD participation failed to reach statistical significance. Additionally, the overall model did not significantly improve when we added these terms. We therefore conclude that, at least for this particular sample, we do not have evidence that the benefits of participating in BUILD differentially impact students based on their racial/ethnic or gender identities.

Discussion

This study intended to describe the characteristics associated with research subjects' likelihood of participating as BUILD Scholars during their first year of college and the

extent to which BUILD Scholars developed significantly stronger science identities during that first year compared to peers who did not become BUILD Scholars. To account for the nonrandom assignment of individuals into the treatment (Scholar) and control (non-Scholar) conditions, we followed the counterfactual, quasi-experimental framework posited by Rosenbaum and Rubin (1983, 1984, 1985) and Guo and Fraser (2010). Our findings suggest that students of color are more likely to become BUILD Scholars in the first year of college than white students, and students who have stronger science identities at college entry have greater likelihoods of becoming BUILD Scholars compared to students who enter college with weaker connections to science. This finding holds for the four sites (out of 10) that offer first-year interventions as part of the local BUILD project. Our findings echo previous research that suggests an interest in science is important for cultivating science identity (Hazari et al., 2010). Future research needs to interrogate this finding further to better understand the extent to which students' science identity or enthusiasm to conduct research or address some health problem facing their community gets conveyed to application screeners. Additionally, learning more about how screeners interpret and translate that enthusiasm into their ratings of applicants would provide additional insight as to some of the implicit or subtle signals that students may convey when seeking out opportunities such as the one offered by becoming a BUILD Scholar.

Students with more substantial concerns about financing their college education participated in BUILD at significantly lower rates. Considering the findings from the final regression model suggesting that BUILD Scholars derive significant benefits with respect to science identity even after controlling for financial concerns, these types of programs may be even more impactful and make more meaningful, lasting contributions toward diversifying the scientific workforce if financial need was made an essential consideration in selecting applicants for the opportunity, especially when the opportunity has the level of financial, academic, personal, and emotional support as many of the BUILD projects.

Turning to the findings related to end-of-first-year science identity, it is encouraging that BUILD Scholars develop significantly stronger science identities in their first year of college compared to their counterparts in the control group. Final models with the ATE and ATU weights applied suggested that BUILD Scholars had a significant advantage in developing stronger science identities in the first year of college relative to their peers in the control group. The significance of the ATU model suggests that the BUILD initiatives at the four sites in this study may provide greater benefits related to changes in science identity for students less likely to seek out or gain access to the intervention.

By contrast, the lack of significance of the ATT model indicates that the intervention may not be as effective for students who are most likely to seek out or gain access to the program, as the ATT effect estimates the effectiveness of treatment among individuals with higher probabilities of receiving the treatment. In other words, this estimate compares science identity outcomes among treated and controlled cases that were quite motivated (represented by degree and career aspirations), already strongly identified with science at college entry, and less concerned with whether they could finance their college education. Collectively, these models suggest the program may have opportunities to give greater consideration to

recruitment efforts targeted to and applications submitted by students who may enter college with greater financial concerns and more modest science identities.

Findings from the model predicting science identity suggest some differences by race and gender that persist even after controlling for first-year experiences and BUILD Scholar participation. Native American and Multiracial students had lower science identity scores than the average students in the sample at the end of the first year of college. Similarly, women's science identity was lower than men. These findings affirm previous studies that point to differences in science identity across race and gender (Carlone & Johnson, 2007; Hazari et al., 2010; Hazari et al., 2013) and require further examination. Although the model also included fixed effects, there are likely relevant climate issues not included in the model that may help to explain the gaps in science identity present between women and men and between white students and those from Native American, Asian, and multiracial backgrounds.

