

Article

Longitudinal Trends in Science Self-Efficacy Among Undergraduate Students: Addressing Disparities in STEM Participation

Sungmin Moon ^{1,*}, Crist Khachikian ², Shu-Sha Angie Guan ³, Judith C. P. Lin ⁴, Jose H. Vargas ⁵, Patchareeya Kwan ⁶ and Gabriela Chavira ⁷

¹ Department of Teacher Education and Professional Development, Morgan State University, Baltimore, MD 21251, USA

² Department of Civil Engineering and Construction Management, California State University, Northridge, CA 91330, USA; crist.khachikian@csun.edu

³ Department of Child and Adolescent Development, California State University, Northridge, CA 91330, USA; angie.guan@csun.edu

⁴ Health Equity Research & Education Center, California State University, Northridge, CA 91330, USA; judith.lin@csun.edu

⁵ College of Education, California State University Dominguez Hills, Carson, CA 90747, USA; jvargas@csudh.edu

⁶ Department of Health Sciences, California State University, Northridge, CA 91330, USA; patty.kwan@csun.edu

⁷ Department of Psychology, California State University, Northridge, CA 91330, USA; gabriela.chavira@csun.edu

* Correspondence: sungmin.moon@morgan.edu

Abstract

This retrospective study examined group-level trends in science self-efficacy among undergraduate students from underrepresented racial/ethnic groups (URGs) over four years (2017–2020) at a public university in California. The study investigated the effectiveness of an undergraduate research experience program designed to increase URG participation in STEM fields. Using propensity score matching to control for selection bias, the final sample included 3146 students (344 intervention participants, 2802 non-participants) who completed the Student Annual Follow-up Survey. The Rasch measurement method was used to convert Likert scale responses into linear logit scores, which enabled group-level trend analysis using linking and equating techniques. A repeated measures analysis examined changes in science self-efficacy while controlling for first-generation status, URG status, and gender. Results revealed a significant overall increase in science self-efficacy across all four years. Intervention participants consistently demonstrated significantly higher science self-efficacy scores than non-participants. The interaction between time and intervention participation was statistically significant, indicating that intervention participants showed greater improvements over time. These findings support the effectiveness of culturally responsive undergraduate research experiences in fostering science self-efficacy among URG students. The study contributes to understanding how structured interventions can address persistent disparities in STEM education and career pathways.

Keywords: group-level trends; science self-efficacy; STEM education; underrepresented groups

Academic Editor: Jack F. Eichler

Received: 25 March 2026

Revised: 17 June 2026

Accepted: 22 June 2026

Published: 28 June 2026

Copyright: © 2026 by the authors.

Submitted for possible open access publication under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Recent decades have witnessed increasing national concern regarding the diversification of the science, technology, engineering, and mathematics (STEM) workforce in the United States. Despite substantial federal investments and policy initiatives aimed at broadening participation in STEM, persistent racial and gender disparities continue to limit equitable access to education and professional opportunities for historically underserved racial/ethnic groups (URGs) (National Science Board, 2022). Researchers have emphasized that structural inequities, limited access to culturally responsive mentorship, and discriminatory institutional practices continue to constrain the recruitment, retention, and persistence of URG students in STEM pathways (McGee, 2016; G. Saw et al., 2018). These disparities are particularly concerning because the development of a diverse STEM workforce is closely associated with scientific innovation, economic competitiveness, and social equity (Hofstra et al., 2020).

To address these inequities, undergraduate research experiences (UREs) have emerged as one of the most widely supported intervention strategies for improving STEM participation and persistence among URG students. Prior research suggests that participation in authentic research environments can strengthen students' science identity, increase motivation, improve academic engagement, and enhance science self-efficacy (Chemers et al., 2011; Estrada et al., 2018). Science self-efficacy, defined as an individual's belief in their ability to successfully perform science-related tasks and succeed in STEM environments, has been consistently identified as a strong predictor of STEM persistence and career aspirations (Bandura, 1997; Lent et al., 1994). Consequently, understanding how research training programs influence the development of science self-efficacy over time is essential for designing equitable and effective STEM education interventions. We first review the relevant literature on science self-efficacy development, then present findings from a group-level quasi-experimental trend study, and finally discuss their broader implications.

1.1. Disparities in STEM

The United States has long sought to maintain and advance its global scientific and technological leadership by broadening participation in STEM. Over the past two decades, federal initiatives such as the America COMPETES Reauthorization Act (Gonzales, 2015) and the Educate to Innovate campaign (Office of the Press Secretary, 2010) have aimed to strengthen STEM education, innovation, and workforce development. More recent legislative efforts, including the Women and Minorities in STEM Booster Act (2021), have similarly sought to increase the participation of women and students from URGs in STEM pathways. Despite these efforts, persistent race- and gender-based disparities remain in STEM education, recruitment, persistence, and employment.

These disparities are rooted in historical and structural racism, which continues to constrain academic, professional, and life opportunities for URGs (see McGee, 2016). For example, McGee's (2016) qualitative study of 38 high-achieving Black and Latino/a STEM college students showed how racially hostile academic spaces and racial microaggressions shaped students' STEM experiences and identity-related coping strategies. Prior research has also shown that students from underrepresented groups encounter multiple barriers that can push them away from STEM careers, including racial and gender-based discrimination, social isolation, limited access to culturally responsive mentorship, inequitable resources, implicit bias, lower faculty expectations, financial pressures, and curricula that lack relevance to students' cultural contexts (G. Saw et al., 2018). In a nationally representative longitudinal study using the High School Longitudinal Study (HSLs) of 2009, G. Saw et al. (2018) examined 20,242 first-time ninth-grade students and found disparities in

STEM career aspirations across gender, race/ethnicity, and socioeconomic status from early ninth grade to late eleventh grade. Geographic inequities further compound these challenges. Using the same nationally representative HSL:09 dataset, G. K. Saw and Agger (2021) examined rural, small-town, suburban, and urban students and found that rural and small-town students were less likely than their suburban peers to enroll in postsecondary STEM programs. Although federal funding and national initiatives have contributed to increased URG participation in STEM pathways, substantial inequities persist and require sustained, intensive efforts to address the enduring effects of racism and structural exclusion (see Carter et al., 2019; Hofstra et al., 2020; National Science Board, 2022). Hofstra et al. (2020), for example, analyzed approximately 1.2 million U.S. doctoral recipients from 1977 to 2015 and found that demographically underrepresented scientists produced novel contributions at high rates, yet their contributions were less likely to translate into equivalent career recognition.

Social scientific research plays a critical role in identifying mechanisms that can promote racial and gender parity in STEM education and career entry. In particular, greater attention is needed to how students from URGs develop and strengthen science self-efficacy, a key factor associated with persistence and success in STEM. This article presents a retrospective study designed to examine changes in science self-efficacy over time among undergraduate URGs who participated in a research experience program. Our research questions are as follows:

1. Does student science self-efficacy change over time?
2. Does student science self-efficacy differ between intervention participants and non-participants?
3. Do student science self-efficacy trajectories differ as a function of the intervention, first-generation, URG status, or gender?

