Editors' Suggestion Featured in Physics

Examining the effect of counternarratives about physics on women's physics career intentions

Geoff Potvin,[1,2] Zahra Hazari,[1,2,3] Raina Khatri,[2] Hemeng Cheng,[1] T. Blake Head,[1]
Robynne M. Lock,[4] Anne F. Kornahrens,[5] Kathryne Sparks Woodle,[6] Rebecca E. Vieyra,[7]
Beth A. Cunningham,[8] Laird Kramer,[1,2] and Theodore Hodapp[9]

[1]*Department of Physics, Florida International University, 11200 SW 8th Street, Miami, Florida 33199, USA*
[2]*STEM Transformation Institute, Florida International University, 11200 SW 8th Street, Miami, Florida 33199, USA*
[3]*Department of Teaching and Learning, Florida International University, 11200 SW 8th Street, Miami, Florida 33199, USA*
[4]*Department of Physics and Astronomy, Texas A&M University-Commerce, 2200 Campbell Street, Commerce, Texas 75428, USA*
[5]*Fannie and John Hertz Foundation, 6701 Koll Center Parkway, Suite 250, Pleasanton, California 94566, USA*
[6]*American Physical Society, One Physics Ellipse, College Park, Maryland 20740, USA*
[7]*Department of Physics, University of Colorado Boulder, 390 UCB, Boulder, Colorado 80309, USA*
[8]*American Association of Physics Teachers, One Physics Ellipse, College Park, Maryland 20740, USA*
[9]*Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, California 94304, USA*

(Received 1 November 2021; revised 13 November 2022; accepted 15 February 2023; published 7 April 2023)

Women and many people of color continue to be minoritized in STEM and notably in physics. We conducted two studies demonstrating that exposure to counternarratives about *who does physics* and *why one does physics* significantly increases high school students—especially women's—physics-related career intentions. These counternarratives facilitate making connections with students' career plans and help in sensemaking causes for the continued minoritization of women in physics. Two separate studies measured the impacts of these interventions on students' physics-related career intentions: first, with an intentionally selected group of teachers (10 teachers, 823 students) across regions and contexts in the U.S.; second, with a randomly sampled group of teachers (13 teachers, 1509 students) from three regions that also included a comparable control group. The results clearly show the importance of exposure to counternarratives in the development of high school students' career interests, particularly for women and minoritized racial or ethnic groups, and that such counternarratives may help to address systemic issues of underrepresentation in STEM.

DOI: 10.1103/PhysRevPhysEducRes.19.010126

I. INTRODUCTION

In the U.S., only 20% of undergraduate physics degrees are awarded to women [1,2]. This percentage has remained stagnant for the past two decades; it stands in stark contrast to chemistry, which awarded 49% of undergraduate chemistry degrees in 2018 to women, and biology, which awarded 63% of its undergraduate degrees in 2018 to women [1,2]. Furthermore, the U.S., Canada, and several western countries have particularly low representation, while Egypt and Iran award a majority of undergraduate physics degrees to women [3–9]. This state of affairs demonstrates that the individual choice to study physics is influenced by cultural associations and complex social dynamics that are often represented in cultural narratives, both implicit and explicit, about who does physics and why one does physics. The impact of these cultural narratives can serve to not only deprive women of a potentially desirable career, but also deny society of the benefits that come from more contributions and diverse viewpoints in solving complex problems [10].

In this study, we examine the effect of physics lessons with counternarratives on students' future physics career intentions. Counternarratives are an approach to countering normative cultural narratives by explicitly providing perspectives of those who have been marginalized and, in so doing, support individuals to challenge inequities they may encounter [11–13]. In physics, these inequities are wide-ranging and systemic, largely shaped by hegemonic perspectives in the development of the field as well as the perpetuating and reinforcing of these perspectives over generations through the continual presence and success of those who purport them (to the exclusion of other perspectives and voices) [14–17]. As such, this work contributes to advancing inclusion in the field by testing an intervention that resists inequitable narratives with counternarratives in introductory high school physics classes. While this intervention will not solve the deep seated inequities in the field, it can provide one possible resource to support cultural change.

II. CULTURAL NARRATIVES IN PHYSICS

Many cultural narratives exist that purport to explain and/or reinforce the overrepresentation of men in particular STEM fields, often based around perceptions of what characteristics are "required" for success in these domains [18,19] or based on traditional narratives about gender-based competency [20–23]. Such narratives, reinforced both explicitly and implicitly in a myriad of different venues, contribute to limiting the participation of women in particular fields such as physics [19,24,25]. Students exposed to such narratives inside and outside the classroom may be driven from further participation in physics [26]. While prior work has highlighted the decreased participation of girls in STEM disciplines as a result of declining perceptions of STEM [27,28], interventions focused on addressing these systemic issues at the secondary level are limited, particularly in physics. Despite the lack of attention to this issue at this grade band, there is clear evidence that the critical period for developing interest in physics careers for women occurs in high school [29], somewhat later than the period critical for general STEM interest [30,31].

Moreover, many stereotypic narratives that reinforce the marginalization of women are refuted by evidence. For example, performance differences between men and women in math and science are regularly found to be null or negligible, including in physics, with women often having higher general academic performance than men [32–34]. It has been noted that associations between brilliance and specific disciplines (including physics) are highly correlated to actual gender representation in these fields [19]. The fact that these associations are connected with actual representation patterns emphasizes the need to disrupt gendered brilliance narratives that influence participation.

III. COUNTERNARRATIVES

In order to neutralize the negative effects of normative and stereotypic cultural narratives, researchers and educators have posited that such narratives need to be disrupted through "counternarratives" [11,12]. Counternarratives are discourses that resist and transform cultural narratives by explicitly providing perspectives of those who have been marginalized [13]. In STEM, much of the current work related to disrupting normative cultural narratives focuses on training or interventions in the workplace (e.g., implicit bias training) or improving faculty members' reflective practice but not with explicit interventions with students in classrooms in the context of specific STEM disciplines [35–38]. Unlike past interventions developed to address inequities in the classroom by unconsciously inoculating students against stereotype threats (e.g., values affirmation [39–41]), counternarratives directly challenge normative cultural narratives, engaging individuals to reflect on systems so that they can recognize and equip themselves to challenge inequities they or others around them face [12]. Despite the fact that many STEM teachers are often trained in equity-oriented pedagogy and culturally relevant practices [42], it may be challenging to put these ideas into practice by engaging students directly in discourse about cultural narratives and counternarratives in STEM. As such, the engagement of students with physics counternarratives is limited. One salient study found that only the explicit discussion of underrepresentation had a significant, positive impact on female students' physics intentions, among other commonly proposed interventions (e.g., having all-female classrooms, having a female teacher or female guest speakers, or discussing the work of women scientists) [43].

The counternarrative interventions used in our study are active-learning lessons including (i) a "Careers in Physics" (CiP) lesson which emphasized the broad range of career opportunities and career goals (foregrounding communal goals such as benefiting society and helping others as emphasized by Diekman *et al.* [44]) that can be achieved through a physics bachelor's degree, and (ii) a "Women in Physics" (WiP) lesson which explicitly engaged students to think about and discuss women's underrepresentation in physics [43] and its underlying causes by drawing on evidence (including cross-cultural comparisons and implicit bias studies). The CiP lesson counters the normative cultural narrative that physicists work on narrow types of problems that have limited connection to people and society [26]. It includes profiles of dozens of individuals (mostly women and people of color from a range of personal backgrounds) who earned bachelor's degrees in physics and who now work across a broad range of sectors. Students are matched with profiles based on their own career values and goals, and are asked to envision how learning physics may connect with their future aspirations. The WiP lesson counters the normative cultural narrative that physicists are primarily white men who innately possess the skills needed to do physics [19]. In this lesson, students reflect on the representation statistics in physics, generate hypotheses as to why underrepresentation remains a persistent problem, examine evidence from studies and their own experiences, and set classroom goals and commitments to support each other in learning physics. Both interventions were developed by a team of physics education researchers, gender researchers, high school teachers, and women students of color. The interventions draw on prior literature in physics and STEM education and gender issues, as well as practical considerations of implementation in physics classrooms. Each lesson was designed to fit within 1–2 class periods and be flexible enough to accommodate a range of classroom contexts and levels of technology support. Further details of the content of the lessons can be found in other sources [45,46], including the complete content of the lessons, at Ref. [47].