Finally, the critical role of mentors in developing students' science identity cannot be overstated. Prior research has found that effective mentorship can enhance science identity among undergraduate STEM majors (Byars-Winston & Rogers, 2019; Estrada et al., 2018; Piatt et al., 2019; Robnett et al., 2018). Outside of racial/ethnic identity and the science identity pretest, the mentorship variable emerged as the strongest predictor of science identity in the final model. This mentorship variable represents a crude way to operationalize the recognition component of Carlone and Johnson's (2007) science identity framework, but we believe there is much more that future research can unpack from this one indicator measure. First-year students who worked with mentors tended to be more likely to report research opportunities and participation in summer training workshops related to mentoring, which partially explains the lack of significance of those measures in predicting science identity.

Importantly, although mentorship is a critical component of the BUILD Scholar experience, our findings raise the question as to whether BUILD sites could do even more with respect to connecting students - both Scholars and those not designated as Scholars – to mentors. It is clear that many students outside the Scholar program successfully identified and began working with mentors during their first year. With respect to institutionalization of useful and effective strategies derived from the BUILD initiatives, sites might explore low-cost initiatives that facilitate early and sustained connections between faculty and students to provide more first-year students with enhanced opportunities to find faculty mentors.

Conclusion

For several decades, federal agencies and private foundations have generously funded programs designed to cultivate and facilitate the success of individuals from underrepresented groups. These efforts have undoubtedly contributed to personal success for the vast majority of the individuals who participated in these initiatives and directly benefited from their involvement. The evidence is less clear regarding the extent to which the effectiveness of the strategies implemented by individual campuses have been broadly tested and implemented, which has curtailed the collective efforts of a number of federal

agencies and private philanthropic foundations over the past several decades to diversify the scientific workforce.

This study suggests that NIH's BUILD Scholar initiative is effective at promoting science identity among first-year students who meet local sites' minimum eligibility requirements. Our findings suggest that these strategies, once scaled at other higher education institutions, may have a more pronounced effect if institutions prioritize applicants who do not fit an imagined "ideal" participant – one who comes to college having had tremendous academic success prior to college, generally does not have concerns about financing college, and already has plans to pursue either a medical doctorate or a Ph.D./other professional doctorate. Our findings suggest that many of the current BUILD Scholars may have been just as successful with respect to science identity had they not encountered the BUILD program. By contrast, when we broaden our analyses to consider all students eligible to participate as BUILD Scholars at the four sites with first-year BUILD interventions, we see a significant, modest benefit by the end of the first year. Future research should consider what other aspects of the BUILD initiative are likely contributing to its success, as our findings make clear that faculty mentorship during the first year is a critical contributor to the program's, and Scholars', success.

Acknowledgments

Appendix A: CFA Factor Loadings and Item Response Theory Parameter Estimates for Science Identity

Item	IRT Population Parameters					CFA Results	
	A	**B1**	**B2**	**B3**	**B4**	**Loading T1**	**Loading T2**
I have a strong sense of belonging to a community of scientists	3.52	−0.72	−0.20	0.71	1.50	0.85	0.85
I think of myself as a scientist	5.54	−0.04	0.10	0.85	1.53	0.84	0.79
I feel like I belong in the field of science	4.43	−0.05	−0.11	0.47	1.06	0.76	0.79
I derive great personal satisfaction from working on a team that is doing important research	1.78	−1.82	−1.11	0.09	1.30	0.64	0.60

Appendix B: Parameter Estimates from Logistic Regression Predicting BUILD Scholar Status

	B	**S.E.**	**Sig.**	**Exp(B)**
Degree aspiration: Masters (ref. bachelor's or less)	1.19	0.64	0.06	3.30