1.2. Science Self-Efficacy as a Shaper of Students' Career Trajectories

According to Social Cognitive Theory (SCT), self-efficacy is an individual's belief in their ability to achieve goals—significantly shapes career choices and development (Bandura et al., 2001; Hackett & Betz, 1981; Lent & Brown, 2006). For instance, broader academic self-efficacy can predict specific occupational self-efficacy, such as science self-efficacy (Bandura et al., 2001). Notably, Bandura et al. (2001) found that self-efficacy, rather than actual academic achievement, is the primary predictor of academic aspirations over time. While these self-beliefs develop early, they remain malleable, influenced by socialization and sociodemographic factors. For example, Benden and Lauermann (2023) demonstrated that self-efficacy is a stronger predictor of academic aspirations and achievements than actual performance, particularly among students from lower socioeconomic backgrounds.

During college, students' academic motivation and self-concept may decline if their psychological, academic, and social needs are unmet within their educational environment (Chemers et al., 2001; Robbins et al., 2004; Tinto, 2012). While institutional support—such as smaller classes, enhanced student-faculty interactions, and targeted advising—can sustain motivation and self-efficacy (Robbins et al., 2004; Tinto, 2012), a clear understanding of these developmental patterns remains complicated by inconsistent sociodemographic analyses and varying self-efficacy measures (Chemers et al., 2001; Robbins et al., 2004).

At the college level, science self-efficacy trajectories are critical because they align with students' science comprehension and competencies. Higher self-efficacy predicts favorable academic and career outcomes, including first-year GPA, STEM career interest, and career readiness (Larose et al., 2006; Lent et al., 2016). These outcomes, however, are often inferred from academic achievements and career progress rather than being directly

self-reported. Furthermore, the development of science self-efficacy varies across domains; for instance, literature summarization skills may develop early, whereas data analysis skills often emerge later (Dalgety & Coll, 2006).

Vocational outcomes and self-efficacy trajectories also vary significantly by demographic. African American and Latinx students often report lower self-efficacy than their Caucasian peers due to structural barriers, such as institutional biases, resource limitations, and a lack of culturally responsive supports or role models. Gender disparities also persist; males initially report higher STEM self-efficacy and career aspirations, though these confidence gaps may level out by the end of the first college semester (Dalgety & Coll, 2006).

Addressing these disparities is crucial, as STEM occupations are projected to grow at twice the average rate over the next decade (Dubina et al., 2021). STEM graduates—including those from historically underrepresented groups like Latinx students—enjoy significantly higher earning than non-STEM peers (Furchtgott-Roth et al., 2009; Melguizo & Wolniak, 2012). Given higher education's role in driving social and economic mobility, understanding the science self-efficacy trajectories of diverse student populations is essential (Furchtgott-Roth et al., 2009; Haskins et al., 2009; Pew Research Center, 2014, 2016).

To address how training stages foster the competencies essential for biomedical research careers (Rivers et al., 2020), our intervention program provides a culturally responsive environment. Previous research shows this program effectively supports URGs in developing science self-efficacy and identity, thereby increasing science career intentions (Camacho et al., 2021; Lin et al., 2024; Vasquez-Salgado et al., 2023). Acting as a “counterspace,” the intervention offers community, support, and identity affirmation often lacking in broader institutional contexts. This counterspace provides proximal contextual support that strengthens career intentions, belonging, and persistence beyond the effects of individual cognitive factors like self-efficacy and academic self-concept (Lin et al., 2024).

While longitudinal studies often suffer from fluctuating participation and incomplete datasets, we address these gaps using a statistical modeling approach that links available data to estimate missing information. This robust analytical method allows us to accurately track developmental patterns despite missing data. In the current study, we leverage this modeling approach to examine group-level trends in science self-efficacy among undergraduate students from URGs over four years and assess the intervention's long-term impact.

2. Method

This retrospective study examines trajectories of student science self-efficacy over four years (2017–2020) at the local Building Infrastructure Leading to Diversity (BUILD) campus. The BUILD program recruited undergraduate students from diverse backgrounds in STEM disciplines to participate in a survey that explored the experiences of STEM undergraduates. The survey participants included both intervention and non-intervention students. They completed a Student Annual Follow-up Survey (SAFS), an online survey administered by national BUILD evaluators at the Coordination and Evaluation Center (CEC) housed at a research-intensive university in California.

2.1. Context

This study was conducted within the context of Building Infrastructure Leading to Diversity; Promoting Opportunities for Diversity in Education and Research (BUILD PODER). Established in 2014 at a public university in California, BUILD PODER is an undergraduate research training initiative funded by the National Institutes of Health (NIH) and managed by the National Institute of General Medical Sciences (NIGMS). This

program's primary objective is to increase the representation and retention of URGs in biomedical and STEM disciplines through structured URE.

A distinguishing feature of BUILD PODER is its integration of Critical Race Theory (CRT) as a guiding framework (Solórzano, 1997, 1998). This framework encourages students and faculty to critically examine systemic oppression, challenge conventional scientific and educational norms, advocate for social justice, and leverage lived experiences as assets in scientific inquiry.

The program supports students through a multi-phase intervention: (a) Summer Jumpstart (SJS), a preliminary four-week intensive training focusing on research skills, professional development, cultural competence, and cohort team-building; (b) Academic Year Support, where students conduct hands-on research under the guidance of trained faculty mentors, present findings at conferences, and attend biweekly group discussions on social justice, community-engaged research, and career pathways (with cohort cohesion further supported through regular social events); and (c) External Summer Research, in which students are encouraged to participate in summer-long research experiences at external doctoral-granting institutions to broaden their scientific skills and networks.

2.2. Participants

Over a four-year period, a total of 4131 students completed the survey at the local BUILD campus. A comparison group was identified using propensity score matching (PSM) to isolate the effects of the intervention from other confounding factors. The rationale behind PSM is to address selection bias, which arises when differences in outcomes may be due to unobserved factors influencing the likelihood of participating in the program rather than the program itself (Murnane & Willett, 2010). For instance, parents who encourage their children to participate in intervention programs may also be more engaged in their children's education, emphasizing the importance of reading or monitoring homework (Dumont et al., 2012; Fan & Chen, 2001). PSM helps control such selection bias, ensuring that the intervention group does not disproportionately include students with more engaged parents. In this study, four matching estimators (formal training or workshop, conducting one's own scientific research, honors and awards, and scientific conference attendance or presentation) were used to generate PSM scores with SPSS 25.0 (IBM Corp, 2017). After matching, 3146 students had similar propensity scores. Among these, 344 were participants in the intervention program, while 2802 were non-participants (see Table 1).

Table 1. Number of Participants in each Survey over time.

