PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH **18,** 020111 (2022)

Longitudinal analysis of women and men's motivational beliefs in a two-semester introductory physics course sequence for students on the bioscience track

Sonja Cwik and Chandralekha Singh
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

(Received 3 December 2021; accepted 10 June 2022; published 22 August 2022)

Societal stereotypes and biases about who belongs in physics and who can excel in it can impact the physics beliefs, including their self-efficacy, interest, and identity, e.g., of women in physics courses. Exploring these beliefs longitudinally and analyzing how different beliefs predict students' physics identity are important for developing a better understanding of the experiences of women who are majoring in science, technology, engineering, and mathematics in their mandatory physics courses and making the learning environment equitable and inclusive. We analyzed beliefs of students longitudinally in introductory physics 1 and physics 2, which are mandatory for students on the bioscience track, and used structural equation modeling (SEM) and a physics identity framework to investigate whether the relation between gender and physics identity was mediated by self-efficacy, interest, and perceived recognition by others. Although women are not underrepresented in this two-semester physics course sequence for students on the bioscience track, societal stereotypes and biases internalized by women over their lifetime can impact their beliefs about physics when they enter these physics classes and gender gaps can persist if the physics learning environment is not equitable and inclusive. Our findings show a gender gap in beliefs disadvantaging women throughout the physics course sequence. Additionally, the SEM model of physics identity shows that physics perceived recognition by others plays a central role in predicting students' physics identity throughout the two-semester course sequence.

DOI: 10.1103/PhysRevPhysEducRes.18.020111

I. INTRODUCTION

Domain specific beliefs, e.g., self-efficacy and identity, in science, technology, engineering, and mathematics (STEM) domains can influence students' participation and persistence in those fields [1–13]. For example, students are more likely to take courses or pursue a career in science if they have higher competency beliefs or self-efficacy [4,5,14–16], display higher interest in science [17], or have a higher science identity [18–20]. However, a gender gap disadvantaging women in student domain specific beliefs [21–27] has been observed in many STEM courses, including physics courses. Most of the research in physics has focused on courses in which women are underrepresented. However, in a study of introductory bioscience courses, in which women made up 60% of the course, women participated less and did not perform as well as men [28]. These trends may signify gender inequities in STEM learning environments even when women are not underrepresented. Since physics is a discipline with one of the worst stereotypes about who belongs in it and can succeed in it [29], it is particularly important to investigate inequitable gender gaps in beliefs (including self-efficacy, interest, recognition by others, and identity) in physics courses in which women are not underrepresented.

Domain-specific identity has been shown to play an important role, particularly in students' participation in courses and their professional choices [18,30–34]. Identity in this context refers to identifying with an academic domain, like physics, or a student's view about whether they see themselves as a "physics person" [34–37]. Prior studies have shown that it can be more difficult for women to form a physics identity than men [37–40]. However, most of the prior studies concerning physics identity and factors that influence it have been conducted in classes in which women are underrepresented. It is important to investigate physics identity in a variety of classes and contexts since there can be differences in each given context. For example, in one study on female undergraduate students, their sense of belonging predicted physics identity for upper-level students but not for introductory students [41]. Therefore, we investigated physics identity and student beliefs such as self-efficacy, interest, and perceived recognition (which have been shown to be related to physics identity in other contexts) in algebra-based physics courses for students on the bioscience track in which women are not underrepresented.

One factor hypothesized to influence physics identity is self-efficacy. Self-efficacy in a particular academic domain is defined as a person's belief that they can succeed in a particular task, activity, or course related to that domain [42,43]. According to Bandura's social cognitive theory, self-efficacy may be derived from four sources: mastery experiences, vicarious learning experiences, social persuasion experiences, and emotional state [42]. Mastery experience means that experiences with successful completion of a task should have a strong positive influence on one's confidence to complete a similar task. Vicarious learning experiences occur when observing someone else's success on a task, which can influence students' own belief in their ability to perform a similar task. Social persuasion experiences mean that verbal suggestions from others such as words of encouragement can positively impact one's self-efficacy. Lastly, one's emotional state can act as a mediating source to amplify or undermine one's confidence in one's ability. Sawtelle *et al.* found that predicting the probability of passing an introductory physics course for women relies primarily on the vicarious learning experiences source while it primarily relies on mastery experiences for men [44]. Students' self-efficacy has been shown to impact students' engagement, learning, and persistence in science courses as well as contribute to students' science identity [14–16,25,44–56]. For example, when tackling difficult problems, students with high self-efficacy tend to view the problems as challenges that can be overcome, whereas people with low self-efficacy tend to view them as personal threats to be avoided [42]. However, in introductory physics courses in which women are underrepresented, studies have found a gender gap in self-efficacy favoring men that widens by the end of the course even in interactive engagement courses [2,25].

Another factor hypothesized to influence physics identity is interest. According to the four-phase model of interest development [57], people's interest in a field is triggered and maintained by external factors first and then becomes a sustained individual interest. Interest in a particular discipline may affect students' perseverance, persistence, and achievement [14,53,58–61]. One study showed that changing the curriculum to stimulate the interest of the female students helped improve all of the students' understanding at the end of the year [62]. Within expectancy value theory, interest and competency beliefs (closely related to self-efficacy) are connected to constructs that predict students' academic outcomes and career expectations [63]. We focus on intrinsic interest, which is an individual's personal interest and enjoyment in engaging with a topic.

The third factor hypothesized to influence physics identity—perceived recognition—has been shown to be an important factor [64,65]. Our prior individual interviews suggest that students' perceived recognition as a physics person by instructors and teaching assistants (TAs) predicts their self-efficacy and interest in physics (see, e.g., Refs. [26,66,67]). When women enter STEM fields, they do not feel validated and recognized as often as men [68–70]. In a study on students' perception of support, teacher support was more strongly linked to the motivation and engagement of girls than boys [64]. Hazari *et al.* found that in high school physics classes, students' physics identity, predicted by students' recognition as a physics person by others, was influenced by teachers' social cues, or how teachers' actions obscured social boundaries between teachers and students [71]. In another study [72], students who received praise for their intelligence were more likely to have a fixed mindset and lower levels of task persistence, enjoyment, and performance than students who received praise for their effort. Prior studies have also shown that female students are not recognized appropriately even before they enter college [38,73,74]. For example, one of the stereotypical views in science is that it is for high achievers or naturally gifted students [38], and in general, being a genius or exceptionally smart is attributed to boys [75].