	B	S.E.	Sig.	Exp(B)
Degree aspiration: Medical doctorate or equivalent (ref. bachelor's degree or less)	2.10	0.62	0.00	8.19
Ph.D. or professional doctorate (ref. bachelor's degree or less)	1.82	0.62	0.00	6.17
Gender: women (ref. men)	0.14	0.21	0.50	1.15
Pre-test: Science Identity score from the Freshman Survey	0.08	0.02	0.00	1.08
Race: Native American	0.95	1.05	0.07	2.59
Race: Asian	−0.72	0.29	0.01	0.49
Race: Black	0.61	0.27	0.03	1.84
Race: Latino	−0.15	0.37	0.67	0.86
Race: Other	0.49	0.54	0.36	1.63
Race: Multiracial	0.11	0.27	0.70	1.11
Race: White	−1.29	0.27	0.00	0.28
Average high school grades	0.04	0.08	0.63	1.04
Status: first-generation student (yes v. no)	0.14	0.22	0.54	1.15
Site: A (ref: Site D)	0.14	0.46	0.77	1.15
Site: B (ref: Site D)	0.54	0.23	0.02	1.71
Site: C (ref: Site D)	−1.08	0.40	0.01	0.34
Cohort: 1 (ref: Cohort 4)	1.01	0.28	0.00	2.74
Cohort: 2 (ref: Cohort 4)	0.37	0.31	0.24	1.45
Cohort: 3 (ref: Cohort 4)	0.22	0.26	0.40	1.25
Major: Natural sciences (ref. non-science majors)	1.46	0.61	0.02	4.31
Major: Social sciences (ref. non-science majors)	1.72	0.69	0.01	5.59
Level of concern about ability to pay for college	−0.43	0.14	0.00	0.65
Constant	−11.25	1.31	0.00	0.00

Appendix C: Unweighted Multiple Linear Regression Results Predicting Science Identity

Independent Variables	b	SE	β
Constant	23.85	1.40	
BUILD Scholar (ref.: control group)	1.29	0.43	0.05[*]
Pre-test: Science Identity	0.41	0.02	0.45[*]
Race: Black	0.53	0.55	0.04
Race: Asian	1.54	0.52	0.15[*]
Race: Native American	−3.65	1.74	−0.20
Race: Multiracial	0.37	0.52	0.03
Race: Latino	0.89	0.59	0.07
Race: Other	0.55	1.00	0.03
Race: Unknown	−0.68	2.38	−0.04
Race: White	0.44	0.49	0.04
Gender: Women (ref.: men)	−0.48	0.28	−0.03

Independent Variables	b	SE	β
First-Generation Student (y/n)	0.07	0.31	0.01
Site: A (ref.: site D)	0.49	0.50	0.02
Site: B (ref.: site D)	−0.25	0.33	−0.02
Site: C (ref.: site D)	0.73	0.49	0.04
Cohort 2 (ref.: cohort 1)	−0.55	0.38	−0.03
Cohort 3 (ref.: cohort 1)	0.33	0.44	0.02
Cohort 4 (ref.: cohort 1)	−1.18	0.32	−0.08 [*]
Major: Natural sciences (ref.: non-science major)	2.75	0.43	0.15 [*]
Major: Social sciences (ref.: non-science major)	1.56	0.59	0.06 [*]
Financial concerns	−0.23	0.20	−0.02
Pell Recipient (ref.: did not receive Pell grant)	0.19	0.23	−0.02
Biomedical science research career: Definitely no (ref.: Uncertain)	−0.44	0.50	−0.02
Biomedical science research career: Probably no (ref.: Uncertain)	−0.77	0.43	−0.04
Biomedical science research career: Probably yes (ref.: Uncertain)	0.34	0.38	0.02
Biomedical science research career: Definitely yes (ref.: Uncertain)	0.65	0.39	0.04
High school GPA	0.21	0.11	0.04
Science self-efficacy from TFS	0.06	0.01	0.08 [*]
Aspiration: Master's degree (ref: bachelor's degree aspiration)	0.11	0.42	0.01
Aspiration: Medical doctorate (ref: bachelor's degree aspiration)	1.97	0.45	0.12 [*]
Aspiration: Ph.D. or other Professional Doctorate (ref: bachelor's degree aspiration)	1.45	0.44	0.09 [*]
Aspiration: Missing (ref: bachelor's degree aspiration)	0.67	1.27	0.01
Summer training for mentors and mentees (y/n)	0.69	0.40	0.03
Conducted hands-on research (y/n)	0.80	0.34	0.04
Conducted hands-on research (missing)	1.23	0.57	0.04
Had at least one mentor during first year (y/n)	1.55	0.28	0.10 [*]
Had a least one mentor during first year (missing)	−0.34	0.62	−0.01 [*]