Survey	Before PSM			After PSM		
	Intervention	Non-Intervention	Total	Intervention	Non-Intervention	Total
SAFS 2017	92	294	386	78	173	251
SAFS 2018	62	621	683	46	368	414
SAFS 2019	87	1675	1762	72	1361	1433
SAFS 2020	194	1106	1300	148	900	1048
Total	435	3696	4131	344	2802	3146

To ensure the robustness of the PSM methodology, recent studies were reviewed. For example, recent advancements in propensity score methods are discussed in the literature on causal inference in observational studies (Kim & Steiner, 2015). Additionally, the use of SPSS for PSM has been validated in educational and psychological research, highlighting the importance of addressing selection bias in intervention studies (Thoemmes & Kim, 2011). However, only 28 participants completed all four surveys from SAFS17 to SAFS20, and even among them, not all responded to every survey item. The number of participants

was too small to draw meaningful conclusions based on their actual responses. Therefore, using the responses of participants from the most recent SAFS20 survey, we employed the “Linking and Equating” methods (Bond & Fox, 2013), which utilized survey item logit values and step logit values (i.e., the thresholds between adjacent response categories) to estimate how SAFS20 participants would have responded to the same survey items had they participated in SAFS17, SAFS18, and SAFS19. In 2020, a total of 1048 participants were asked to respond to constructs measuring science identity, science self-efficacy, academic self-concept, sense of belonging, science-related career intention, and other factors. However, not all participants completed every construct. For example, seven participants did not respond to the survey items related to science identity, and 10 participants did not respond to the items related to science self-efficacy. Since the missing rate for each construct was less than 5% of the total number of participants, it did not affect the results, and thus, listwise deletion was adopted. By analyzing these estimated values, we aimed to identify meaningful patterns of change in science self-efficacy (see the ‘Linking and Equating’ section below → see the “Section 2.4.1”).

University Institutional Research (IR) data provided demographic information for students, including self-reported binary gender, first-generation student status, and whether they belonged to a URG, such as African American/Black, American Indian/Alaskan Native, Latinx, and Native Hawaiian/Pacific Islander. The demographic characteristics of the participants are presented in Table 2. In our model, we included gender, first-generation status, and URG-related information as covariates to account for the influence of these demographic factors on student outcomes.

Table 2. Demographic characteristics of the participants in SAFS20.

	Frequency	Percent
Gender		
Female	552	63.74
Male	314	36.26
First Gen		
Non-First Gen	237	27.69
First Gen	629	72.31
URG		
Non-URG	258	26.99
URG	698	73.01

Note. Gender is self-identified. A first-generation college student is defined as a student whose parent(s)/legal guardian(s) have not completed a bachelor’s degree at a four-year college or university. URG represents historically underserved racial/ethnic group (e.g., African American/Black, American Indian/Alaskan Native, Latinx, and Native Hawaiian/Pacific Islander).

2.3. Instruments

Student science self-efficacy was measured using six items from the SAFS, as shown in Table 3. Each item asked students to rate their confidence in performing science-related tasks using a five-point Likert-type response scale ranging from “not at all confident” (1) to “absolutely confident” (5). Higher scores indicated higher levels of perceived science self-efficacy.

Because Likert-type responses are ordinal, raw scores do not necessarily represent equal intervals between response categories. For example, a response of “very confident” coded as 4 does not mean that the student is twice as confident as a student who selected “a little confident” coded as 2. Rather, the numerical codes indicate ordered categories of confidence. To address this limitation, Rasch measurement modeling was used to transform students’ ordinal responses into logit scores, which place student science self-

efficacy estimates on a common interval-level scale under the Rasch model assumptions (Boone et al., 2014).

Rasch analyses were conducted using Winsteps 5.3.2 (Linacre, 2022). Student responses to the six self-efficacy items were converted from raw ordinal scores into person measures expressed in logits. In addition, rating-scale step calibrations were estimated to examine the relative distances between adjacent response categories, such as the transition from “not at all confident” (1) to “a little confident” (2), and from “a little confident” (2) to the next higher category. These step values, expressed in logits, are reported in Table 4.

The psychometric quality of the science self-efficacy measure was evaluated using both reliability and validity evidence. Reliability was examined using Cronbach’s alpha, the Rasch separation index, and the Rasch reliability index. Cronbach’s alpha was used to assess internal consistency among the six items, while the separation and reliability indices were used to evaluate how well the instrument distinguished students with different levels of science self-efficacy. Following Boone et al. (2014), Cronbach’s alpha values above 0.70, separation index values above 2.00, and reliability index values above 0.80 were interpreted as evidence of acceptable to strong reliability. Values below these thresholds were interpreted as indicating weaker internal consistency, limited separation, or lower measurement reliability.

Table 3. Items, their Logit Scores, and Reliability Indices for this Study.

Construct	Item	Logit Score	Reliability Cronbach’s α /Separa- tion/Reliability
Science self-efficacy	1. Use technical science skills (use of tools, instruments, and/or techniques)	0.24	0.92/2.70/0.88
	2. Generate a research question	−0.13	
	3. Determine how to collect appropriate data	−0.26	
	4. Explain the results of a study	−0.54	
	5. Use scientific literature to guide research	0.46	
	6. Integrate results from multiple studies	0.23	

Table 4. Summary of Category and Rasch-Andrich Step Threshold.

Category Label	Category Description	Rasch-Andrich Threshold
1	Not at all confident	None
2	A little confident	−3.3
3	Somewhat confident	−1.51
4	Very confident	1.1
5	Absolutely confident	3.71

Validity evidence was examined using confirmatory factor analysis (CFA) in Mplus 8.8 (Muthén & Muthén, 1998–2023). The CFA tested whether the six items represented a single latent construct: student science self-efficacy. All six items demonstrated standardized factor loadings above the commonly accepted minimum threshold of 0.40, suggesting that the items adequately contributed to the measurement of the same underlying construct (Moon et al., 2024).

For subsequent group-level analyses, the Rasch-derived logit scores were used as the outcome variable. Intervention participation was included as the between-subjects factor, coded as 0 for non-intervention students and 1 for intervention students. First-generation college status, URG status, and gender were included as covariates to account for background characteristics that could be associated with students' science self-efficacy.

2.4. Analysis

2.4.1. Linking and Equating

The Rasch measurement method was used to link science self-efficacy scores across four administrations of the SAFS from 2017 to 2020. Linking was necessary because the survey was administered to different groups of students across years, while the same six SAFS items were used each year to measure the same latent construct: student science self-efficacy. Thus, although the individual participants were not necessarily the same across survey years, the repeated use of the same items made it possible to place students' science self-efficacy estimates from different years onto a common measurement scale.

In Rasch measurement, raw scores from different survey administrations are not directly comparable unless they are placed on the same logit scale. To establish this common scale, item and rating-scale step parameters were estimated and anchored across years. Anchoring means that selected item or step parameters are fixed at previously estimated values so that scores from different administrations are calibrated using the same measurement frame of reference. In this study, all rating-scale step values were anchored to maintain consistent category functioning across years. In addition, one of the six science self-efficacy items was anchored to serve as a reference point for the construct, while the remaining five items were freely estimated within the anchored scale.

This linking procedure allowed student science self-efficacy estimates from different years to be expressed on the same logit metric. Once the yearly measures were linked, equating was used to identify comparable score values across survey administrations. Equating places scores from different time points onto a shared scale so that scores can be interpreted as representing equivalent levels of the latent construct. For example, a raw science self-efficacy score of 20 in 2020 corresponded to an estimated Rasch measure of 0.68 logits. On the equated 2019 scale, the same logit value of 0.68 corresponded to an estimated raw score of 18.84. This means that a score of 20 in 2020 and a score of 18.84 in 2019 represented approximately the same level of science self-efficacy on the common Rasch scale.