To understand how engagement with counternarratives may impact students' intentions towards physics careers, we conducted two studies in 2017–2018 and 2018–2019 to test the effect of our interventions on students' future physics intentions, particularly for female students. Furthermore, although discussions of racial or ethnic inequity are not a core focus of the lesson materials, counternarratives can benefit minoritized groups beyond the particular group the counternarrative centers as exemplars of resistance against a dominant narrative. As such, we were motivated to investigate the impacts of the lessons on students who identified with minoritized racial or ethnic groups (MRE). The research questions we seek to address in this paper are the following:

1. ***Study 1:*** How does experiencing counternarratives embedded in classroom lessons impact high school students' future physics intentions, particularly for female-identifed students and students who identify from minoritized racial or ethnic groups?
2. ***Study 2:*** How does experiencing counternarratives embedded in classroom lessons impact high school students' future physics intentions, particularly for female-identifed students and students who identify from minoritized racial/ethnic groups, by comparison to similar groups of students who experience a control lesson (without counternarratives)?

In the remainder of this paper, we report on the results of these two separate studies. We first describe the context of each study including the data collected, describe the analysis used (including measurement validation), present the main findings of each study, and, lastly, discuss the importance and limitations of this work.

IV. DATA COLLECTION AND METHODS

A. Study 1

Study 1 included the classes of 10 high school physics teachers in eight states (teacher and school profiles appear in Tables VI and VII in Appendix A) who implemented the lessons in their physics classes during the fall of 2017 and collected pre- and postsurvey data from their students, including students' future physics intentions. These teachers taught physics in a broad range of contexts including urban, suburban, and rural schools; public and private schools; diverse student racial or ethnic identities, including classes with primarily white students and those with primarily Black and Latino/a students; and a range of socioeconomic backgrounds. Further, all of these teachers had previously participated in other studies on inclusivity in the classroom and/or were identified as having a strong commitment to working on inclusivity. They participated directly in the initial intervention development to ensure the practical feasibility of the interventions for different contexts before implementing them in their classes. Thus, Study 1 teachers were familiar with the lessons. However, there was no training provided for the lessons since we wanted the lessons to be effectively implementable without requiring extensive training for participants (since professional development and training can be practically costly and difficult to provide and attend). Pre- and postsurveys were administered to 823 students across the 10 teachers' classes before and after each intervention.

While Study 1 examined the effect of counternarratives for the intended persistence of students in physics, particularly female students, there was a need to replicate the study for a few reasons. First, the teachers involved in Study 1 were not randomly selected, may have been especially motivated to teach for inclusivity in their classrooms, and were generally highly experienced in teaching physics. Furthermore, in this first study, there was no control group of teachers to which we could compare the outcomes, which left open the possibility that any measured changes in future physics intentions may not have been significantly different than those that would come from other classroom experiences. (However, this would be surprising, given the prior literature on declining attitudes in typical introductory physics courses [48]). This motivated us to conduct Study 2, in which we recruited a random sample of teachers and included a control group which taught the Notable Physicists lesson.

B. Study 2

Study 2 included the classes of 13 randomly selected and recruited teachers in three regions across four states (representing rural and urban settings with a diversity of student populations and school contexts), who were randomly assigned to a treatment group that used the counternarrative interventions or to a control group (sample details appear in Tables IX and X in Appendix A). To recruit classes for participation, a list of high school physics teachers was generated from public school districts in the three regions, which was then stratified and randomized to capture varying socioeconomic profiles of schools (as measured by free-and-reduced lunch rates). In Study 2, teachers were not familiar with the lessons and no training was provided for the lessons again since we wanted the lessons to be effectively implementable without requiring extensive training for participants. However, each lesson did include a detailed lesson plan and a brief video explaining the parts of the lesson.

The control group implemented a "Notable Physicists" (NP) lesson, which was also an active learning lesson with similar pedagogical structure and length as the treatment interventions, but did not incorporate physics counternarratives. In the control lesson, students explored profiles of historical physicists conducting traditional research and/or academic physicists. Women and individuals from minoritized racial or ethnic groups were included in the lesson but they were few in number and worked in historically traditional physics fields, reflecting a more normative depiction of physics practice [19].

Beginning in the fall of 2018, teachers in both treatment and control groups were given the lessons and their students were surveyed pre and post each lesson, and then again at a delayed time point later in the semester (the length of the delay varied according to the teachers' schedules), resulting in a sample of 1509 student responses. The purpose of the delayed measure was to try to account for possible declines in attitudes some time after the end of the lessons. Also of note is that within the treatment group, some teachers were randomly assigned to teach the interventions in a particular order (the CiP lesson followed by the WiP lesson), and the rest of the treatment group of teachers were assigned to teach the interventions in the reverse order. Comparing the student outcomes of these two subgroups, no meaningful differences between the two subgroups was found, so the treatment group is considered as a single condition throughout the analysis presented in this paper.

C. Handling of demographic information

Throughout this work, we deliberately chose not to treat any group of students (e.g., white, male students) as a "normative" baseline [14,49]. So, in all of the instruments of this study, gender and race or ethnicity were collected with explicitly multicategorical items that allowed students to self-identify with any, all, or none of several different gender and racial or ethnic identities. In the analysis that follows, we assess outcomes from those students who explicitly identified as female (regardless of any other gender identification, if any) from nonfemales (e.g., those who did not explicitly identify as female, regardless of any other gender identification), and in Study 2 we compare the different condition groups (e.g., treatment versus control) for students who identified as female *separately* from those who did not. We chose not to exclude any students from our analysis due to smaller sample sizes of some groups of students or make binary female-to-male comparisons when our main focus was on female students. Similarly, we assess outcomes for students who identified with a minoritized racial or ethnic group in STEM—African-American, Black, Hispanic, Latino/a, American Indian or Alaskan Native—regardless of other racial or ethnic identifications (if any) *separately* from those who did not identify with any of these groups [49].

D. Instrument validation

In this paper, the primary dependent (outcome) variable is a measure of future physics intentions, taken before and after each lesson (and, in Study 2, once more at a delayed time point). See Tables VIII and XI in Appendix A for a descriptive summary of the pre- and postfuture physics intentions in both studies, broken down by demographic groups. As we report in detail below, the measure of students' future physics intentions was constructed from four survey items focused on whether students intended to persist in physics, and is strongly correlated with students' intentions to apply to undergraduate physics programs. Moreover, STEM intentions as early as middle school have been previously shown to be highly predictive of STEM degree outcomes [30,31].

In order to establish the validity of the future physics intentions construct, we conducted confirmatory factor analysis with four survey items as indicators of the latent variable of future physics intentions. Table I summarizes the factor loadings, item reliabilities, construct reliability, and average variance extracted (AVE) using the initial survey data for Study 1. Table II summarizes the same results for Study 2.

TABLE I. CFA results for Study 1.

Item	Std. factor loading	Std. error	Item reliability	Construct reliability	AVE
I can see myself as a physicist.	0.91***	0.09	0.83	0.96	0.86
A future in physics is a possibility for me.	0.94***	0.09	0.88		
I am likely to major in physics in college/university.	0.93***	0.08	0.87		
I could see myself pursuing a physics-related career.	0.93***	0.09	0.86		

*** Represents a significance level of $p < 0.001$.

TABLE II. CFA results for Study 2.