[*] indicates $p < 0.01$

References

Byars-Winston A, & Rogers JG (2019). Testing intersectionality of race/ethnicity × gender in a social–cognitive career theory model with science identity. Journal of Counseling Psychology, 66(1), 30. [PubMed: 30284846]

Carlone HB, & Johnson A (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. Journal of Research in Science Teaching, 44(8), 1187–1218.

Chang MJ, Eagan MK, Lin MH, & Hurtado S (2011). Considering the impact of racial stigmas and science identity: Persistence among biomedical and behavioral science aspirants. The Journal of Higher Education, 82(5), 564–596. [PubMed: 23226874]

DesJardins SL, McCall BP, Ahlburg DA, & Moye MJ (2002). Adding a timing light to the "tool box". Research in Higher Education, 43(1), 83–114. doi:10.1023/A:1013022201296

Eagan MK Jr, Hurtado S, Chang MJ, Garcia GA, Herrera FA, & Garibay JC (2013). Making a difference in science education: the impact of undergraduate research programs. American Educational Research Journal, 50(4), 683–713. [PubMed: 25190821]

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.

Estrada M, Woodcock A, Hernandez PR, & Schultz PW (2011). Toward a model of social influence that explains minority student integration into the scientific community. Journal of educational psychology, 103(1), 206. [PubMed: 21552374]

Ginther DK, Schaffer WT, Schnell J, Masimore B, Liu F, Halak LL, & Kington R (2011). Race, ethnicity, and research awards. Science, 333(6045), 1015–1019. [PubMed: 21852498]

Guo S, & Fraser MW (2010). Propensity score analysis: Statistical methods and applications. Thousand Oaks, CA: Sage.

Hazari Z, Sonnert G, Sadler PM, & Shanahan MC (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. Journal of research in science teaching, 47(8), 978–1003.

Hazari Z, Sadler PM, & Sonnert G (2013). The science identity of college students: Exploring the intersection of gender, race, and ethnicity. Journal of College Science Teaching, 42(5), 82–91.

Hernandez PR, Bloodhart B, Barnes RT, Adams AS, Clinton SM, Pollack I, Godfre E, Burt M, & Fischer EV (2017) Promoting professional identity, motivation, and persistence: Benefits of an informal mentoring program for female undergraduate students. PLoS ONE 12(11): 0187531. 10.1371/journal.pone.0187531

Hirano K, & Imbens GW (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. Health Services and Outcomes Research Methodology, 2(3–4), 259–278. doi:10.1023/A:1020371312283

Hurtado S, White-Lewis D, & Norris K (2017). Advancing inclusive science and systemic change: The convergence of national aims and institutional goals in implementing and assessing biomedical science training. BMC Proceedings, 11(Suppl. 12), 17. 10.1186/s12919-017-0086-5 [PubMed: 31851727]

Lane TB (2016). Research environments as counterspaces? Examining spaces that inhibit and support science identity development for black students in STEM. Urban Education Research & Policy Annuals, 4(1).

Laursen S, Seymour E, Hunter AB, Thiry H, & Melton G (2010). Undergraduate research in the sciences: Engaging students in real science. San Francisco, CA: Jossey-Bass.

Lopatto D (2004). Survey of Undergraduate Research Experiences (SURE): First findings. Cell Biology Education, 3(4), 270–277. doi:10.1187/cbe.04-07-0045 [PubMed: 15592600]

Lu C (2015). Finding los científicos within: Latino male science identity development in the first college semester. Journal of College Student Development, 56(7), 740–745.