This procedure enabled the study to examine group-level trends in science self-efficacy across years, even though the survey samples differed from one year to another. The purpose of linking and equating was not to track individual student growth over time, but to ensure that group-level science self-efficacy estimates from different years were placed on a comparable measurement scale. Therefore, the findings should be interpreted as evidence of changes in group-level trends across survey years, rather than as direct evidence of within-person developmental change. The linked and equated results are presented in Tables 5 and 6.

Table 5. Intervention Participants' Science Self-Efficacy (SSE) Estimates across years based on the Equated Scaling.

SSE	SSE	SSE	SSE	SSE
2017	2018	2019	2020	Logit Score
5.63	5.73	5.86	6.00	−6.84
6.64	6.73	6.86	7.00	−5.53
7.45	7.59	7.79	8.00	−4.68
8.21	8.40	8.70	9.00	−4.10
9.04	9.26	9.64	10.00	−3.63
9.95	10.17	10.61	11.00	−3.21
10.90	11.13	11.60	12.00	−2.81
11.87	12.10	12.58	13.00	−2.43
12.89	13.11	13.61	14.00	−2.04
13.92	14.13	14.62	15.00	−1.65
14.98	15.17	15.63	16.00	−1.24
16.05	16.21	16.65	17.00	−0.81
17.07	17.24	17.66	18.00	−0.37
18.09	18.26	18.66	19.00	0.09
19.13	19.29	19.68	20.00	0.57
20.15	20.31	20.69	21.00	1.06
21.23	21.37	21.73	22.00	1.59
22.29	22.41	22.74	23.00	2.16
23.36	23.48	23.77	24.00	2.80
24.35	24.48	24.75	25.00	3.44
25.31	25.44	25.74	26.00	4.06
26.31	26.44	26.75	27.00	4.68
27.39	27.51	27.79	28.00	5.36
28.58	28.64	28.85	29.00	6.28
29.73	29.76	29.91	30.00	7.63

Table 6. Non-participants' Science Self-Efficacy (SSE) Estimates across years based on the Equated Scaling.

SSE	SSE	SSE	SSE	SSE
2017	2018	2019	2020	Logit Score
5.76	5.84	5.94	6.00	−6.33
6.77	6.84	6.95	7.00	−5.03
7.65	7.75	7.92	8.00	−4.20
8.49	8.64	8.89	9.00	−3.64
9.39	9.56	9.87	10.00	−3.19
10.30	10.49	10.85	11.00	−2.80
11.28	11.49	11.86	12.00	−2.42
12.28	12.47	12.86	13.00	−2.06
13.28	13.47	13.86	14.00	−1.70
14.28	14.46	14.84	15.00	−1.34
15.33	15.50	15.87	16.00	−0.95
16.35	16.51	16.85	17.00	−0.55
17.38	17.52	17.86	18.00	−0.13
18.40	18.53	18.86	19.00	0.30
19.37	19.51	19.86	20.00	0.72
20.37	20.52	20.86	21.00	1.15
21.37	21.52	21.86	22.00	1.57
22.38	22.52	22.86	23.00	1.99

23.40	23.53	23.86	24.00	2.42
24.41	24.55	24.86	25.00	2.86
25.43	25.57	25.87	26.00	3.32
26.50	26.61	26.90	27.00	3.83
27.59	27.67	27.92	28.00	4.43
28.73	28.77	28.94	29.00	5.30
29.83	29.85	29.97	30.00	6.62

2.4.2. Repeated Measures Analysis

We adopted a quasi-experimental repeated-measures trend design because the same survey forms were administered to different groups of participants across years, and participants' science self-efficacy estimates were expressed on a common scale. Because the survey participants differed across the four years, the analysis used response estimates rather than repeated observations from the same individuals (see Tables 5 and 6). Thus, the analysis examined year-to-year trends in scale-equated science self-efficacy estimates rather than within-person developmental change.

To examine trends in student science self-efficacy over time, we conducted a repeated-measures between-within-subjects analysis in SPSS (Version 25.0). The four yearly science self-efficacy estimates were entered as the repeated-measures dependent variables, with year specified as the within-subjects factor. Intervention participation was entered as the between-subjects factor to examine whether science self-efficacy differed between intervention and non-intervention participants across time. First-generation status, URG status, and gender were included as covariates.

The model tested the main effect of year, the main effect of intervention participation, and the year-by-intervention interaction. Assumptions were examined by reviewing descriptive statistics, residual distributions, profile plots, Levene's tests for homogeneity of variance, Box's M test for homogeneity of covariance matrices, and Mauchly's test of sphericity. When the sphericity assumption was violated, Greenhouse–Geisser corrected results were interpreted. When significant main or interaction effects were found, follow-up pairwise comparisons were conducted using estimated marginal means with Bonferroni adjustments. Effect sizes were reported using partial eta squared, and adjusted means, standard errors, F statistics, degrees of freedom, and *p*-values were used to interpret the direction and magnitude of the findings.

3. Results

3.1. Research Question 1: Does Student Science Self-Efficacy Change over Time?

Within-subjects effects tests revealed a significant overall increase in student science self-efficacy across the four years ($F_{(3, 864)} = 1360.484$, $p < -0.001$; $M_{2017} = 19.12$, $SE = 0.26$; $M_{2018} = 19.27$, $SE = 0.26$; $M_{2019} = 19.60$, $SE = 0.26$; $M_{2020} = 19.80$, $SE = 0.26$) (see Table 7).

Table 7. Pairwise comparisons of the science self-efficacy variable.

Source	Time				Mean	Mean	Mean
	T1	T2	T3	T4	Difference 1	Difference 2	Difference 3
Science Self-Efficacy	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	T2-T1 (SE)	T3-T2 (SE)	T4-T3 (SE)
(2017–2020)	19.123 (0.26)	19.267 (0.26)	19.597 (0.26)	19.799 (0.26)	0.144 *** (0.002)	0.475 *** (0.007)	0.676 *** (0.003)

*** $p < 0.001$.

3.2. Research Question 2: Does Student Science Self-Efficacy Differ Between Intervention Participants and Non-Participants?

The tests of between-subjects effects indicated a significant difference between intervention and non-intervention participants regarding their overall science self-efficacy ($F_{(1, 866)} = 84.659, p < 0.001$). Intervention participants showed significantly higher scores for science self-efficacy ($M = 21.81, SE = 0.477$) than non-intervention participants ($M = 17.08, SE = 0.190$).

3.3. Research Question 3: Do Student Science Self-Efficacy Trajectories Differ as a Function of the Intervention, First-Generation, URG Status, or Gender?

3.3.1. Intervention

When comparing intervention and non-intervention students, the tests of within-subjects effects showed a significant interaction between the overall change in science self-efficacy and intervention participation ($F_{(1, 866)} = 124.140, p < 0.001$). A post hoc analysis examining changes across each of the four years revealed that science self-efficacy increased significantly during each year (see Figure 1 and Table 8).

Table 8. Summary table of the results of repeated measures analysis.