Item	Std. factor loading	Std. error	Item reliability	Construct reliability	AVE
I can see myself as a physicist.	0.92***	0.06	0.84	0.95	0.84
A future in physics is a possibility for me.	0.94***	0.06	0.88		
I am likely to major in physics in college/university.	0.92***	0.06	0.85		
I could see myself pursuing a physics-related career.	0.89***	0.07	0.79		

*** Represents a significance level of $p < 0.001$.

To support convergent validity, at a minimum, all standardized factor loadings should be significant while it is also recommended that factor loadings be 0.5 or higher [50]. To assess convergent validity with construct reliabilities, a threshold greater than 0.7 is considered good reliability. Furthermore, for convergent validity AVE should be greater than 0.5. Both Study 1 and Study 2 CFA results meet the requirements for convergent validity (see Tables I and II). Note that the construct reliability for Study 1 is 0.96 and 0.95 for Study 2, which well exceed the recommended minimum reliability of 0.7 [50].

For the used fit indices, the minimum recommended cutoffs are $\chi 2 \;\; p > 0.05$; $\mathrm{GFI} > 0.95$; $\mathrm{AGFI} > 0.95$; $\mathrm{RMSEA} < 0.07$; $\mathrm{SRMR} < 0.08$; $\mathrm{CFI} > 0.97$; and $\mathrm{NNFI} > 0.95$ [50]. For Study 1, the CFA fit indices are $\chi 2$ p $= 0.060$, $\mathrm{GFI} = 0.997$, $\mathrm{AGFI} = 0.974$, $\mathrm{RMSEA} = 0.061$, $\mathrm{SRMR} = 0.004$, $\mathrm{CFI} = 0.999$, and $\mathrm{NNFI} = 0.995$. These indices all meet the recommended thresholds and thus the measurements are well fit to the underlying latent variable. For Study 2, the CFA fit indices are $\chi 2 \;\; p = 0.267$, $\mathrm{GFI} = 0.999$, $\mathrm{AGFI} = 0.995$, $\mathrm{RMSEA} = 0.014$, $\mathrm{SRMR} = 0.002$, $\mathrm{CFI} = 1.000$, and $\mathrm{NNFI} = 1.000$. Again, these indices all meet the recommended thresholds and thus the measurements are well fit to the underlying latent variable.

We also assessed the criterion-related validity of the composite measure for future physics intentions (combining the 4 observed measures) by examining its correlation with the likelihood that students intended to apply to an undergraduate program in physics. The composite for Study 1 had a strong correlation with intending to apply to physics programs of $r = 0.724$ ($p < 0.001$). The composite for Study 2 also had a strong correlation with intending to apply to physics programs of $r = 0.814$ ($p < 0.001$). This further supports the validity of the dependent variable.

Finally, since we examine the effects on certain groups (female or nonfemale and MRE or non-MRE) in this work, it was important to ensure that the measures functioned similarly for the respective groups. Thus, we tested the measurement invariance for female or nonfemale and for MRE/non-MRE students. The results are summarized in Table III for Study 1 and in Table IV for Study 2. For both Study 1 and 2 across both comparison groups (female or nonfemale and MRE/non-MRE), the ΔCFIs are all less than 0.01 which is the cutoff recommended by most for establishing strong invariance [51]. Thus, the results indicate that the latent construct (dependent variable) does not have statistically significantly different structures or meanings for the groups compared and can be used to assess effects for the separate groups.

TABLE III. CFI and ΔCFI for Study 1 invariance tests.

Groups	Model	CFI	ΔCFI
Female/Nonfemale	Configural	0.996	⋯
	Metric (weak)	0.996	0.000
	Scalar (strong)	0.994	0.002
MRE/Non-MRE	Configural	0.997	⋯
	Metric (weak)	0.997	0.000
	Scalar (strong)	0.998	0.001

TABLE IV. CFI and ΔCFI for Study 2 invariance tests.

Groups	Model	CFI	ΔCFI
Female/Non-Female	Configural	1.000	⋯
	Metric (weak)	0.999	0.001
	Scalar (strong)	0.999	0.000
MRE/Non-MRE	Configural	0.998	⋯
	Metric (weak)	0.999	0.000
	Scalar (strong)	0.996	0.003

E. Multiple imputation of data

Because of the challenges of collecting data multiple times during the semester from students in real high school classrooms, there was missingness in the sample that needed to be accounted for. In Study 1, the fraction of missingness for the variables used in this manuscript ranges from 0.17 to 0.28, with a mean of 0.20. In Study 2, the fraction of missingness for the variables used in this analysis ranges from 0.32 to 0.49, with a mean of 0.38. To properly handle missingness, multiple imputation [52] methods were used throughout, using the Amelia package in **R** [53,54]. Multiple imputation is a best practices approach to estimating missing responses while also properly accounting for various sources of variance. Amelia implements a bootstrapping algorithm that is run multiple times to properly represent the uncertainty in missing values. In the case of the current analysis, the data was imputed 100 times ($m = 100$) which is significantly above modern recommendations for multiple imputation [55]. Also, to ensure consistency of chain length in multiple imputation, a ridge prior of 1% was used [54]. This is consistent with best-practices recommendations and well below the maximum recommended ridge prior of 10%. In order to ensure that the imputation model incorporated sufficient variance, extra variables beyond the ones directly used in this analysis were included in the imputation step, including: students' reported likelihood of pursuing college physics, their year of enrollment in high school, and their recognition beliefs regarding physics [56]. Finally, note that the fully imputed results that are presented next are not substantially different from a similar analysis of the raw, unimputed findings, which present nearly the same picture with only minor adjustments to estimated effects; all of the primary inferential results appearing in the main body are similar. However, the multiply imputed inferences are much more robust and so are reported here.

V. RESULTS

A. Study 1

As seen in Fig. 1 (the left-most bars shaded yellow-orange), both female and nonfemale-identified students have statistically significant increases in their future physics intentions after the intervention. Comparing students' pre-to-post change in future physics intentions using paired t-tests, we find significant increases for female-identified students with an effect size of 0.29 [$t(N) = 5.61(382)$, $p < 0.001$, gain of 8.2, 95% confidence interval [5.3–11.1], Cohen's $d = 0.29$]. Similarly, for nonfemale students, there is a significant change before and after the lessons with an effect size of 0.19 [$t(N) = 3.97(441)$, $p < 0.001$, gain of 5.0, CI [2.5–7.4], $d = 0.19$]. Figure 2 (again, both yellow-orange shaded bars) shows that MRE students reported significant increases in future physics intentions with an effect size of 0.24 [$t(N) = 4.32(334)$, $p < 0.001$, gain of 6.9, CI [3.7–10.0], $d = 0.24$], as did non-MRE students with an effect size of 0.23 [$t(N) = 5.19(489)$, $p < 0.001$, gain of 6.2, CI [3.8–8.5], $d = 0.23$].