II. THEORTICAL FRAMEWORK

In this study, we use a physics identity framework to investigate whether the relation between gender and physics identity was mediated by self-efficacy, interest, and perceived recognition by others longitudinally in introductory physics 1 and physics 2, which are mandatory for students on the bioscience track. We build on the physics identity framework developed by Hazari *et al.*, which adapted the science identity framework by Carlone and Johnson [30]. The science identity framework by Carlone and Johnson [30] includes three dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others"). Hazari *et al.* modified the framework specifically for physics. "Competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a physics person. Lastly, a fourth dimension, interest, was added to the framework since students can have highly varying levels of interest in physics [35,76]. In recent studies by Hazari *et al.* of introductory students, performance and competence are combined into one variable [41]. In a slightly reframed version of Hazari *et al.*'s physics identity framework by Kalender *et al.* [77] that we use, performance or competence was framed as self-efficacy (closely related to competency belief). Figure 1 shows the schematic representation of the model used here. In this framework, students' physics identity has been shown to be influenced by their self-efficacy, interest, and perceived recognition [39,40,44,78,79].

Stereotypes and biases about who belongs in physics and can excel in it are one of the absolute worst out of all of the STEM disciplines [29] and this could be impacting women even in physics courses in which they are not underrepresented. Boys and girls are exposed to these fixed

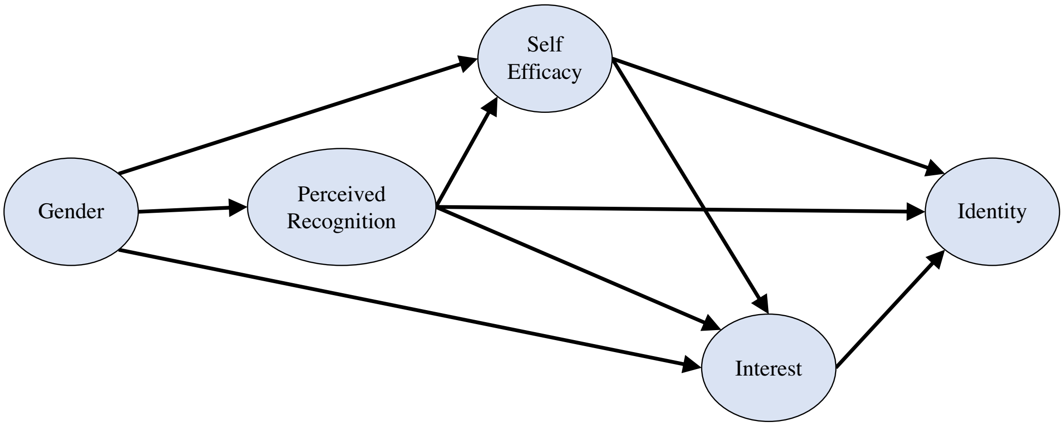

FIG. 1. Schematic representation of the model based on the theoretical framework. From left to right, all possible paths were considered. Some, but not all, of the regression paths are shown.

intelligence views starting from an early age [74]. Prior studies have found that by the age of six, girls are less likely than boys to believe they are "really really smart" and less likely to choose activities that are made for "brilliant people" [74]. As these students get older, these stereotypes can negatively impact women. One study found that science faculty members in biological and physical sciences exhibit biases against female students by rating a male student with an identical resume to a female student significantly more competent and being more willing to hire the male student and pay him more [80]. All of these stereotypes and biases can influence female students' perception of their ability to engage in physics problem solving even before they enter the physics classroom. Therefore, it is possible that although women are the majority in algebra-based physics courses for students on the bioscience track, these societal stereotypes can still influence their outcomes and identity in these courses.

Thus, our study focuses on the hypothesis that male and female students' physics beliefs including their physics identity may be different even in courses in which women are not underrepresented and the physics learning environment in this situation may also exacerbate the situation and make the gender gap worse. In the research presented here, we used structural equation modeling (SEM) and a physics identity framework to investigate whether the relation between gender and physics identity is mediated by self-efficacy, interest, and perceived recognition by others longitudinally in introductory physics 1 and physics 2, which are mandatory for students on the bioscience track. In the framework used, physics self-efficacy, interest, and perceived recognition can play important roles in predicting the physics identity. In general, these beliefs that predict physics identity (i.e., physics self-efficacy, interest, and perceived recognition) are related to each other. However, it is not clear in the context of a two-semester introductory physics course sequence in which a majority of the students are women, whether there are gender differences in these beliefs at the end of the first physics course and the end of the entire two-semester physics course sequence.

Therefore, we administered a validated survey about physics beliefs in algebra-based physics courses and used mediation analysis in SEM to investigate progression in physics identity and beliefs (self-efficacy, perceived recognition, and interest) that mediate the relation between gender and physics identity in introductory physics 1 and physics 2. We recognize that there can be multiple statistically equivalent SEM models with the same constructs mediating the relationship between gender and physics identity. We chose our particular statistically equivalent SEM model in which perceived recognition precedes self-efficacy and interest in which the potential instructional implications (positive recognition can play a role in supporting students' self-efficacy and interest) can empower instructors or teaching assistants to positively recognize students and make their classes more equitable and inclusive. Thus, our model puts recognition by instructors and TAs first before self-efficacy and interest. Figure 1 shows the schematic representation of the physics identity framework for this investigation since no gender moderation effects were found (identical to the one used for calculus-based introductory physics courses [77]). We fit the model with the data collected at the end of physics 1 and physics 2 and compare the predictive relationships among the constructs in the two courses. Our research questions are as follows for the study spanning the two-semester course sequence:

RQ1 Are there gender differences in the student physics beliefs (self-efficacy, interest, perceived recognition, and identity) and do they change from physics 1 to physics 2 for bioscience students?