Source	Time	Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
time	time 1 vs. time 2	2.983	1	2.983	2260.987	0.000	0.723
	time 2 vs. time 3	15.445	1	15.445	2966.558	0.000	0.774
	time 3 vs. time 4	5.795	1	5.795	4055.949	0.000	0.824
time * First-generation	time 1 vs. time 2	0.004	1	0.004	2.885	0.090	0.003
	time 2 vs. time 3	0.016	1	0.016	3.003	0.083	0.003
	time 3 vs. time 4	0.002	1	0.002	1.451	0.229	0.002
time × Gender	time 1 vs. time 2	0.008	1	0.008	5.798	0.016	0.007
	time 2 vs. time 3	0.005	1	0.005	0.884	0.347	0.001
	time 3 vs. time 4	0.002	1	0.002	1.166	0.281	0.001
time × URG	time 1 vs. time 2	0.003	1	0.003	1.902	0.168	0.002
	time 2 vs. time 3	0.010	1	0.010	1.918	0.166	0.002
	time 3 vs. time 4	0.002	1	0.002	1.682	0.195	0.002
time × Intervention	time 1 vs. time 2	0.019	1	0.019	14.149	0.000	0.016
	time 2 vs. time 3	0.023	1	0.023	4.393	0.036	0.005
	time 3 vs. time 4	1.896	1	1.896	1326.649	0.000	0.605
Error (time)	time 1 vs. time 2	1.143	866	0.001			
	time 2 vs. time 3	4.509	866	0.005			
	time 3 vs. time 4	1.237	866	0.001			

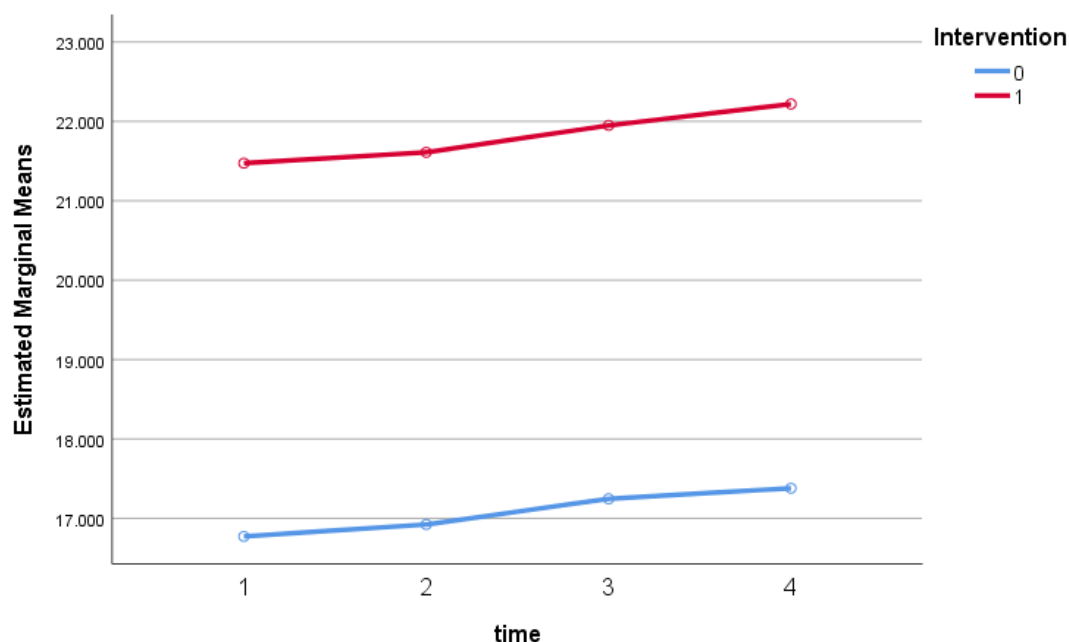


Figure 1. Estimated Marginal Means of Science Self-efficacy.

3.3.2. First-Generation

The tests of within-subjects effects did not show a significant interaction between the overall change in science self-efficacy and first-generation status ($F_{(1, 866)} = 2.788, p = 0.095$). Furthermore, post hoc analyses for each of the four yearly intervals revealed no significant changes in science self-efficacy during any of the periods examined (Table 8).

3.3.3. URG

The tests of within-subjects effects did not show a significant interaction between the overall change in science self-efficacy and URG status ($F_{(1, 866)} = 1.984, p = 0.159$). A post hoc analyses for each of the four yearly intervals revealed no significant changes in science self-efficacy during any of the periods examined (Table 8).

3.3.4. Gender

The tests of within-subjects effects did not show a significant interaction between the overall change in science self-efficacy and gender ($F_{(1, 866)} = 1.766, p = 0.184$). A post hoc analyses of changes across the four yearly intervals revealed that a significant change in science self-efficacy occurred only between timepoint 1 and timepoint 2 (Table 8).

4. Discussion

4.1. Summary of Research Questions and Results

The primary objective of this study was to examine the group-level trends in science self-efficacy among students from diverse backgrounds during their college years. Analysis of the SAFS data showed a statistically significant increase in science self-efficacy across the four yearly estimates from 2017 to 2020, with adjusted means increasing from 19.12 in 2017 to 19.18 in 2020. Because these estimates included both intervention and non-intervention students, the results suggest that students as a whole demonstrated higher confidence in their science skills across the study period.

This finding is consistent with prior research showing that students' science self-efficacy can be strengthened when learning environments provide meaningful opportunities to develop, practice, and receive feedback on science-related skills. From a social cognitive

perspective, science self-efficacy is shaped by mastery experiences, mentoring, social persuasion, and supportive learning contexts. Thus, the upward trend may reflect the broader university environment, including research-related opportunities and STEM-supportive programming that helped students build confidence in their science abilities.

The results also showed a significant overall difference between intervention and non-intervention participants. Intervention participants reported substantially higher science self-efficacy than non-intervention participants across the study period, even after controlling for first-generation status, URG status, and gender. The adjusted mean for intervention participants was 21.81, compared with 17.08 for non-intervention participants. This finding extends prior literature on undergraduate research experiences by showing that students who participated in the intervention consistently demonstrated stronger science self-efficacy than their peers who did not participate. It suggests that structured research experiences may provide academic, social, and mentoring support that is not always available through traditional coursework alone.

A significant interaction between year and intervention participation further indicated that the pattern of science self-efficacy change differed between intervention and non-intervention participants. Follow-up analyses showed significant changes across the yearly intervals, suggesting that the intervention group's advantage was sustained across the four-year period rather than limited to a single year. However, because the analysis was based on linked and equated responses from different participant groups rather than repeated observations of the same individuals, the findings should be interpreted as group-level longitudinal trends rather than evidence of individual-level developmental growth.

The nonsignificant interactions for first-generation status and URG status indicate that the overall science self-efficacy trend did not differ significantly by these characteristics after accounting for the other variables in the model. Similarly, the gender interaction was not significant overall, although follow-up analyses showed a significant change only between the first and second timepoints. These findings suggest that the intervention effect was more pronounced than the effects associated with first-generation status, URG status, or gender. In relation to the literature on STEM inequities, this pattern suggests that well-structured research interventions may support science self-efficacy across diverse student groups rather than benefits only one demographic subgroup.

The findings also contribute to theory by supporting the importance of stage-environment fit in STEM learning contexts. Intervention participants may have experienced stronger alignment between their developmental needs and the opportunities provided by the program, including research engagement, mentoring, and skill-building activities. Such experiences may have strengthened mastery experiences and reinforced students' belief that they could successfully perform scientific work. Given that self-efficacy is a significant predictor of students' academic persistence, career choices, and career development (Bandura et al., 2001; Hackett & Betz, 1981; Lent & Brown, 2006), these findings emphasize the importance of undergraduate research experiences as a mechanism for strengthening students' long-term STEM pathways.