B. Study 2

The results of Study 2 are also summarized in Figs. 1 and 2 (treatment group appears in blue bars, control group in gray bars). Similar to Study 1, the results indicate gains in future physics intentions for students in the treatment group. Specifically, female-identified students had statistically significant gains with an effect size of 0.29 [$t(N) = 6.98(604)$, $p < 0.001$, gain of 8.6, CI [6.1–11.0], $d = 0.29$]. Non-female students in the Treatment group also had significant gains with an effect size of 0.28 [$t(N) = 6.74(574)$, $p < 0.001$, gain of 8.2, CI [5.8–10.6], $d = 0.28$]. Importantly, the results for the Control group (gray bars) contrast strongly with the outcomes in the

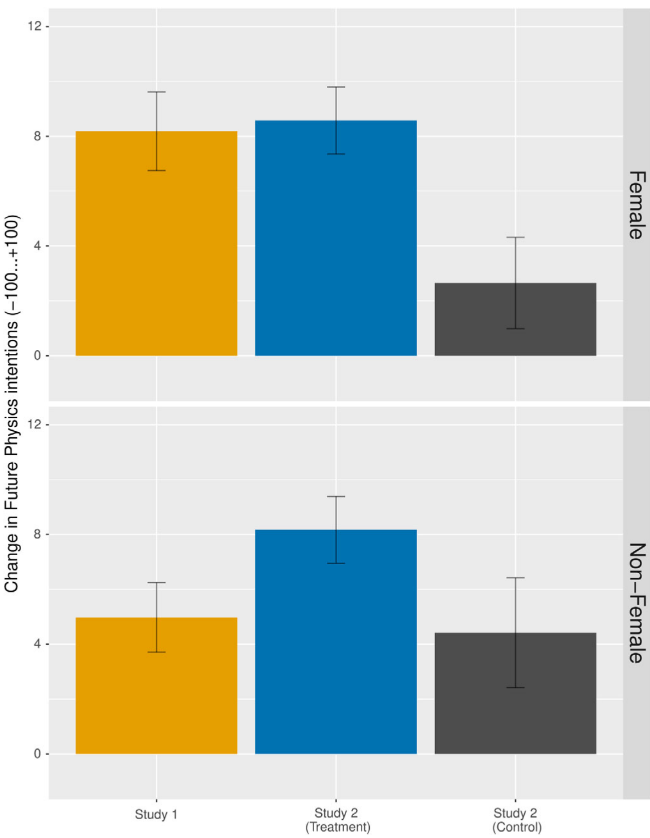

FIG. 1. Outcome of Study 1 (yellow-orange) and Study 2 (treatment in blue, control in gray). Female-identified students (top panel) show statistically significant gains in future physics intentions in Study 1 and Study 2 treatment. Non-female-identified students (bottom panel) show gains in Study 1, Study 2 treatment, and Study 2 control. The difference between treatment and control is statistically significant for female students ($p = 0.005$); it is not statistically significantly different for nonfemale students ($p = 0.11$). The height of each bar is pre-post change in physics intentions (scale −100 to +100); error bars represent standard errors.

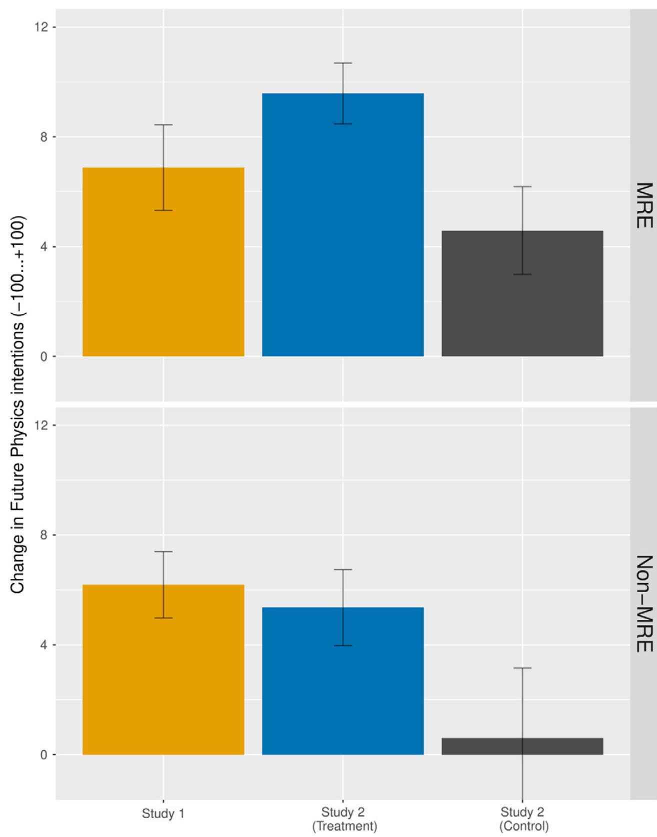

FIG. 2. Outcome of Study 1 (yellow-orange) and Study 2 (treatment in blue, control in gray). MRE-identified students (top panel) in Study 1, Study 2 treatment, and Study 2 control show statistically significant gains in future physics intentions; non-MRE-identified students (bottom panel) also show gains in Study 1 and in the Study 2 treatment. The difference between treatment and control is statistically significant for MRE ($p = 0.008$), but not for non-MRE students ($p = 0.11$). The height of each bar represents pre-post change in physics intentions (scale −100 to +100); error bars represent standard errors.

TABLE V. Summary of unpaired t tests for Study 2. Each row represents the independent samples Welch's t test after combining 100 multiply imputed datasets according to Rubin's rules.

Group 1 (N_{ave})	Group 1 Mean (SD)	Group 2 (N_{ave})	Group 2 Mean (SD)	t	Sig.	Cohen's d of Diff. [95% CI]
Treatment—Female (604)	8.6 (30.0)	Control—Female (148)	2.7 (20.2)	2.85	$p = 0.005$	0.23 [0.07, 0.39]
Treatment—Non-F (574)	8.2 (29.2)	Control—Non-F (183)	4.4 (27.1)	1.60	$p = 0.11$	0.13 [−0.03, 0.30]
Treatment—MRE (842)	9.6 (32.2)	Control—MRE (252)	4.6 (25.3)	2.67	$p = 0.008$	0.17 [0.05, 0.30]
Treatment—Non-M (336)	5.4 (25.3)	Control—Non-M (79)	0.6 (22.7)	1.60	$p = 0.11$	0.20 [−0.05, 0.44]

treatment group. Female-identified students in the control had no significant changes for their future physics intentions [$t(N) = 1.56(148)$, p $= 0.95$], while nonfemale-identified students reported statistically *significant* gains with an effect size of 0.16 [$t(N) = 2.24(183)$, $p = 0.03$, gain of 4.4, CI [0.5–8.3], $d = 0.16$]. Directly comparing the Control and Treatment groups of Study 2 using unpaired t-tests (summarized in Table V), female-identified students in the Treatment gained significantly more in their future physics intentions than the Control with an effect size of 0.23 [$t(N_1 \text{ and } N_2) = 2.85(604 \text{ and } 148)$, $p = 0.005$, mean difference in gain 5.9, CI [1.8–10.0], $d = 0.23$] while nonfemale-identified students had nonsignificant differences in gains [$t(N_1 \text{ and } N_2) = 1.60(574 \text{ and } 183)$, $p = 0.11$].

Similarly, as shown in Fig. 2, there were gains in future physics intentions for students who identified as MRE and those who did not. In the treatment group, gains for MRE students were statistically significant with an effect size of 0.30 [$t(N) = 8.59(842)$, $p < 0.001$, gain of 9.6, CI [7.4–11.8], $d = 0.30$] as were those for non-MRE students with an effect size of 0.21 [$t(N) = 3.92(336)$, $p < 0.001$, gain of 5.4, CI [2.7–8.1], $d = 0.21$]. In the control group, MRE students did show a significant gain with an effect size of 0.18 [$t(N) = 2.90(252)$, $p = 0.004$, gain of 4.6, CI [1.5–7.7], $d = 0.18$] while non-MRE students did not [$t(N) = 0.23(79)$, $p = 0.82$]. Directly comparing the control and treatment groups (Table V), the gains for MRE students were significantly greater in the treatment group with an effect size of 0.17 [$t(N_1 \text{ and } N_2) = 2.67(842 \text{ and } 252)$, $p = 0.008$, mean difference in gain 5.0, CI [1.3–8.7], $d = 0.17$], but non-MRE students did not have significantly different gains [$t(N_1 \text{ and } N_2) = 1.60(336 \text{ and } 79)$, $p = 0.11$].