Maton KI, & Hrabowski FA (2004). Increasing the number of African American PhDs in the sciences and engineering: A strengths-based approach. American Psychologist, 59(6), 547–556. doi:10.1037/0003-066X.59.6.547 [PubMed: 15367090]

Mayhew MJ, & Simonoff JS (2015). Non-white, no more: Effect coding as an alternative to dummy coding with implications for higher education researchers. Journal of College Student Development, 56(2), 170–175.

National Academies of Sciences, Engineering, and Medicine. (2007). Rising above the gathering storm: Energizing and employing America for a brighter future. Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2019). The science of effective mentorship in STEMM. Washington, DC: The National Academies Press. 10.17226/25568.

National Science Foundation. (2019). Higher education in science and engineering. Science and Engineering Indicators 2020. NSB-2019-7. Alexandria, VA. Available at https://ncses.nsf.gov/pubs/nsb20197/.

Nichols A (2007). Causal inference with observational data. Stata Journal, 7(4), 507–541.

Nichols A (2008). Erratum and discussion of propensity-score reweighting. Stata Journal, 8(4), 532–539.

Piatt E, Merolla D, Pringle E, & Serpe RT (2019). The role of science identity salience in graduate school enrollment for first-generation, low-income, underrepresented students. The Journal of Negro Education, 88(3), 269–280.

Robnett RD, Nelson PA, Zurbriggen EL, Crosby FJ, & Chemers MM (2018). Research mentoring and scientist identity: insights from undergraduates and their mentors. International journal of STEM education, 5(1), 1–14. [PubMed: 30631691]

Rodriguez S, Cunningham K, & Jordan A (2019). STEM identity development for Latinas: The role of self-and outside recognition. Journal of Hispanic Higher Education, 18(3), 254–272.

Rosenbaum PR, & Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41–55. doi:10.1093/biomet/70.1.41

Rosenbaum PR, & Rubin DB (1984). Reducing Bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 79(387), 516–524. doi:10.2307/2288398

Rosenbaum PR, & Rubin DB (1985). Constructing a comparison group using multivariate matched sampling methods that incorporate the propensity score. American Statistician, 39, 33–38. doi:10.2307/2683903

Schneider B, Carnoy M, Kilpatrick J, Schmidt WH, & Shavelson RJ (2007). Estimating causal effects: Using experimental and observational designs. Washington, DC: American Educational Research Association.

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. Science Education, 88(4), 493–534.

Shadish WR, Clark MH, & Steiner PM (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random to nonrandom assignment. Journal of the American Statistical Association, 103(484), 1334–1344.

Sharkness J, DeAngelo L, & Pryor J (2010). CIRP construct technical report. Retrieved from http://www.gseis.ucla.edu/heri/PDFs/technicalreport.pdf

Spell RM, Guinan JA, Miller KR, & Beck CW (2014). Redefining authentic research experiences in introductory biology laboratories and barriers to their implementation. CBE-LIFE Sciences Education, 13(1), 102–110. 10.1187/cbe.13-08-0169 [PubMed: 24591509]

Stets JE, Brenner PS, Burke PJ, & Serpe RT (2017). The science identity and entering a science occupation. Social science research, 64, 1–14. [PubMed: 28364837]

Sullivan Commission (2004). Missing persons: Minorities in the health professions. http://www.sullivancommission.org/.

Trujillo G, & Tanner KD (2014). Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. CBE—Life Sciences Education, 13(1), 6–15. [PubMed: 24591497]

White AM, DeCuir-Gunby JT, & Kim S (2019). A mixed methods exploration of the relationships between the racial identity, science identity, science self-efficacy, and science achievement of African American students at HBCUs. Contemporary Educational Psychology, 57, 54–71.