This study advances existing knowledge in two ways. First, it demonstrates that science self-efficacy can be examined over time using linked and equated survey estimates when the same individuals are not available across all years. Second, it shows that intervention participants consistently reported higher science self-efficacy than non-intervention participants across multiple years, suggesting that structured undergraduate research experiences may help reduce opportunity gaps by providing students from diverse backgrounds with sustained access to research skill development, mentoring, and confidence-building experiences.

4.2. Limitations

While the findings highlight the effectiveness of the intervention, several limitations must be addressed. We acknowledge the concern that the data do not represent a conventional individual-level longitudinal design in which the same participants are followed across all four years. Because the participant samples varied each year, the 2017 scores should not be interpreted as baseline data, and the year-to-year increases should not be interpreted as evidence that the same students' science self-efficacy increased over four years. Students in any given year may have represented different class levels and may have had varying degrees of prior exposure to research opportunities.

However, the purpose of this study was to explore whether a longitudinal trend analytic approach could still be meaningfully applied when the number of participants observed across all time points is very small or limited. This approach is conceptually similar to large-scale educational trend studies such as TIMSS, which has assessed mathematics and science achievement generally every four years since 1995 and allows participating countries to monitor changes in achievement over time across assessment cycles (IEA, n.d.; NCES, n.d.). TIMSS uses scaling, linking, and related procedures to support score comparability and trend reporting across assessment cycles and, in recent cycles, across assessment modes (Martin et al., 2020; von Davier, 2020).

In this study, linking and equating were used to place responses from different cohorts and time points on a common scale, allowing us to examine changes in intervention-related outcomes over time at the group level rather than claiming individual developmental trajectories. This use of linking and equating is consistent with psychometric principles that such procedures can support score comparability across forms, occasions, or groups when appropriate design and measurement assumptions are met (Dorans et al., 2007; Kolen & Brennan, 2014). Therefore, the findings should not be interpreted as evidence of within-person change, but rather as evidence of group-level longitudinal trends among comparable groups of intervention participants. From this perspective, the study demonstrates that, even when very few participants are available across all four years, a TIMSS-like approach can provide useful evidence regarding whether intervention participants show more desirable outcomes over time. The results suggest that intervention participants generally demonstrated more favorable patterns in the measured outcomes, although these findings should be interpreted cautiously as group-level longitudinal trends rather than direct evidence of individual developmental change.

In addition, the study relied solely on self-reported measures of science self-efficacy, which are susceptible to social desirability bias and self-perception errors. Since self-reports do not directly measure performance, future studies should include assessments of student work and performance to better understand whether intervention participants demonstrate greater mastery of specific science skills (Timmerman et al., 2011; Willison & O'Regan, 2007). Future study should also follow the same students over time to examine individual-level change more directly and to test how specific intervention components, such as mentoring quality, research engagement, and culturally responsive support, contribute to science self-efficacy development.

Intervention participants are matched with mentors, who are known to play a critical role in supporting students' development of science self-efficacy. However, the study did not measure the direct impact of mentoring on science self-efficacy trajectories. Future research should examine whether science self-efficacy trajectories differ for students with mentors in both intervention and non-intervention groups (Larose, 2013). While this study focuses on changes in science self-efficacy, it does not examine participants' long-term career outcomes. Future research should track participants' STEM career pathways to assess how increases in science self-efficacy influence their future academic and professional choices (Guay et al., 2006). Family background, such as having relatives in STEM fields,

could influence students' science self-efficacy trajectories. Future studies should consider how these external social and family influences shape students' development (Bandura, 1997; Carter et al., 2019). The broader institutional environment may play a role in shaping students' experiences and their science self-efficacy. Future studies should examine how institutional policies, climate, and available resources contribute to student development and self-efficacy.

5. Conclusions

This study contributes to the growing body of literature on UREs by examining group-level longitudinal trends in students' self-reported science self-efficacy from 2017 to 2020. Using linked and equated data, the study found that intervention participants generally demonstrated higher estimated science self-efficacy scores than non-intervention participants across the study period. Although science self-efficacy appeared to increase over time, these findings should be interpreted cautiously because the analysis was based on estimated responses from different participant groups rather than repeated observations of the same individuals.

The findings suggest that participation in the intervention program was associated with more favorable patterns of self-reported confidence in science-related tasks. However, because the study used a quasi-experimental design, the results should not be interpreted as definitive causal evidence that the intervention produced the observed differences. In addition, the findings reflect students' perceived confidence in their science abilities rather than direct measures of actual science skill development.

The study also contributes to SCT by illustrating how structured undergraduate research experiences may provide learning conditions that support the development of science self-efficacy. From an SCT perspective, intervention experiences such as research participation, mentoring, feedback, and exposure to scientific practices may serve as sources of self-efficacy by offering students opportunities for mastery experiences, social persuasion, vicarious learning, and supportive academic engagement. Thus, the study provides preliminary evidence that research-based interventions may be linked to stronger self-efficacy beliefs among undergraduate students.

Future studies should examine these relationships using designs that follow the same individuals over time and include direct measures of science skills, academic outcomes, and career trajectories. Additional research should also explore how mentoring quality, family support, peer networks, and institutional culture shape students' science self-efficacy and longer-term participation in STEM pathways.

Author Contributions: Conceptualization, S.M.; methodology, S.M.; software, S.M.; validation, S.M.; formal analysis, S.M.; investigation, S.M.; resources, S.M. and G.C.; data curation, S.M.; writing-original draft preparation, S.M., C.K., S.-S.A.G., J.C.P.L., and J.H.V.; writing-review and editing, S.M., C.K., S.-S.A.G., J.C.P.L., J.H.V., P. K., and G.C.; visualization, S.M.; supervision, G.C. and C.K.; project administration, G.C., P.K., C.K., and S.-S.A.G.; funding acquisition, G.C., P.K., and C.K. All authors have read and agreed to the published version of the manuscript.

Funding: Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers RL5GM118975, UL1GM118976, and TL4GM118977 And The APC was funded by NIH BUILD PODER at California State University, Northridge, RL5GM118975, UL1GM118976, and TL4GM118977. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Institutional Review Board Statement: This study, which involves the secondary analysis of data, was conducted in accordance with Declaration of Helsinki, and approved by the Institutional Review Board of California State University Northridge (IRB-FY21-222, 5/11/2021).

Informed Consent Statement: The data used in our study were not collected by us but were obtained from the Diversity Program Consortium Co-ordination and Evaluation Center (CEC) at UCLA and are not owned by the authors. As such, we do not possess the consent forms for the participants.

Data Availability Statement: The data analyzed in this study were obtained from the Diversity Program Consortium Coordination and Evaluation Center (CEC) at UCLA and are not owned by the authors. Data may be available from the provider upon reasonable request and with permission.

Conflicts of Interest: The authors have declared that no conflicts of interest exist.