Lastly, while we consider the overall gain from the combined lessons to be the most salient outcome for purposes of this work, we conducted further inferences to assess the impacts of each individual lesson (e.g., CIP, WIP, and NP lessons) on students; the detailed outcomes of those tests are presented in Appendix B for both Study 1 and 2. In particular, it is noteworthy that for all of the groups tested, the immediate pre-post gains in future physics intentions for those students experiencing the CiP or WiP lessons in Study 1 and 2 were positive (in most cases, statistically significantly so), while the immediate pre-post gains from the control lesson in Study 2 were not significantly different from zero. See Appendix B for complete details.

VI. DISCUSSION

These two studies demonstrate that engaging with counternarratives in physics can help students to see a future in physics, particularly students from systemically marginalized groups. The results of Study 1 and Study 2 show that for *all* groups tested, there were statistically significant gains in future physics intentions for those students who had engaged with the counternarrative lessons. The contrast with the outcomes of the Control group in Study 2 is stark: the changes in physics intentions of female students and MRE students are significantly greater in the treatment than control group. Although our primary intention was to focus on female students and female underrepresentation in physics, we were able to explore the impact of these lessons on students who identify with minoritized racial or ethnic groups, seeing clear effects for MRE students. Furthermore, Study 2 extended Study 1 by showing that classroom practices that use counternarratives can be effective for randomly selected teachers, not just those who are highly experienced or specifically vested in equity issues. Finally, this work clearly demonstrates that high school classrooms can be an effective place to engage in equity discussions surrounding science participation, and these results have broad applicability to other STEM classrooms.

These findings support prior work, which found that explicit discussion of underrepresentation had a significant, positive impact on female students' physical science career intentions [43], which are strongly predictive of the likelihood of STEM career choice [30,31]. In addition, there is a considerable body of qualitative research which has theorized and demonstrated that counternarratives have a substantive and meaningful impact on women and minoritized racial or ethnic groups [11–13]. Drawing on this theoretical work, this study provides quantitative evidence that strongly supports what qualitative researchers have theorized—the importance of counternarratives as a form of resistance that can disrupt normative cultural narratives about physics and physicists, and in doing so, help support marginalized students to envision a future for themselves in a field like physics. However, while counternarratives do present an avenue for increasing students' ability to envision themselves in physics and resist dominant, marginalizing narratives, they do not address the deep-seated systemic inequities in the field. It is more likely that they create a disruption to a narrow culture (a difficult task in itself) that allows more diverse voices with greater knowledge of the culture to be heard in order to challenge and drive the field towards change. Subsequently, more work is needed on other resources and programs to support individuals from diverse backgrounds and allow their perspectives to flourish (e.g., [15]), thereby moving the culture towards greater inclusivity and dismantling systemic inequities.

A limitation of this study is that it is not possible to follow significant numbers of students through college to understand how many chose to study physics or physics-related majors. Additional work exploring the long-term impacts of counternarratives and similar experiences with resistance (e.g., counterspaces) is necessary. Furthermore, while we did randomly select teachers in Study 2 and randomly assigned them to the groups, we were not able to collect and examine their classroom practices. Moreover, while our studies show the promise of counternarratives for combating prevailing stereotypical discourse around physics and physicists, it is important to note that multiple, persistent classroom approaches are necessary to achieve equity. In many cases, students entering physics classrooms have already been exposed to normative discourse around physics, and the lessons tested here are only the beginning of deeper, necessary conversations about the systemic problems associated with cultural inequities that lead to the reproduction of marginalization in physics. While the gains in physics intentions measured in this study persisted through the end of the data collection period, students are likely to face reiterated and/or reinforced normative cultural narratives and practices in college. Thus, there is an urgent need for post-secondary educators to also challenge marginalizing narratives and practices promoted in college science classrooms and STEM programs with approaches that expose and address systemic cultural issues in STEM, including counternarratives. Future work should also examine the ways in which counternarratives interact with physics identities, given that these identities are typically formed in ways that reinforce normative narratives. Moving forward, it is not enough to promote the agency and resistance of marginalized groups; practices that uphold, reinforce, and reproduce marginalizing narratives need to be challenged through policy and structural changes that reflect the perspectives of those who have historically had little or no voice in shaping those policies and structures.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation through Grants No. 1721021, No. 1720810, No. 1720917, and No. 1720869. Thanks to all of the many individuals who have contributed to the STEP UP project, and to the many teachers and students who participated in these studies. Thank you in particular to the STEP UP development team, propagation team, and research team. Publication of this article was funded in part by the University of Colorado Boulder Libraries Open Access Fund.

APPENDIX A: SAMPLE DETAILS

A. Study 1

10 teachers in eight states participated in Study 1, representing 823 students. These teachers taught in a range of school contexts and had a range of student populations,

TABLE VI. Summary of teacher or school characteristics, Study 1. NA represents a response not self-reported by the teacher. Free or reduced lunch rate (FRL%) not available for private schools (denoted "…").

Teacher	Years teaching	Years teaching physics	Teacher gender	Teacher race or ethnicity	School FRL%	Urban, suburban, rural	Public or private
1	32	32	Male	White, Hispanic/Latino/a	43	Urban	Public
2	5	5	Male	White	31	Suburban	Public
3	15	15	NA	NA	⋯	Rural	Private
4	10	10	Male	Black	30	Suburban	Public
5	13	11	Female	Caucasian	86	Urban	Public
6	22	12	NA	NA	31	Suburban	Public
7	13	13	Female	White, Non-Hispanic	80	Urban	Public
8	3	3	Female	White	45	Suburban	Public
9	20	17	Male	White	⋯	Urban	Private
10	20	20	Male	White	36	Suburban	Public

TABLE VII. Summary of Study 1 sample, by teacher. Percentages quoted are the combined results after imputation ($m = 100$).

Teacher	N	Female identified (%)	MRE identified (%)
1	178	39	70
2	176	48	25
3	7	43	29
4	70	43	24
5	80	40	66
6	192	44	22
7	19	32	95
8	44	100	57
9	49	54	17
10	8	25	12
Total	823	46	41

as summarized in Tables VI and VII. Of note is that 9 of the 10 teachers taught the CiP lesson first, followed by the WiP lesson. Teacher 8 in Tables VI and VII taught the lessons in reverse order, but the ordering was not found to significantly impact the results (in either study). Table VIII summarizes the pre- and post-future physics intentions for Study 1, broken down by the demographic groups that are analyzed.

B. Study 2

Thirteen teachers were recruited from three regions of the U.S. and represent the public school systems in these regions (See Table IX). In total, this study involved a total of 1509 students (See Table X). Table XI summarizes the pre- and postfuture physics intentions for Study 2, for both conditions and broken down by the demographic groups that are analyzed. To recruit teachers, firstly a list was generated of all high school teachers who taught physics in that region and then the list was stratified according to

TABLE VIII. Summary of Study 1 measures, by demographic group. Percentages quoted represent pre- and post-future physics intentions (representing the entire range of possible responses, scaled from 0 to 100), after imputation ($m = 100$).

	Future physics intentions—pre (Mean)	Future physics intentions—pre (SD)	Future physics intentions—post (Mean)	Future physics intentions—post (SD)
All	36	28	42	30
Female	28	27	37	30
Nonfemale	42	28	47	29
MRE	37	30	44	31
Non-MRE	35	28	41	29

TABLE IX. Summary of teacher or school characteristics, Study 2. NA represents a response not self-reported by the teacher.