RQ2 Can gender differences in students' physics identity at the end of physics 1 and physics 2 for bioscience students be explained with gender differences in physics perceived recognition, self-efficacy, and interest?

RQ3 How do the predictive relationships between the student beliefs such as self-efficacy, interest, and perceived recognition mediate the relation between gender and physics identity and how do they change between physics 1 and physics 2 for bioscience students?

III. METHODOLOGY

A. Participants

We analyzed data from a physics belief survey of 563 matched students (the same students took the survey at the end of physics 1 and 2) that was administered at a large public research university at the beginning and end of two semesters of introductory algebra-based physics 1 and 2 over the course of two years. The lecture-based physics courses were taught by 5 male instructors and 1 female instructor in physics 1 and 3 male instructors in physics 2. The classes consisted of 3 h of lecture per week taught by the instructor and 1 h of recitation per week taught by a teaching assistant (TA) (in which, typically, the students could ask questions about the homework or course material and complete group work). The classes were similar in terms of their grading policy in that students' grades heavily consisted of 2–3 midterm exams and a final exam involving primarily quantitative problems. However, using hierarchical modeling, we found the intraclass correlation coefficient (ICC) [81], which measures the proportion of the variance in the different constructs investigated in our model between instructors to be at most 4%. This is below 10%, the usual threshold cited for warranting the use of multilevel models and thus we grouped all the students together, regardless of instructor. We also note that the difference between the course taught by the female instructor and male instructors was comparable to those between different male instructors.

These introductory physics courses are typically taken by students on the bioscience track in their junior or senior year of undergraduate studies, with most students expressing a desire to pursue future careers in health professions. The university provided demographic information such as age, gender, and ethnic or racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. The gender data provided by the university include only binary options of "male" and "female." We recognize that gender is a sociocultural and a nonbinary construct, however, we are limited to the binary data in this study. Based on the university data; the participants were 36% male and 64% female students.

B. Instrument validity

The survey items were constructed from items validated by others [82–84] and revalidated in our own context using one-on-one student interviews [23], exploratory factor analysis (EFA), confirmatory factor analysis (CFA) [85], analyzing the Pearson correlation between different constructs [85], and using Cronbach's alpha [86]. At the beginning of the survey, students were instructed to answer the questions on the survey with regard to the physics course they were in. The survey items asked about different beliefs at the beginning and end of the course. These constructs included students' physics identity (1 item), self-efficacy (4 items), interest (4 items), and perceived recognition (3 items). The *physics self-efficacy* questions measured students' confidence in their ability to answer and understand physics problems [36,82,83,87,88]. The self-efficacy questions are mainly drawing on students' mastery experiences. The *interest in physics* questions measured students' enthusiasm and curiosity to learn physics and ideas related to physics [87]. The *perceived recognition* questions measured the extent to which the students thought that other people see them as a physics person [36]. The *physics identity* question evaluated whether the students see themselves as a physics person [34,36,89]. The physics identity instrument only included one question, which is consistent with past studies since it has been challenging to make other questions that factor in this category in exploratory factor analysis [34,40,90,91].

The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [92]. A lower score is indicative of a negative endorsement of the survey construct while a higher score is related to a positive belief of the construct. Some of the questions were reverse coded (e.g., I feel like an outsider in this class). A CFA was conducted to establish a measurement model for the constructs and used in SEM. The square of CFA factor loadings (lambda) indicates the fraction of variance explained by the factor. The model fit indices were good and all of the factor loadings (lambda) were above 0.50, which indicates good loadings [85]. The results of the CFA model are shown in Table I. The survey questions for each construct and factor loadings for each question from the CFA are given in Table I. The Cronbach's alpha was used to measure the internal consistency of the items. The Cronbach's alpha is 0.79 for the self-efficacy questions, 0.76 for interest questions, and 0.89 for perceived recognition questions, which are considered reasonable [86].

Zero-order pairwise Pearson correlations of the average values of each construct are given in Table II. Pearson's r values signify the strength of the relationship between variables. The intercorrelations vary in the strength of their correlation, but none of the correlations are so high that the constructs cannot be separately examined. The only high intercorrelation was the value between physics identity and perceived recognition (0.83). Perceived recognition questions ask about external identity, whereas physics identity asks about internal identity so there tends to be a high correlation between the constructs; however, the correlation is low enough that they can be considered separate constructs.

TABLE I. Survey questions corresponding to each of the constructs, along with factor loadings from the confirmatory factor analysis for all matched students ($N = 563$) in physics 1 and physics 2. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [92]. The rating scale for most of the self-efficacy and interest questions was: NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree. All p values (for the significance test of each item loading) are $p < 0.001$.

	Lambda	
Construct and item	Physics 1	Physics 2
Physics identity		
I see myself as a physics person	1.00	1.00
Physics self-efficacy		
I am able to help my classmates with physics in the laboratory or recitation	0.52	0.61
I understand concepts I have studied in physics	0.67	0.72
If I study, I will do well on a physics test	0.77	0.78
If I encounter a setback in a physics exam, I can overcome it	0.77	0.72
Physics interest		
I wonder about how physics works	0.68	0.70
In general, I find physics[a]	0.79	0.80
I want to know everything I can about physics	0.79	0.78
I am curious about recent discoveries in physics	0.62	0.74
Physics perceived recognition		
My family sees me as a physics person	0.86	0.90
My friends see me as a physics person	0.91	0.91
My physics instructor and/or TA sees me as a physics person	0.72	0.69

[a] the rating scale for this question was very boring, boring, interesting, very interesting.

C. Analysis

Initially, we compared female and male students' mean scores for each construct for statistical significance using t tests and for the effect sizes using Cohen's d [85]. In order to investigate the effect of the introductory physics course on the gender differences in student beliefs, multiple one-way analysis of covariance (ANCOVAs) were conducted to determine whether there was a statistically significant gender difference in each post belief controlling for the prebelief. We note that the assumptions were met to carry out the ANCOVA including normality and homogeneity conditions.