References

- (Bandura, 1997) Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- (Bandura et al., 2001) Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, 72(1), 187–206. <https://doi.org/10.1111/1467-8624.00273>.
- (Benden & Lauermann, 2023) Benden, D. K., & Lauermann, F. (2023). Searching for short-term motivational alignment and spillover effects: A random intercept cross-lagged analysis of students' expectancies and task values in math-intensive study programs. *Contemporary Educational Psychology*, 73, 102166. <https://doi.org/10.1016/j.cedpsych.2023.102166>.
- (Bond & Fox, 2013) Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- (Boone et al., 2014) Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>.
- (Camacho et al., 2021) Camacho, T. C., Vasquez-Salgado, Y., Chavira, G., Khachikian, C. S., Boyns, D., Cabello, B., & Saetermoe, C. L. (2021). Building science identity and improving equity in STEM through critical race theory and culturally responsive mentorship: An evaluation of the BUILD PODER program. *CBE—Life Sciences Education*, 20(2), ar19.
- (Carter et al., 2019) Carter, D. F., Razo Dueñas, J. E., & Mendoza, R. (2019). Critical examination of the role of STEM in propagating and maintaining race and gender disparities. In M. B. Paulsen & L. W. Perna (Eds.), *Higher education: Handbook of theory and research* (Vol. 34, pp. 39–97). Springer. https://doi.org/10.1007/978-3-030-03457-3_2.
- (Chemers et al., 2001) Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology*, 93(1), 55–64. <https://doi.org/10.1037/0022-0663.93.1.55>.
- (Chemers et al., 2011) Chemers, M. M., Zurbruggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, 67(3), 469–491. <https://doi.org/10.1111/j.1540-4560.2011.01710.x>
- (Dalgety & Coll, 2006) Dalgety, J., & Coll, R. K. (2006). Exploring first-year science students' chemistry self-efficacy. *International Journal of Science and Mathematics Education*, 4(1), 97–116. <https://doi.org/10.1007/s10763-005-1080-3>.
- (Dorans et al., 2007) Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.). (2007). *Linking and aligning scores and scales*. Springer. <https://doi.org/10.1007/978-0-387-49771-6>.
- (Dubina et al., 2021) Dubina, K. S., Ice, L., Kim, J.-L., & Rieley, M. J. (2021). Projections overview and highlights, 2020–30. *Monthly Labor Review*, 1–38. <https://doi.org/10.21916/mlr.2021.20>.
- (Dumont et al., 2012) Dumont, H., Trautwein, U., Lüdtke, O., Neumann, M., Niggli, A., & Schnyder, I. (2012). Does parental home-work involvement mediate the relationship between family background and educational outcomes? *Contemporary Educational Psychology*, 37(1), 55–69. <https://doi.org/10.1016/j.cedpsych.2011.09.004>.
- (Estrada et al., 2018) Estrada, M., Burnett, M., Campbell, A. G., Campbell, P. B., Denetclaw, W. F., Gutiérrez, C. G., Hurtado, S., John, G. H., Matsui, J., Okpodu, C. M., Robinson, T. J., Rogers, J. R., Ruiz-Harrington, V. M., Silva, C. L., Tsingalia, A., & Zavala, M. (2018). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education*, 17(1), es1. <https://doi.org/10.1187/cbe.16-01-0038>.
- (Fan & Chen, 2001) Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review*, 13(1), 1–22.

- (Furchtgott-Roth et al., 2009) Furchtgott-Roth, D., Jacobson, L., & Mokher, C. (2009). *Strengthening community colleges' influence on economic mobility*. Hudson Institute Center for Employment Policy. Available online: <https://files.eric.ed.gov/fulltext/ED507430.pdf> (accessed on 21 June 2026).
- (Gonzales, 2015) Gonzalez, H. B. (2015, May 12). *The America COMPETES Acts: An overview* (CRS Report No. R43880). Congressional Research Service. <https://crsreports.congress.gov/product/pdf/R/R43880>
- (Guay et al., 2006) Guay, F., Ratelle, C. F., Senécal, C., Larose, S., & Deschênes, A. (2006). Distinguishing developmental from chronic career indecision: Self-efficacy, autonomy, and social support. *Journal of Career Assessment*, 14(2), 235–251.
- (Hackett & Betz, 1981) Hackett, G., & Betz, N. E. (1981). A self-efficacy approach to the career development of women. *Journal of Vocational Behavior*, 18(3), 326–339. [https://doi.org/10.1016/0001-8791\(81\)90019-1](https://doi.org/10.1016/0001-8791(81)90019-1).
- (Haskins et al., 2009) Haskins, R., Holzer, H. J., & Lerman, R. I. (2009, May 12). *Promoting economic mobility by increasing postsecondary education*. Economic Mobility Project, The Pew Charitable Trusts. <https://www.brookings.edu/articles/promoting-economic-mobility-by-increasing-postsecondary-education/> (accessed on 21 June 2026).
- (Hofstra et al., 2020) Hofstra, B., Kulkarni, V. V., Galvez, S. M.-N., He, B., Jurafsky, D., & McFarland, D. A. (2020). The diversity–innovation paradox in science. *Proceedings of the National Academy of Sciences*, 117(17), 9284–9291. <https://doi.org/10.1073/pnas.1915378117>.
- (IBM Corp, 2017) IBM Corp. (2017). *IBM SPSS statistics for windows* (Version 25.0) [Computer software]. IBM Corp.
- (IEA, n.d.) International Association for the Evaluation of Educational Achievement. (n.d.). TIMSS. Available online: <https://www.iea.nl/studies/iea/timss> (accessed on 21 June 2026).
- (Kim & Steiner, 2015) Kim, J. S., & Steiner, P. M. (2015). Multilevel propensity score methods for estimating causal effects: A latent class modeling strategy. In L. A. van der Ark, D. M. Bolt, W. C. Wang, J. A. Douglas, & S. M. Chow (Eds.), *Quantitative psychology research: The 79th annual meeting of the Psychometric Society, Madison, Wisconsin, 2014* (pp. 293–306). Springer. https://doi.org/10.1007/978-3-319-19977-1_21.
- (Kolen & Brennan, 2014) Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer. <https://doi.org/10.1007/978-1-4939-0317-7>.
- (Larose, 2013) Larose, S. (2013). Trajectories of mentors' perceived self-efficacy during an academic mentoring experience: What they look like and what are their personal and environmental correlates? *Mentoring & Tutoring: Partnership in Learning*, 21(2), 150–174. <https://doi.org/10.1080/13611267.2013.813728>.
- (Larose et al., 2006) Larose, S., Ratelle, C. F., Guay, F., Senécal, C., & Harvey, M. (2006). Trajectories of science self-efficacy beliefs during the college transition and academic and vocational adjustment in science and technology programs. *Educational Research and Evaluation*, 12(4), 373–393. <https://doi.org/10.1080/13803610600765836>.
- (Lent & Brown, 2006) Lent, R. W., & Brown, S. D. (2006). On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *Journal of Career Assessment*, 14(1), 12–35. <https://doi.org/10.1177/1069072705281364>.
- (Lent et al., 1994) Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, 45(1), 79–122. <https://doi.org/10.1006/jvbe.1994.1027>.
- (Lent et al., 2016) Lent, R. W., Ezeofor, I., Morrison, M. A., Penn, L. T., & Ireland, G. W. (2016). Applying the social cognitive model of career self-management to career exploration and decision-making. *Journal of Vocational Behavior*, 93, 47–57. <https://doi.org/10.1016/j.jvb.2015.12.007>.
- (Lin et al., 2024) Lin, J. C. P., Moon, S., Guan, S. A., Kwan, P. P., Flores, G. E., & Chavira, G. (2024). Factors influencing science career intention: The power of counterspaces. *Innovative Higher Education*, 49(2), 641–663. <https://doi.org/10.1007/s10755-024-09752-2>.
- (Linacre, 2022) Linacre, J. M. (2022). *Winsteps® Rasch measurement computer program* (Version 5.3.2) [Computer software]. Winsteps.com. Available online: <https://www.winsteps.com> (accessed on 21 June 2026).
- (Martin et al., 2020) Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.). (2020). *Methods and procedures: TIMSS 2019 technical report*. TIMSS & PIRLS International Study Center, Boston College. Available online: <https://timssandpirls.bc.edu/timss2019/methods/> (accessed on 21 June 2026).
- (McGee, 2016) McGee, E. O. (2016). Devalued Black and Latino racial identities: A by-product of STEM college culture? *American Educational Research Journal*, 53(6), 1626–1662. <https://doi.org/10.3102/0002831216676572>.
- (Melguizo & Wolniak, 2012) Melguizo, T., & Wolniak, G. C. (2012). The earnings benefits of majoring in STEM fields among Latino and Black college graduates. *Research in Higher Education*, 53(4), 383–405. <https://doi.org/10.1007/s11162-011-9238-z>.
- (Moon et al., 2024) Moon, S., Guan, Y.-S. A., Vargas, J. H., Lin, J. C. P., Kwan, P. P., Saetermoe, C. L., Flores, G. E., & Chavira, G. (2024). Critical mentorship in undergraduate research experience builds science identity and self-efficacy. *International Journal of Science and Mathematics Education*, 23(2), 321–341. <https://doi.org/10.1007/s10763-024-10476-0>.