Teacher	Years teaching	Years teaching physics	Teacher gender	Teacher race or ethnicity	School FRL%	Urban, suburban, rural
Treatment 1	1	0	Female	American Indian/Alaskan Native, White	71	Urban
Treatment 2	2	1	Male	White	47	Rural
Treatment 3	NA	NA	NA	NA	36	Urban
Treatment 4	15	15	Male	White	59	Suburban
Treatment 5	9	9	Male	White	11	Urban
Treatment 6	1	0	Female	White	71	Urban
Treatment 7	10	10	Female	White	14	Suburban
Treatment 8	NA	NA	NA	NA	88	Suburban
Treatment 9	10	10	Female	Black/African-American	71	Urban
Treatment 10	17	12	Male	Black/African-American, White	57	Suburban
Control 1	22	7	Female	White	31	Rural
Control 2	15	15	Male	Hispanic/Latino/a	68	Suburban
Control 3	18	18	Female	White	60	Suburban

TABLE X. Summary of Study 2 sample, by teacher. Percentages quoted are the combined results after imputation ($m = 100$).

Teacher	N	Female identified (%)	MRE identified (%)
Treatment 1	74	46	81
Treatment 2	42	58	23
Treatment 3	41	42	68
Treatment 4	36	49	79
Treatment 5	147	51	34
Treatment 6	183	48	91
Treatment 7	61	38	33
Treatment 8	247	64	89
Treatment 9	146	47	83
Treatment 10	201	49	69
Treatment Total	1178	51	71
Control 1	119	46	50
Control 2	46	40	98
Control 3	166	45	89
Control Total	331	45	76

schools' free or reduced lunch rates (FRL%) into high, medium, and low bins in order to ensure representation of schools from a spectrum of socioeconomic contexts. Teachers from each bin were then contacted in order to recruit them to participate in this study, without knowledge of which lessons would be taught. After the entire sample of teachers had agreed to participate, they were randomly assigned by the research team to one of the condition groups. This ensured that teachers who ended up being assigned to the control group would be similarly motivated as those in the treatment group, and that no teacher would have prior knowledge of the lessons (which were not available ahead of time).

APPENDIX B: SUPPLEMENTARY ANALYSIS

In addition to the results appearing in the main body, we also conducted paired t-tests to assess the change for each group separately for each lesson used (e.g., pre to post of the CiP lesson, the WiP lesson, and the NP lesson, respectively). Figures 3 and 4 represent the effects of each lesson on future physics intentions (Study 1 appears as shades of yellow-orange, Study 2 Treatment appears as shades of blue, and Study 2 Control appears as shades of gray). Further, Tables XII and XIII summarize the details of these t tests including the overall gains over the study duration (these also appear in the main body). Critically, for the CiP and WiP lessons, in no case was the gain for an individual lesson negative and in most cases each lesson was associated with a statistically significant, positive gain in future physics intentions for both female and nonfemale students. By contrast, female students in the Study 2 control group saw no statistically significant change due to the NP lesson, nor did nonfemale students (though there was a significant gain in

TABLE XI. Summary of Study 2 measures, by demographic group and condition. Percentages quoted represent pre- and postfuture physics intentions (representing the entire range of possible responses, scaled from 0 to 100), after imputation ($m = 100$).

	Future physics intentions—pre (Mean)	Future physics intentions—pre (SD)	Future physics intentions—post (Mean)	Future physics intentions—post (SD)
Treatment				
All	29	26	37	28
Female	26	25	34	28
Nonfemale	32	26	40	28
MRE	27	25	36	28
Non-MRE	34	27	39	29
Control				
All	24	26	27	28
Female	17	24	20	25
Nonfemale	29	27	33	29
MRE	24	27	28	29
Non-MRE	23	23	23	25

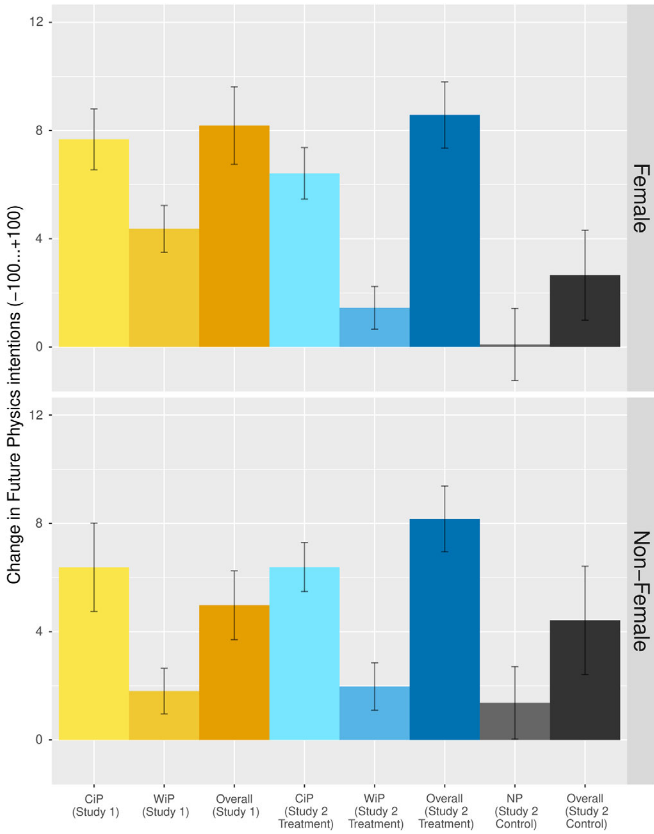

FIG. 3. Summary of results including per-lesson results (Careers in Physics, Women in Physics, and Notable Physicists lessons, in addition to the overall impact on students), female-identified students on top, nonfemale identified students on bottom. For each lesson (CiP, WiP, or NP), the bar represents the immediate post- minus immediate prescore. For the overall measure, the bar represents the latest post- minus prescore. Error bars represent standard errors.

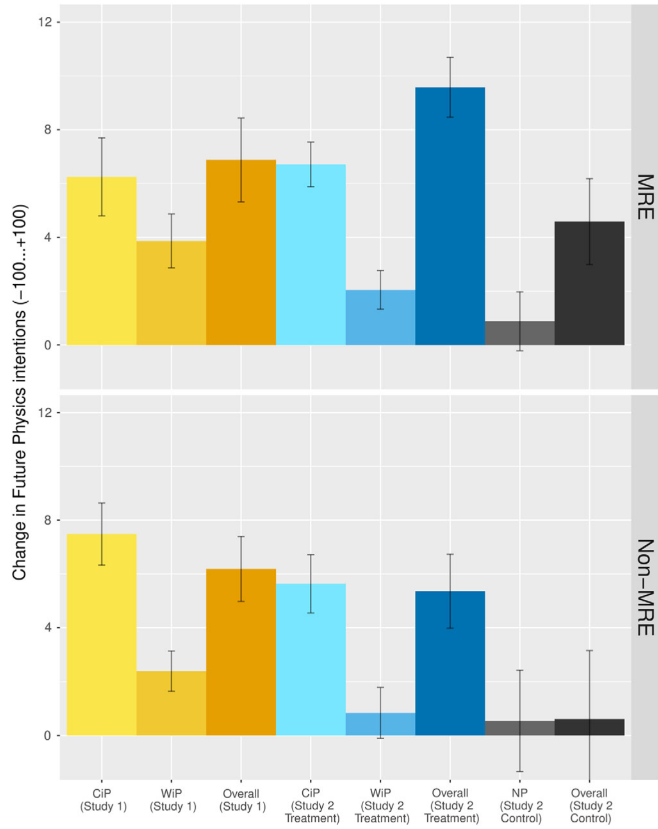

FIG. 4. Summary of results including per-lesson results (Careers in Physics, Women in Physics, and Notable Physicists lessons, in addition to the overall impact on students), MRE-identified students on top, non-MRE identified students on bottom. For each lesson (CiP, WiP, or NP), the bar represents the immediate post- minus immediate prescore. For the overall measure, the bar represents the latest post- minus prescore. Error bars represent standard errors.

their overall future physics intentions in this group). Similarly, MRE and non-MRE students saw a positive gain for each of the CiP and WiP lesson in both Study 1 and Study 2 treatment (in all but one case, statistically significant) while MRE and non-MRE students saw no gain from the NP lesson (though MRE students did see a statistical gain in their overall future physics intentions in this group).