To quantify the significance and relative strength of our framework's links, we used strict structural equation modeling as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [93]. SEM is a statistical method consisting of two parts that are completed simultaneously; a measurement part which consists of CFA and a structural part which consists of path analysis. Path analysis can be considered an extension of multiple regression and conducts several multiple regressions simultaneously between variables in one estimation model in addition to the measurement part involving CFA. This is an improvement over multiple regression since it allows us to calculate the overall goodness of fit and allows for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. We report model fit for SEM by using the comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [94].

TABLE II. Pearson intercorrelations are given between all the predictors and outcomes for the postsurvey in physics 2. All p values <0.001.

	Pearson correlation coefficient			
Observed variable	1	2	3	4
1. Perceived recognition	⋯	⋯	⋯	⋯
2. Self-efficacy	0.56	⋯	⋯	⋯
3. Interest	0.57	0.59	⋯	⋯
4. Physics identity	0.83	0.56	0.54	⋯

Initially, we performed gender moderation analysis by conducting multigroup SEM, i.e., the model estimates were performed separately for men and women to check whether any of the relations between variables show differences across gender by using "lavaan" [95]. In particular, our moderation analysis was similar to our mediation model in Fig. 1 except there was no link from gender, instead, multigroup SEM was performed separately for women and men simultaneously.

In order to explain what moderation analysis means, we start with a simple moderation analysis example. In a simple moderation analysis involving the predictive relation between only two variables, the predictive relationship (the regression path) between those two variables is tested for two or more different groups (e.g., men and women) simultaneously. If the predictive relationship is different for the groups [i.e., the values of the regression coefficients (β) are not the same for the correlation between the two constructs for different groups], then there is a moderation effect in the model. For example, in a study focusing on how smoking predicts lung cancer, if there was a moderation effect by gender, the predictive relation (regression coefficient) between smoking and lung cancer would be different for women and men. However, if the regression coefficients for how smoking predicts lung cancer were exactly the same for women and men, then there is no moderation by gender and one can just focus on mediation analysis by gender (in other words, we need not separately calculate the regression coefficients for women and men since they are equal, and we can introduce gender as an additional categorical variable in the model to do gender mediation analysis).

When the model is more complex than the preceding example of smoking and lung cancer as in our SEM model (which has a measurement part involving CFA and a structural part involving path analysis), checking to make sure there are no gender moderation effects involves checking that there are no gender moderation effects for both the measurement and structural parts. For the measurement part, to check for measurement invariance in each step of gender moderation analysis, we fixed different elements of the measurement part of the model to equality across gender and compared the results to the previous step when they were allowed to vary between groups (i.e., for women and men) separately using the likelihood ratio test [95]. A nonsignificant p value at each step indicates that the fit of this model is not appreciably worse than that of the model in the previous step, so the more restrictive invariance hypothesis (when the parameters are set to the same values for women and men) is retained. Therefore, setting those different elements of the measurement part of the model to equality across gender is valid, which means that estimates are not statistically significantly different across groups (i.e., women and men).

First, we tested for “weak” measurement invariance, which determines if survey items have similar factor loadings for men and women. We compared two models, one in which the factor loadings (which represent the correlation between each item and its corresponding construct) for women and men were predicted independently, and the other in which the factor loadings were forced to be equal between the groups (i.e., for women and men). Next, we tested for “strong” measurement invariance, which determines if survey items have similar factor loadings as well as similar intercepts [95] for men and women. Similar to weak invariance testing, we compared the models in which these factors were allowed to vary between groups separately for women and men and when they were set equal for women and men. If measurement invariance passes the weak and strong invariance test, i.e., there is no statistically significant difference between models when those parameters for women and men are set equal, then we must check for differences in the path analysis part, i.e., regression coefficients (β) among different latent variables in the model between women and men.

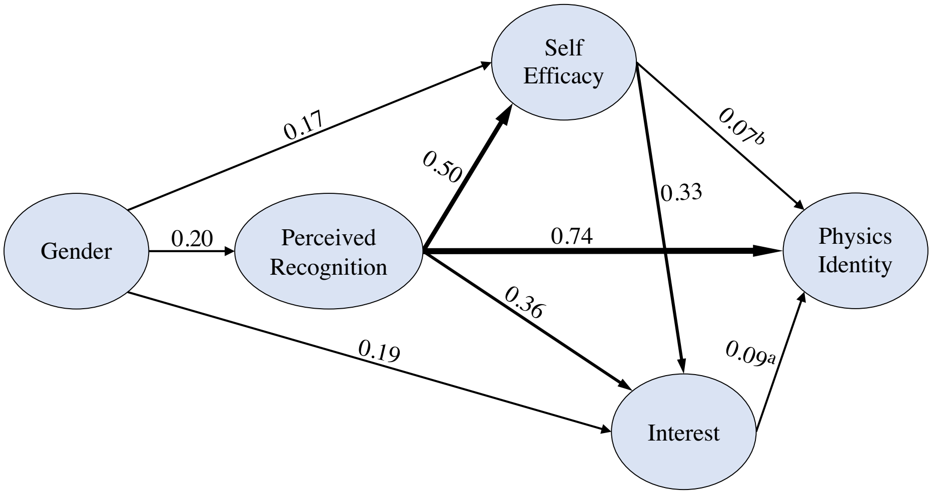

FIG. 2. Result of the path analysis part of the SEM with the relationship between gender and physics identity being mediated through perceived recognition, self-efficacy, and interest for physics 1. The gender variable was coded as 0 for women and 1 for men. The line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. All p values for β are indicated by no superscript for $p < 0.001$ “a” for $0.001 \leq p < 0.05$, and “b” for $p \geq 0.05$ (nonsignificant). Gender does not directly predict physics identity.