- (Murnane & Willett, 2010) Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.
- (Muthén & Muthén, 1998–2023) Muthén, L. K., & Muthén, B. O. (1998–2023). *Mplus* (Version 8.9) [Computer software]. Muthén & Muthén. Available online: <https://www.statmodel.com> (accessed on 21 June 2026).
- (NCES, n.d.) National Center for Education Statistics. (n.d.). *Trends in International Mathematics and Science Study (TIMSS)*. U.S. Department of Education, Institute of Education Sciences. Available online: <https://nces.ed.gov/timss/> (accessed on 21 June 2026).
- (National Science Board, 2022) National Science Board. (2022). *Science and engineering indicators 2022: The state of U.S. science and engineering* (NSB-2022-1). National Science Foundation. Available online: <https://nces.nsf.gov/pubs/nsb20221> (accessed on 21 June 2026).
- (Office of the Press Secretary, 2010) Office of the Press Secretary. (2010). *President Obama expands “Educate to Innovate” campaign for excellence in science, technology, engineering, and mathematics (STEM) education*. The White House. Available online: <https://www.presidency.ucsb.edu/documents/press-release-president-obama-expands-educate-innovate-campaign-for-excellence-science> (accessed on 21 June 2026).
- (Pew Research Center, 2014) Pew Research Center. (2014). *The rising cost of not going to college*. Available online: <https://www.pewresearch.org/social-trends/2014/02/11/the-rising-cost-of-not-going-to-college/> (accessed on 21 June 2026).
- (Pew Research Center, 2016) Pew Research Center. (2016). *5 facts about Latinos and education*. Available online: https://www.pewresearch.org/short-reads/2016/07/28/5-facts-about-latinos-and-education/?gad_source=1&gad_campaignid=23853897313&gclid=EAIaIQobChMIj6-Kp6KplQMVXW0PAh2DJgy0EAAYASAAEgJddfd_BwE (accessed on 21 June 2026).
- (Rivers et al., 2020) Rivers, R., Norris, K. C., Hui, G., Halpern-Felsher, B., Dodge-Francis, C., Guerrero, L. R., Golshan, A., Brinkley, K., Tran, K., McLaughlin, S., Antolin, N., Yoshida, T., Caffey-Fleming, D. E., & Agodoa, L. (2020). The NIDDK high school short-term research experience for underrepresented persons. *Ethnicity & Disease*, 30(1), 5–14. <https://doi.org/10.18865/ed.30.1.5>.
- (Robbins et al., 2004) Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288. <https://doi.org/10.1037/0033-2909.130.2.261>.
- (G. Saw et al., 2018) Saw, G., Chang, C.-N., & Chan, H.-Y. (2018). Cross-sectional and longitudinal disparities in STEM career aspirations at the intersection of gender, race/ethnicity, and socioeconomic status. *Educational Researcher*, 47(8), 525–531. <https://doi.org/10.3102/0013189X18787818>.
- (G. K. Saw & Agger, 2021) Saw, G. K., & Agger, C. A. (2021). STEM pathways of rural and small-town students. *Journal of Research in Science Teaching*, 58(9), 1244–1271. <https://orcid.org/0000-0002-9328-2830>.
- (Solórzano, 1997) Solórzano, D. G. (1997). Images and words that wound: Critical race theory, racial stereotyping, and teacher education. *Teacher Education Quarterly*, 24(3), 5–19.
- (Solórzano, 1998) Solórzano, D. G. (1998). Critical race theory, race and gender microaggressions, and the experience of Chicana and Chicano scholars. *International Journal of Qualitative Studies in Education*, 11(1), 121–136. <https://doi.org/10.1080/095183998236926>.
- (Thoemmes & Kim, 2011) Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, 46(1), 90–118. <https://doi.org/10.1080/00273171.2011.540475>.
- (Timmerman et al., 2011) Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a “universal” rubric for assessing undergraduates’ scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, 36(5), 509–547.
- (Tinto, 2012) Tinto, V. (2012). *Completing college: Rethinking institutional action*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226804545.001.0001>.
- (Vasquez-Salgado et al., 2023) Vasquez-Salgado, Y., Saetermoe, C. L., Chavira, G., Khachikian, C. S., Boyns, D., & Cabello, B. (2023). Culturally informed critical race mentoring to support undergraduate students in science. *Journal of Diversity in Higher Education*, 16(3), 306–319. <https://doi.org/10.1037/dhe0000431>.
- (von Davier, 2020) von Davier, M. (2020). TIMSS 2019 scaling methodology: Item response theory, population models, and linking across modes. In M. O. Martin, M. von Davier, & I. V. S. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report* (pp. 11.1–11.25). TIMSS & PIRLS International Study Center, Boston College. Available online: <https://timssandpirls.bc.edu/timss2019/methods/chapter-11.html> (accessed on 21 June 2026).

(Willison & O'Regan, 2007) Willison, J., & O'Regan, K. (2007). Commonly known, commonly not known, totally unknown: A framework for students becoming researchers. *Higher Education Research & Development*, 26(4), 393–409. <https://doi.org/10.1080/07294360701658609>.

(Women and Minorities in STEM Booster Act, 2021) Women and Minorities in STEM Booster Act of 2021, H.R. 3855, 117th Cong. (2021). Available online: <https://www.congress.gov/bill/117th-congress/senate-bill/2217> (accessed on 21 June 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.