TABLE XII. Summary of paired t tests based on gender group (Female-identified or Nonfemale-identified) for Study 1 and Study 2 (treatment and control groups). Each row represents the paired t test after combining 100 multiply imputed datasets according to Rubin's rules.

Condition	Lesson	Group	Gain [95% CI]	Sig.	Std. Err.	t	Cohen's d [95% CI]	N_{avg}
Study 1	CiP	Female	7.7 [5.4, 9.9]	$p < 0.001$	1.12	6.73	0.35 [0.25, 0.45]	382
Study 1	WiP	Female	4.4 [2.6, 6.1]	$p < 0.001$	0.97	4.97	0.26 [0.16, 0.36]	382
Study 1	Overall	Female	8.2 [5.3, 11.1]	$p < 0.001$	1.44	5.61	0.29 [0.19, 0.39]	382
Study 1	CiP	Non-F	6.4 [3.2, 9.6]	$p < 0.001$	1.63	3.96	0.19 [0.09, 0.28]	441
Study 1	WiP	Non-F	1.8 [0.2, 3.4]	$p = 0.03$	0.84	2.18	0.10 [0.01, 0.20]	441
Study 1	Overall	Non-F	5.0 [2.5, 7.4]	$p < 0.001$	1.27	3.97	0.19 [0.09, 0.28]	441

(Table continued)

TABLE XII. *(Continued)*

Condition	Lesson	Group	Gain [95% CI]	Sig.	Std. Err.	t	Cohen's d [95% CI]	N_{avg}
Study 2 treatment	CiP	Female	6.4 [4.5, 8.3]	$p < 0.001$	0.95	6.73	0.27 [0.19, 0.36]	604
Study 2 treatment	WiP	Female	1.4 [−0.1, 3.0]	$p = 0.069$	0.79	1.82	0.07 [−0.1, 0.16]	604
Study 2 treatment	Overall	Female	8.6 [6.1, 11.0]	$p < 0.001$	1.22	6.98	0.29 [0.20, 0.37]	604
Study 2 treatment	CiP	Non-F	6.4 [4.6, 8.2]	$p < 0.001$	0.91	7.08	0.29 [0.21, 0.38]	574
Study 2 treatment	WiP	Non-F	2.0 [0.3, 3.7]	$p = 0.025$	0.88	2.26	0.09 [0.01, 0.18]	574
Study 2 treatment	Overall	Non-F	8.2 [5.8, 10.6]	$p < 0.001$	1.22	6.74	0.28 [0.20, 0.36]	574
Study 2 control	NP	Female	0.1 [−2.6, 2.8]	$p = 0.95$	1.33	0.07	0.1 [−0.16, 0.17]	148
Study 2 control	Overall	Female	2.7 [−0.7, 6]	p = 0.12	1.66	1.56	0.13 [−0.04, 0.30]	148
Study 2 control	NP	Non-F	1.4 [−1.2, 4.0]	p = 0.30	1.34	1.04	0.08 [−0.07, 0.22]	183
Study 2 control	Overall	Non-F	4.4 [0.5, 8.3]	$p = 0.03$	2.00	2.24	0.16 [0.02, 0.31]	183

TABLE XIII. Summary of paired t tests based on MRE-identification (MRE-identified or non-MRE-identified) for Study 1 and Study 2 (treatment and control groups). Each row represents the paired t test after combining 100 multiply imputed data sets according to Rubin's rules.

Condition	Lesson	Group	Gain [95% CI]	Sig.	Std. Err.	t	Cohen's d [95% CI]	N_{avg}
Study 1	CiP	MRE	6.2 [3.3, 9.2]	$p < 0.001$	1.45	4.23	0.24 [0.13, 0.35]	334
Study 1	WiP	MRE	3.9 [1.9, 5.9]	$p < 0.001$	0.99	3.81	0.21 [0.10, 0.32]	334
Study 1	Overall	MRE	6.9 [3.7, 10.0]	$p < 0.001$	1.56	4.32	0.24 [0.13, 0.35]	334
Study 1	CiP	Non-M	7.5 [5.2, 9.7]	$p < 0.001$	1.16	6.55	0.29 [0.20, 0.38]	489
Study 1	WiP	Non-M	2.4 [0.9, 3.9]	$p = 0.001$	0.75	3.23	0.14 [0.06, 0.23]	489
Study 1	Overall	Non-M	6.2 [3.8, 8.5]	$p < 0.001$	1.21	5.19	0.23 [0.14, 0.32]	489
Study 2 treatment	CiP	MRE	6.7 [5.1, 8.3]	$p < 0.001$	0.83	8.07	0.28 [0.21, 0.35]	842
Study 2 treatment	WiP	MRE	2.0 [0.6, 3.5]	$p = 0.005$	0.72	2.84	0.10 [0.03, 0.17]	842
Study 2 treatment	Overall	MRE	9.6 [7.4, 11.8]	$p < 0.001$	1.11	8.59	0.30 [0.23, 0.37]	842
Study 2 treatment	CiP	Non-M	5.6 [3.5, 7.8]	$p < 0.001$	1.08	5.24	0.28 [0.18, 0.39]	336
Study 2 treatment	WiP	Non-M	0.8 [−1.0, 2.7]	$p = 0.37$	0.95	0.89	0.05 [−0.06, 0.16]	336
Study 2 treatment	Overall	Non-M	5.4 [2.7, 8.1]	$p < 0.001$	1.38	3.92	0.21 [0.11, 0.32]	336
Study 2 control	NP	MRE	0.9 [−1.3, 3.0]	p = 0.42	1.10	0.81	0.05 [−0.07, 0.17]	252
Study 2 control	Overall	MRE	4.6 [1.5, 7.7]	$p = 0.004$	1.60	2.90	0.18 [0.06, 0.30]	252
Study 2 control	NP	Non-M	0.5 [−3.4, 4.4]	$p = 0.78$	1.9	0.28	0.03 [−0.20, 0.27]	79
Study 2 control	Overall	Non-M	0.6 [−4.6, 5.9]	p = 0.82	2.6	0.23	0.03 [−0.20, 0.26]	79