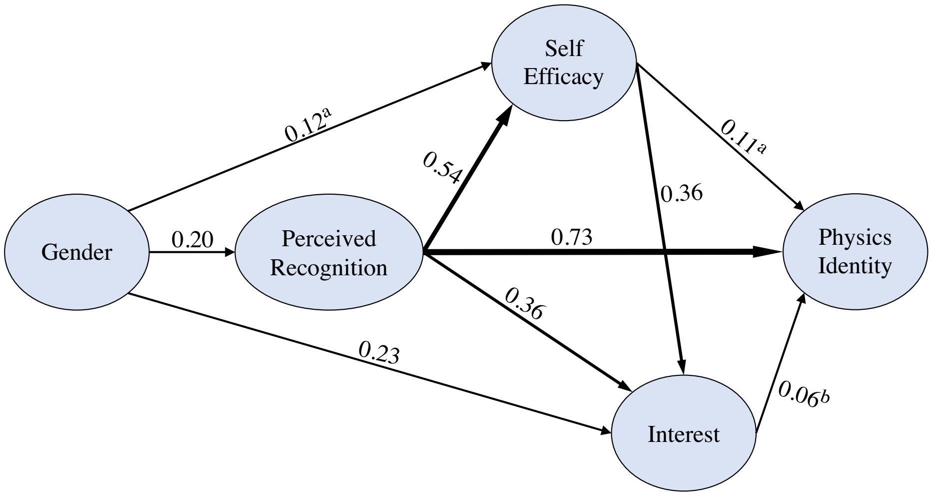

FIG. 3. Result of the path analysis part of the SEM with the relationship between gender and physics identity being mediated through perceived recognition, self-efficacy, and interest for physics 2. The gender variable was coded as 0 for women and 1 for men. The line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. All p values for β are indicated by no superscript for $p < 0.001$ "a" for $0.001 \leq p < 0.05$, and "b" for $p \geq 0.05$ (nonsignificant). Gender does not directly predict physics identity.

This is because differences between the groups could occur at the factor (latent variable) level in regression coefficients (β).

Similar to "weak" and "strong" measurement invariance for the measurement part, when testing moderation effect in path analysis, the predictive relationship (regression path) between two variables is tested for the two groups (e.g., women and men) simultaneously. If the predictive relationship is different for the groups (i.e., the values of the regression coefficients β are not the same for the predictive relationship between the two constructs for women and men), then there is a gender moderation effect in the model. If moderation does not show differences by gender in any of these steps (measurement invariance holds and testing for regression coefficients shows that they can be set equal for women and men), we can utilize a gender mediation model (see Fig. 1). In other words, we can interpret our model the same way for both men and women, and any gender differences can be modeled using a separate gender variable as in Fig. 1.

In our multigroup SEM model, we found a nonsignificant p value in each step, and thus measurement invariance holds and the regression coefficients for women and men can be set equal, i.e., there are no moderation effects by gender (for men and women) in our models. Thus, we concluded that our SEM model can be interpreted similarly for men and women and we can use gender mediation analysis (instead of doing moderation by gender). Therefore, we tested the theoretical model in mediation analysis, using gender as a variable (1 for male and 0 for female) directly predicting items to examine the resulting structural paths between constructs (a schematic representation of the path analysis for the gender mediation model is shown in Fig. 1). In the mediation analysis, if there are paths from gender to any of the constructs as we found in our results (Figs. 2 and 3) discussed in the next section, it implies that women and men did not have the same average value for those constructs controlling for all constructs to the left. However, it is important to note that all of the item factor loadings and regression coefficients between the constructs are the same for women and men (as found from the gender moderation analysis which preceded the mediation analysis).

IV. RESULTS

To answer RQ1 we analyzed the means of the beliefs in our model. According to Tables III and IV, women had statistically significantly lower mean values than men in all constructs in our model. Both women and men scored lowest in physics identity and women scored below 2. Although we focus on the beliefs at the end of the course, we have included data for students' beliefs at the beginning

TABLE III. Mean postpredictor and outcome values in physics 1 by gender as well as statistical significance (p values) and effect sizes (Cohen's d) by gender. 563 matched students are included with 361 women and 202 men. All p values <0.001.

	Mean		
Predictors and outcomes	Male	Female	Cohen's d
Perceived recognition	2.17	1.89	0.45
Self-efficacy	2.98	2.73	0.50
Interest	2.81	2.37	0.73
Physics identity	2.08	1.78	0.39

TABLE IV. Mean postpredictor and outcome values in physics 2 by gender as well as statistical significance (p values) and effect sizes (Cohen's d) by gender. 563 matched students are included with 361 women and 202 men. All p values <0.001.

	Mean		
Predictors and outcomes	Male	Female	Cohen's d
Perceived recognition	2.24	1.98	0.39
Self-efficacy	2.93	2.71	0.42
Interest	2.77	2.29	0.78
Physics identity	2.24	1.85	0.46

(pre) and end (post) of the course in Appendix A (note the presurvey in physics 1 did not include the identity or perceived recognition constructs and thus in order to match pre- and postsurvey responses, only data for the second year of survey administration for physics 1 are included). In addition, the percentages of men and women who selected each choice for each survey item are included in Appendix B.

The results of the ANCOVAs show that there was a significant effect of gender on students' postinterest [$F(1,383) = 1.85$ ($p = 0.002$)], post self-efficacy [$F(1,383) = 1.25$ ($p = 0.029$)], and postperceived recognition [$F(1,382) = 1.96$ ($p = 0.010$)] in physics 1 and on postinterest [$F(1,419) = 0.84$ ($p = 0.019$)] and postidentity [$F(1,419) = 1.92$ ($p = 0.014$)] in physics 2 controlling for the prevalues in each case.

To answer RQ2, we used SEM to investigate the relationships between the constructs in physics 1 and physics 2 and to unpack whether the constructs contributed toward explaining the gender difference in physics identity. We first tested gender moderation between different constructs using multi-group SEM (between male and female students) to investigate whether the relationships between the variables were different across gender both for the measurement part (CFA) and structural part (path analysis) of the SEM. There were no group differences at the level of weak and strong measurement invariance as well as at the level of regression coefficients. Therefore, we proceeded to gender mediation analysis, where gender is a precursor, to understand how the relationship between gender and physics identity is mediated through self-efficacy, interest, and perceived recognition at the end of the yearlong introductory physics sequence.

The results of the SEM for physics 1 are presented visually in Fig. 2. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): $\mathrm{CFI} = 0.977$ (>0.90), $\mathrm{TLI} = 0.969$ (>0.90), $\mathrm{RMSEA} = 0.048$ (<0.08), and $\mathrm{SRMR} = 0.033$ (<0.08). All three of the intervening variables (perceived recognition, self-efficacy, and interest) predict physics identity at the end of the physics course similar to past models [34,77]. Perceived recognition has the largest direct effect with smaller effects from self-efficacy and interest. In addition to the direct effect, perceived recognition also indirectly predicts physics identity through self-efficacy and interest and self-efficacy indirectly predicts physics identity through interest.