Table 1

Descriptive Statistics of the Sample before and After Weighting

	BUILD Scholars			Control Group - Pre-Weighting			Control Group - Post-Weighting			Bias Red.
	%	Mean	SD	%	Mean	SD	%	Mean	SD	
Race: Native American	0.7			0.4			0.6			0.67
Race: Asian	17.3			19.3			17.8			0.75
Race: Black	29.5			11.4			29.8			0.98
Race: Latino	10.1			15.0			9.3			0.84
Race: White	18.7			36.4			19.2			0.97
Race: Other	2.9			1.4			2.6			0.80
Race: Multiracial	20.9			16.1			20.7			0.96
Gender: Man (including trans*)	30.9			33.0			31.1			0.90
Gender: Woman (including trans*)	69.1			66.4			68.9			0.93
First-Generation in College	28.1			27.4			27.9			0.71
High school GPA: D	0.0			0.0			0.0			
High school GPA: C	0.0			0.5			0.4			
High school GPA: C+	0.7			0.7			1.3			0.01
High school GPA: B-	3.6			2.6			1.0			-1.60
High school GPA: B	7.9			10.7			9.1			0.57
High school GPA: B+	14.4			17.9			18.3			-0.11
High school GPA: A-	34.5			29.3			27.2			-0.40
High school GPA: A or A+	38.8			38.2			42.4			-5.00
Average High School GPA (8-point scale ranging from D (1) to A/A+ (8))		6.95	1.14		6.85	1.23		6.93	1.26	0.80
Pre-test: Science Identity		61.33	6.43		56.29	7.96		61.21	6.76	0.98
Cohort: 1	30.2			18.2			30.9			0.94
Cohort: 2	17.3			18.4			16.7			0.45
Cohort: 3	32.4			36.6			33.9			0.64
Cohort: 4	20.1			26.8			18.5			1.00
Site: A	4.3			9.4			5.0			0.86
Site: B	42.4			40.2			44.6			0.00
Site: C	9.4			20.8			9.0			0.96

Eagan et al. Page 24

	BUILD Scholars			Control Group - Pre-Weighting			Control Group - Post-Weighting			Bias Red.
	%	Mean	SD	%	Mean	SD	%	Mean	SD	
Site: D	43.9			29.6			41.4			0.83
Pursue science-related biomedical research career: Definitely No	7.9			10.7			9.6			0.39
Pursue science-related biomedical research career: Probably No	7.2			15.4			9.8			0.68
Pursue science-related biomedical research career: Uncertain	13.7			21.9			15.7			0.76
Pursue science-related biomedical research career: Probably yes	22.3			25			22.9			0.78
Pursue science-related biomedical research career: Definitely yes	48.9			27			42.1			0.69
Degree aspiration: Bachelor's or less	2.2			12			1.6			0.94
Degree aspiration: Master's degree	15.1			35.5			15.8			0.97
Degree aspiration: Ph.D. or professional doctorate	37.4			27.9			37.4			1.00
Degree aspiration: Medical degree	45.3			23.1			44.9			0.98
End-of-first-year Science Self-Efficacy		52.75	8.81		49.02	9.65		51.92	10.15	0.78
End-of first-year science identity		60.68	6.37		55.42	7.07		58.17	6.97	0.52
Financial concerns: None	36.0			24.3			32			0.66
Financial concerns: Some	52.5			60.4			55.2			0.66
Financial concerns: Major	10.8			14.9			10.4			0.90
Intended major: Natural sciences	90.6			80			89.1			0.86
Intended major: Social sciences	7.9			7.8			9.1			
Intended major: Non-science	1.4			12.2			1.9			