[1] P. J. Mulvey and S. Nicholson, *Physics Bachelor's Degrees: 2018* (American Institute of Physics, College Park, MD, 2020).
[2] C. de Brey, A. Zhang, and S. Duffy, Digest of Education Statistics 2020 (NCES 2022-009), National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC. (2022).
[3] C. P. Foley, Status of women in physics in Australia, AIP Conf. Proc. **1517**, 72 (2013).
[4] A. M. Porter and R. Ivie, *Women in physics and astronomy, 2019* (AIP Statistical Research Center, College Park, MD, 2019).
[5] M. Kilfoil, J. McKenna, A. Predoi-Cross, and M. Steinitz, Women in physics in Canada: Progress and shortcomings, AIP Conf. Proc. **795**, 103 (2005).
[6] M. Mohsen, H. Hosni, H. Mohamed, A. Gadalla, H. Kahil, and H. Hashem, Egyptian women in physics: Progress and challenges, AIP Conf. Proc. **1697**, 060015 (2015).
[7] A. Irajizad, F. Roshani, and D. Izadi, Improving the status of Iranian women in physics, AIP Conf. Proc. **1697**, 060024 (2015).
[8] A. Sommerfeld, A. Krankl, and B. Sandow, Area of actions: Equal opportunities for women in physics in Germany, AIP Conf. Proc. **1697**, 060020 (2015).
[9] S. Moshfeghyeganeh and Z. Hazari, Effect of culture on women physicists' career choice: A comparison of Muslim majority countries and the West, Phys. Rev. Phys. Educ. Res. **17**, 010114 (2021).
[10] L. Hong and S. E. Page, Groups of diverse problem solvers can outperform groups of high-ability problem solvers, Proc. Natl. Acad. Sci. U.S.A. **101**, 16385 (2004).
[11] J. B. Guerra, A. Calabrese Barton, E. Tan, H. Kang, and C. Brecklin, Kay's coat of many colors: Out-of-school figured worlds and urban girls' engagement with science, *Identity Construction and Science Education Research* (Brill Sense, Leiden, Netherlands, 2012), pp. 43–60.
[12] D. G. Solórzano and T. J. Yosso, Critical race methodology: Counter-storytelling as an analytical framework for education research, Qual. Inq. **8**, 23 (2002).
[13] *Considering Counternarratives: Narrating, Resisting, Making Sense*, edited by M. Bamberg and M. Andrews (John Benjamins Publishing Company, Philadelphia, 2004).
[14] Z. Hazari, D. Chari, G. Potvin, and E. Brewe, The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates, J. Res. Sci. Teach. **57**, 1583 (2020).
[15] The AIP National Task Force to Elevate African American Representation in Undergraduate Physics & Astronomy (TEAM-UP), *THE TIME IS NOW: Systemic Changes to Increase African Americans with Bachelor's Degrees in Physics and Astronomy*, (American Institute of Physics, College Park, MD 2020).
[16] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).
[17] S. Kanim and X. C. Cid, Demographics of physics education research, Phys. Rev. Phys. Educ. Res. **16**, 020106 (2020).
[18] L. Bian, S. J. Leslie, and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, Science **355**, 389 (2017).
[19] S. J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).
[20] U. Kessels, Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects, Eur. J. Psychol. Educ. **20**, 309 (2005).
[21] C. Leaper, T. Farkas, and C. S. Brown, Adolescent girls' experiences and gender-related beliefs in relation to their motivation in math/science and English, J. Youth Adolescence **41**, 268 (2012).
[22] C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman, Science faculty's subtle gender biases favor male students, Proc. Natl. Acad. Sci. U.S.A. **109**, 16474 (2012).
[23] G. Potvin, Z. Hazari, R. H. Tai, and P. M. Sadler, Unraveling bias from student evaluations of their high school science teachers, Sci. Educ. **93**, 827 (2009).
[24] K. Due, Who is the competent physics student? A study of students' positions and social interaction in small-group discussions, Cult. Studies Sci. Educ. **9**, 441 (2014).
[25] U. Kessels, M. Rau, and B. Hannover, What goes well with physics? Measuring and altering the image of science, Br. J. Educ. Psychol. **76**, 761 (2006).
[26] M. Bruun, S. Willoughby, and J. L. Smith, Identifying the stereotypical who, what, and why of physics and biology, Phys. Rev. Phys. Educ. Res. **14**, 020125 (2018).
[27] C. Riegle-Crumb, C. Moore, and A. Ramos-Wada, Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity, Sci. Educ. **95**, 458 (2011).
[28] P. M. Sadler, G. Sonnert, Z. Hazari, and R. H. Tai, Stability and volatility of STEM career interest in high school: A gender study, Sci. Educ. **96**, 411 (2012).
[29] Z. Hazari, E. Brewe, R. M. Goertzen, and T. Hodapp, The importance of high school physics teachers for female students' physics identity and persistence, Phys. Teach. **55**, 96 (2017).
[30] R. H. Tai, C. Q. Liu, A. V. Maltese, and X. Fan, Planning early for careers in science, Science **312**, 1143 (2006).
[31] A. V. Maltese and R. H. Tai, Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students, Sci. Educ. **95**, 877 (2011).
[32] M. Dew, J. Perry, L. Ford, W. Bassichis, and T. Erukhimova, Gendered performance differences in introductory physics: A study from a large land-grant university, Phys. Rev. Phys. Educ. Res. **17**, 010106 (2021).
[33] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, Sci. Educ. **91**, 847 (2007).
[34] J. S. Hyde, S. M. Lindberg, M. C. Linn, A. B. Ellis, and C. C. Williams, Gender similarities characterize math performance, Science **321**, 494 (2008).
[35] T. L. Killpack and L. C. Melón, Toward inclusive STEM classrooms: What personal role do faculty play?, CBE Life Sci. Educ. **15**, es3 (2016).

[36] S. M. Jackson, A. L. Hillard, and T. R. Schneider, Using implicit bias training to improve attitudes toward women in STEM, Social Psychol. Educ. **17**, 419 (2014).

[37] C. Pritlove, C. Juando-Prats, K. Ala-Leppilampi, and J. A. Parsons, The good, the bad, and the ugly of implicit bias, Lancet **393**, 502 (2019).

[38] S. N. C. Liu, S. E. Brown, and I. E. Sabat, Patching the "leaky pipeline": Interventions for women of color faculty in STEM academia, Archives Sci. Psychol. **7**, 32 (2019).

[39] G. L. Cohen, J. Garcia, and N. Apfel, A. Master, Reducing the racial achievement gap: A social-psychological intervention, Science **313**, 1307 (2006).

[40] G. M. Walton and G. L. Cohen, A brief social-belonging intervention improves academic and health outcomes of minority students, Science **331**, 1447 (2011).

[41] G. L. Cohen and D. K. Sherman, The psychology of change: Self-affirmation and social psychological intervention, Ann. Rev. Psychol. **65**, 333 (2014).

[42] T. O. Jackson and G. S. Boutte, Exploring culturally relevant/responsive pedagogy as praxis in teacher education, New Edu. **14**, 87 (2018).

[43] Z. Hazari, G. Potvin, R. M. Lock, F. Lung, G. Sonnert, and P. M. Sadler, Factors that affect the physical science career interest of female students: Testing five common hypotheses, Phys. Rev. Phys. Educ. Res. **9**, 020115 (2013).

[44] A. B. Diekman, E. R. Brown, A. M. Johnston, and E. K. Clark, Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers, Psychol. Sci. **21**, 1051 (2010).

[45] H. Cheng, G. Potvin, G. R. Khatri, L. H. Kramer, R. M. Lock, and Z. Hazari, Examining physics identity development through two high school interventions, in *Proceedings of PER Conf. 2018, Washington, DC*, 10.1119/perc.2018.pr.Cheng.

[46] P. Sabouri, Z. Hazari, R. Khatri, and B. Barnett Dreyfuss, Considerations for Inclusive and Equitable Design: The Case of STEP UP Counternarratives in High School Physics, Phys. Teach. **60**, 740 (2022).

[47] http://stepupphysics.org.

[48] A. Madsen, S. B. McKagan, and E. C. Sayre, How physics instruction impacts students' beliefs about learning physics: A meta-analysis of 24 studies, Phys. Rev. ST Phys. Educ. Res. **11**, 010115 (2015).

[49] D.-L. Stewart, Racially minoritized students at U.S. four-year institutions, J. Negro Educ. **82**, 184 (2013).

[50] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, Pearson New International Edition (Pearson Education Limited, Essex, 2014).

[51] D. L. Putnick and M. H. Bornstein, Measurement invariance conventions and reporting: The state of the art and future directions for psychological research, Developmental Rev. **41**, 71 (2016).

[52] D. B. Rubin, Multiple imputation after 18+ years, J. Am. Stat. Assoc. **91**, 473 (1996).

[53] R Core Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, 2021).

[54] J. Honaker, G. King, and M. Blackwell, Amelia II: A program for missing data, J. Stat. Softw. **45**, 1 (2011).

[55] P. T. von Hippel, How many imputations do you need? A two-stage calculation using a quadratic rule, Sociological Methods Res. **49**, 699 (2020).

[56] Z. Hazari, G. Sonnert, P. M. Sadler, and M.-C. Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, J. Res. Sci. Teach. **47**, 978 (2010).