Additionally, gender is directly connected to perceived recognition (P.R.), self-efficacy, and interest. The relation between gender and physics identity was considered initially, however, the pathway was nonsignificant statistically (and therefore not shown for clarity). The relation between gender and physics identity is mediated only by the mediating constructs, and after accounting for these indirect paths, there is no direct path from gender to identity. In other words, women appear to have a lower physics identity because they have lower perceived recognition, self-efficacy, and interest.

In order to understand how these relationships change over the introductory physics course sequence, the results of the SEM for physics 2 are presented visually in Fig. 3. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): $\mathrm{CFI} = 0.972$ (>0.90), $\mathrm{TLI} = 0.963$ (>0.90), $\mathrm{RMSEA} = 0.055$ (<0.08), and $\mathrm{SRMR} = 0.038$ (<0.08). Similar to the model in physics 1, all three of the intervening variables (perceived recognition, self-efficacy, and interest) predict physics identity at the end of the yearlong physics course sequence. The students' perceived recognition is the strongest predictor of physics identity ($\beta = 0.73$). In addition to the direct effect, perceived recognition also indirectly predicts physics identity through self-efficacy and interest and self-efficacy indirectly predicts physics identity through interest.

Gender is directly connected to perceived recognition, self-efficacy, and interest similar to physics 1. Similar to the results shown in physics 1, the relation between gender and physics identity was considered initially, and the pathway was nonsignificant. Therefore, the relation between gender and physics identity is mediated only by the mediating constructs, and after accounting for these indirect paths, there is no direct path from gender to identity. In other words, similar to physics 1, women appear to have a lower physics identity because they have lower perceived recognition, self-efficacy, and interest.

To answer RQ3, we compared the path analysis results of the SEM model in physics 1 (Fig. 2) and physics 2 (Fig. 3). Most of the pathways in each model did not change or changed by a small amount. The exception was that the effect of physics self-efficacy on identity becomes larger in physics 2 while the effect of physics interest on identity becomes smaller and nonsignificant in physics 2. We also ran a multigroup SEM analysis using the bootstrap technique for physics 1 and 2 and compared the regression paths in physics 1 and physics 2 using confidence intervals. The confidence intervals for each regression pathway are shown in Table V. A prior study [96] suggests that the difference is statistically significant ($p \leq 0.05$) when the overlap of the 95% confidence intervals is no more than about half the average margin of error, that is when

TABLE V. Confidence intervals for each regression pathway in physics 1 and physics 2. C.I. is the confidence interval, P.R. is perceived recognition, Lower is lower bound and Upper is the upper bound. The proportion of overlap is above 0.5 for each of the regression pathway confidence intervals, and thus the regression pathways are not statistically significantly different from one another.

	Physics 1		Physics 2	
Regression path	C.I. Lower	C.I. Upper	C.I. Lower	C.I. Upper
P.R. → Self-efficacy	0.27	0.40	0.27	0.39
P.R. → Interest	0.26	0.46	0.25	0.45
Self-efficacy → Interest	0.42	0.79	0.45	0.83
P.R. → Identity	0.79	0.98	0.76	0.93
Self-efficacy → Identity	−0.03	0.28	0.06	0.37
Interest → Identity	0.01	0.21	−0.03	0.16

proportion overlap is about 0.5 or less. We use this rule to compare the regression paths in physics 1 and physics 2. For example, when comparing the confidence intervals for the regression pathway from perceived recognition (P.R.) to identity, the confidence interval is (0.79, 0.98) for physics 1 and (0.76, 0.93) for physics 2. The midpoint of the first confidence interval is 0.89, which is lower than the upper bound of the second confidence interval. Thus, the proportion of overlap between the confidence intervals is more than 0.5 and the regression pathways between physics 1 and 2 are not statistically significantly different from one another. None of the differences in the confidence intervals of the regression pathways (shown in Table V) are statistically significantly different from one another and thus none of these regression pathways for physics 1 and physics 2 are statistically significantly different from one another.

V. SUMMARY AND DISCUSSION

In this study, we investigated women's and men's physics beliefs longitudinally in a two-semester introductory physics course sequence for students on the bioscience track. While other studies have investigated physics identity and other student beliefs in calculus-based physics courses in which women are underrepresented [77,97], physics identity has not been investigated in physics courses in which women are not underrepresented. This new context is important since students' physics identity is context dependent and the factors that predict physics identity in a calculus-based physics course may not be the same in a course for students on the bioscience track. For example, in one recent study on female undergraduate students' physics identity, sense of belonging predicted physics identity for upper level students but not for introductory students [41]. Therefore, researchers have called for this type of research to be conducted in multiple contexts [41].

We find gender gaps in the beliefs disadvantaging women in both semesters of the physics course sequence. This finding is similar to prior research in calculus-based introductory physics courses in which women are underrepresented [77]. However, men's and women's beliefs do not significantly decrease over the course of the physics course sequence for students on the bioscience track, unlike in calculus-based physics courses (in which women's beliefs decrease more than men's) [23]. One potential reason is that women are not underrepresented in the physics courses for students on the bioscience track. However, gender inequity is maintained in these traditional lecture-based courses in which no explicit equity-related efforts are made, signifying that the learning environment is not equitable and inclusive. In addition, controlling for students' pre belief, gender differences were found in students' self-efficacy, interest, and perceived recognition at the end of physics 1 and in students' identity and interest at the end of physics 2.