Table 2

Results of Linear Regression on Science Identity

Independent Variables	Model 1			Model 2			Model 3		
	b	SE	β	b	SE	β	b	SE	β
Constant	27.90	0.75		25.47	1.13		26.44	1.08	
BUILD Scholar (ref.: control group)	2.67	0.19	0.19*	2.45	0.19	0.18*	1.38	0.22	0.10*
Pre-test: Science Identity	0.49	0.01	0.52*	0.39	0.02	0.42*	0.38	0.02	0.41*
Race: Black				0.09	0.46	0.01	−0.11	0.44	−0.01
Race: Asian				2.24	0.44	0.24*	2.22	0.42	0.24*
Race: Native American				−4.51	0.80	−0.31*	−3.60	1.40	−0.25*
Race: Multiracial				−0.66	0.44	−0.07	−0.49	0.42	−0.05*
Race: Latino				0.98	0.50	0.09	0.16	0.48	0.02
Race: Other				1.15	0.80	0.08	1.61	0.77	0.12
Race: Unknown				−0.11	2.24	−0.01	−0.29	2.15	−0.02
Race: White				0.82	0.42	0.09	0.50	0.41	0.05
Gender: Women (ref.: men)				−0.50	0.21	−0.04*	−0.59	0.20	−0.04*
First-Generation Student (y/n)				0.24	0.23	0.02	0.37	0.22	0.02
Site: A (ref.: site D)				0.15	0.37	0.01	−0.17	0.36	−0.01
Site: B (ref.: site D)				−0.98	0.24	−0.07*	−0.23	0.23	−0.02
Site: C (ref.: site D)				1.75	0.37	0.10*	2.19	0.36	0.12*
Cohort 1 (ref.: cohort 4)				−0.40	0.27	−0.02	−0.14	0.26	−0.01
Cohort 2 (ref.: cohort 4)				0.84	0.34	0.05	1.17	0.33	0.07*
Cohort 3 (ref.: cohort 4)				0.03	0.23	−0.01	−0.38	0.23	−0.03
Major: Natural sciences (ref.: non-science major)				1.89	0.38	0.10*	2.30	0.36	0.12*
Major: Social sciences (ref.: non-science major)				1.32	0.47	0.06*	1.21	0.45	0.05*
Financial concerns				−0.50	0.15	−0.05*	−0.55	0.14	−0.05*
Pell Recipient (y/n)				−0.08	0.16	−0.01	−0.06	0.16	−0.01
Biomedical science research career: Definitely no (ref.: Uncertain)				−0.57	0.41	−0.02	−0.29	0.39	−0.01
Biomedical science research career: Probably no (ref.: Uncertain)				−1.94	0.35	−0.09*	−1.71	0.33	−0.08*

Independent Variables	Model 1			Model 2			Model 3		
	b	SE	β	b	SE	β	b	SE	β
Biomedical science research career: Probably yes (ref.: Uncertain)				-0.30	0.29	-0.02	-0.33	0.28	-0.02
Biomedical science research career: Definitely yes (ref.: Uncertain)				1.28	0.30	0.09*	1.21	0.29	0.08*
High school GPA				0.41	0.08	0.07*	0.36	0.08	0.07*
Science self-efficacy from TFS				0.05	0.01	0.06*	0.03	0.01	0.04*
Aspiration: Master's degree (ref: bachelor's degree aspiration)				0.64	0.35	0.04	0.34	0.34	0.02
Aspiration: Medical doctorate (ref: bachelor's degree aspiration)				2.74	0.37	0.17*	2.35	0.36	0.15*
Aspiration: Ph.D. or other Professional Doctorate (ref: bachelor's degree aspiration)				2.42	0.36	0.16*	2.15	0.35	0.14*
Aspiration: Missing (ref: bachelor's degree aspiration)				0.27	0.88	0.01	-0.91	0.84	-0.01
Summer training for mentors and mentees (y/n)							0.44	0.25	0.03
Conducted hands-on research (y/n)							0.33	0.23	0.02
Conducted hands-on research (missing)							-0.92	0.37	-0.03
Had at least one mentor during first year (y/n)							2.77	0.21	0.20*
Had a least one mentor during first year (missing)							-1.84	0.42	-0.06*

*
indicates p < 0.01