One hypothesis for the gender gap in these constructs (indicated by women's lower scores) is that women may be affected by previous experiences, stereotypes, and biases about who belongs in physics and who can excel in it, which can accumulate over their lifetime and in the absence of an equitable and inclusive learning environment, the gender gaps are maintained. For example, when validating this survey with individual interviews, women often noted that when they asked questions to their physics instructor or TA, they responded by saying that the questions were trivial, easy, or obvious. This made them feel stupid and feel like their questions were devalued, often in front of their peers. Some of them also acknowledged that they never asked another question after that. These types of interactions show the important role that instructors or TAs play in ensuring that women do not feel disparaged and positively recognizing them and helping them feel like they belong in the physics classes. While the instructor or TA may not have meant to belittle women and could have responded with this type of response to anyone who asked those questions, the societal stereotypes, and biases about who belongs in physics and can excel in it, particularly impact women's interpretation of these types of negative interactions. What instructors and TAs must internalize is that what is important is not their intentions but the impact they have on the students. We hypothesize that these gender gaps that persist through the two-semester introductory physics course sequence, even though women are not underrepresented in these courses, may signify the impact of deep-rooted societal stereotypes and biases pertaining to physics as well as a noninclusive and inequitable culture of these physics classes that do not focus on uprooting these inequities.

Moreover, we find that our identity model using SEM for the two-semester introductory physics sequence suggests that perceived recognition, self-efficacy, and interest predict students' physics identity in a qualitatively similar manner in physics 1 and physics 2. This model is similar to models in past studies [34] and calculus-based physics courses in which women are underrepresented [41,77,97]. We did not find any statistically significant difference between the SEM models in physics 1 and physics 2. We note that the physics self-efficacy to identity pathway becomes significant in physics 2 while the pathway of physics interest on identity becomes smaller and nonsignificant in physics 2. However, since these differences were not significant, this indicates that the models do not significantly change from physics 1 to physics 2. Therefore, it is important that instructors provide opportunities for students to improve their self-efficacy and interest in both the physics courses. Moreover, the path from gender to each motivational belief is very concerning since it shows that women have lower motivational beliefs at the end of physics 1 and physics 2. If these courses were equitable and inclusive, there would not be a path from gender to any of the motivational beliefs at the end of physics 1 and physics 2. Additionally, our findings suggest that perceived recognition as a physics person by others is the strongest predictor of physics identity and gender differences in perceived recognition disadvantage women's physics identity in both physics 1 and physics 2.

The strong contribution of perceived recognition in predicting physics identity throughout the two-semester course sequence shows that TAs and instructors can play a critical role to increase students' physics self-efficacy, interest, and identity. We note that the physics 1 and physics 2 courses in this study were traditionally taught lecture-based physics courses in which student grades heavily depended on two or three midterm exams and a final exam. The courses consisted of 3 h of lecture per week taught by the instructor and 1 h of recitation per week taught by a TA. It is important to recognize that even in these traditional lecture-based courses, students often receive feedback from their instructors in multiple ways, including receiving praise for asking or answering a question in class (which often advantages male students since they dominate these situations) and their interactions with students during office hours or over email. In addition, students interact with their TAs in recitation by asking questions about the homework or class material at the start of recitation, when completing group work during recitation, and during the TA's office hours. However, the gender differences in these beliefs throughout the two-semester course sequence suggest that physics instructors and TAs are not doing enough to create an equitable and inclusive learning environment. Our view of an equitable and inclusive learning environment is that it should provide adequate support to all students and close the initial gaps, e.g., in the beliefs of students from different demographic groups such as female and male students discussed here. The focus on inclusivity of the learning environment is important since one study showed that the student perception of the inclusivity of the physics learning environment, consisting of students' perceived recognition, sense of belonging, and perception of the effectiveness of interaction with their peers predicted students' physics beliefs at the end of the course [97]. Thus, instructors and TAs should strive to make the learning environment in both physics 1 and physics 2 classes equitable and inclusive.

One study showed that incorporating discussion sessions about the underrepresentation of women in physics improved women's physics identities [98]. In addition, explicitly holding everyone to the same standards, not letting men dominate the conversations in class or office hours, and explicitly praising women when they do well or make progress on various components of the class could help decrease the gender gap in students' perceived recognition while also boosting their self-efficacy and interest. One strategy to increase students' recognition by instructors is to offer opportunities for student-centered learning where students can serve as leaders in problem solving and encouraging students to persist in their efforts by normalizing struggle and framing struggle as an important stepping stone to learning and developing a solid grasp of physics [76].

It is especially important for instructors to focus on equity and inclusion when implementing active-learning pedagogy. For example, one study showed that teacher practices, including their encouragement of cooperative activities in an inclusive environment, were related to students' engagement in the course [99]. In particular, if active-engagement pedagogies are not implemented using teaching strategies that are equitable and inclusive, men have been shown to not only dominate responding to questions in class but also while working in groups which can lower women's self-efficacy [100]. Additionally, instructors can improve equity and inclusion in their courses by making their courses student-centered, e.g., by adopting pedagogy that focuses on societal implications of physics [62] in addition to providing mentoring or support for students who are underrepresented [101]. Short social-psychological classroom interventions, e.g., sense of belonging and mindset interventions, have also been shown to positively impact students from underrepresented groups including women [102–105]. Beyond the physics classroom, informal science activities, like participation in science fairs or talking science with family and friends could increase students' physics identity in college [106].

In summary, even in physics courses in which women are not underrepresented, women have lower physics beliefs, including physics identity than men. While having a larger cohort of women helps these constructs stay relatively similar over the two-semester course sequence compared to calculus-based courses [23], negative stereotypes about who belongs in physics and can excel in it disadvantage women throughout the yearlong physics sequence, hinting at lack of effort to create an equitable and inclusive learning environment in both these physics classes. Instructors and TAs must make concerted efforts to not let men dominate in class and create a learning environment that emphasizes recognizing and validating all their students in these physics courses, particularly women and other underrepresented students who have been stigmatized due to societal stereotypes and biases about physics for too long.

VI. LIMITATIONS

One limitation of our study is that we did not measure whether the students in these courses internalized the societal stereotypes and biases about who belongs in physics and can excel in it based upon their gender. However, since these stereotypes and biases are in multiple facets at every stage of life, at least some of the women in the introductory physics classes may be impacted by the stereotypes. Future studies would investigate how female and male students endorsing gender-based stereotypes pertaining to physics impacts their beliefs in these courses. Additionally, we note that the physics 1 and physics 2 courses in this study were traditionally taught lecture-based physics courses in which there were 3 hours of lecture per week taught by the instructor, 1 hour of recitation per week taught by a TA and student grades heavily depended on midterm exams and a final exam. In these lecture style courses, students may receive positive or negative recognition from their instructor in different ways, e.g., the instructor may praise students who answer or ask questions during lectures. In addition, the students can ask the TA questions during recitation about homework or group work they need to complete. The male students often dominate in these situations. However, there were no research-based active engagement strategies used in the classroom. Therefore, it would be beneficial to investigate these student beliefs in courses in which active engagement strategies are used. In particular, it would be useful to investigate how these findings are impacted by research-based active engagement courses in which there is no explicit focus on equity and inclusion and also those in which equity and inclusion are at the center. Also, physics identity is context dependent and thus may not be generalizable across institutions of different types. Thus, studies that investigate physics identity and other beliefs related to it such as the one presented here for different student populations should be conducted, e.g., at different types of institutions including four-year colleges, community colleges, and minority serving institutions.

ACKNOWLEDGMENTS

This work was supported by Grant No. NSF DUE-152457. We thank all students who participated in this research and Dr. Robert Devaty for his constructive feedback on this manuscript.

APPENDIX A: STUDENTS' PRE- AND POSTBELIEFS

The descriptive statistics of the students' pre- and postbeliefs in each course are shown in Tables VI and VII. Since the perceived recognition and identity constructs were not included in our survey at the beginning of physics 1 in the first year of study, we only include data for matched students (students who took both the pre and post survey) in the second year. However, the students are not matched from physics 1 to physics 2. Therefore, the sample size is smaller than that in the main text; however, the results from one year shown below are qualitatively similar to the results for two years (discussed in the main text).

TABLE VI. Mean physics 1 pre- and postpredictor and outcome values by gender as well as statistical significance (p values) and effect sizes (Cohen's d) by gender for 260 women and 126 men. All p values <0.001.

	Premean				Postmean			
Predictors and outcomes	Male	Female	Cohen's d	p value	Male	Female	Cohen's d	p value
Perceived recognition	2.11	1.98	0.20	0.065	2.16	1.95	0.35	0.001
Self-efficacy	3.04	2.85	0.48	<0.001	2.90	2.70	0.38	<0.001
Interest	2.88	2.50	0.68	<0.001	2.80	2.35	0.73	<0.001
Physics identity	2.26	1.94	0.44	<0.001	2.08	1.81	0.37	0.001

TABLE VII. Mean physics 2 pre and postpredictor and outcome values by gender as well as statistical significance (*p* values) and effect sizes (Cohen's d) by gender for 274 women and 148 men. All *p* values <0.001.

	Premean				Postmean			
Predictors and outcomes	Male	Female	Cohen's *d*	*p* value	Male	Female	Cohen's *d*	*p* value
Perceived recognition	2.21	2.03	0.28	0.006	2.26	2.05	0.31	0.003
Self-efficacy	3.02	2.81	0.53	<0.001	2.94	2.74	0.38	<0.001
Interest	2.80	2.31	0.91	<0.001	2.79	2.26	0.87	<0.001
Physics identity	2.16	1.91	0.37	<0.001	2.19	1.88	0.43	<0.001

APPENDIX B: PERCENTAGES OF MALE AND FEMALE STUDENTS WHO SELECTED EACH CHOICE FOR EACH SURVEY ITEM

Below, we provide the percentages of men and women who selected each answer choice for each question in the pre- and postsurvey in physics 1 and physics 2. This distribution provides a sense of how students shifted their answers from pre to post survey. Tables VIII and IX are for women and men, respectively, in physics 1 while Tables X and XI are for women and men, respectively, in physics 2.

TABLE VIII. Percentages of 260 women in physics 1 who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO!, no, yes, YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

	Women							
	Presurvey				Postsurvey			
Question	1	2	3	4	1	2	3	4
Physics identity								
1	27%	59%	13%	1%	45%	45%	9%	1%
Physics perceived recognition								
2	31%	53%	15%	1%	40%	46%	13%	1%
3	30%	55%	13%	2%	38%	46%	13%	3%
4	16%	52%	29%	3%	26%	38%	32%	4%
Physics self-efficacy								
5	14%	45%	39%	2%	11%	29%	54%	6%
6	4%	19%	68%	9%	6%	29%	59%	6%
7	1%	6%	66%	27%	11%	39%	41%	9%
8	1%	14%	66%	19%	7%	39%	48%	6%
Physics interest								
9	24%	53%	20%	3%	23%	35%	33%	9%
10	5%	32%	58%	5%	13%	38%	46%	3%
11	5%	44%	45%	6%	18%	54%	26%	2%
12	5%	33%	54%	8%	15%	40%	41%	4%

TABLE IX. Percentages of 126 men in physics 1 who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO!, no, yes, YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

	Men							
	Presurvey				Postsurvey			
Question	1	2	3	4	1	2	3	4
Physics identity								
1	9%	59%	28%	4%	19%	58%	19%	4%
Physics perceived recognition								
2	20%	60%	17%	3%	17%	59%	23%	1%
3	21%	59%	18%	2%	18%	62%	18%	2%
4	12%	53%	30%	5%	8%	54%	32%	6%
Physics self-efficacy								
5	5%	33%	59%	3%	4%	26%	60%	10%
6	1%	10%	77%	12%	5%	14%	66%	15%
7	0%	3%	56%	41%	2%	13%	60%	25%
8	0%	5%	69%	26%	2%	22%	64%	12%
Physics interest								
9	8%	37%	36%	19%	3%	26%	42%	29%
10	1%	16%	66%	17%	5%	20%	64%	11%
11	0%	19%	67%	14%	7%	30%	54%	9%
12	3%	20%	56%	21%	4%	31%	52%	13%

TABLE X. Percentages of 274 women in physics 2 who answered each question in physics 2 by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO!, no, yes, YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

	Women							
	Presurvey				Postsurvey			
Question	1	2	3	4	1	2	3	4
Physics identity								
1	28%	54%	17%	1%	30%	53%	15%	2%
Physics perceived recognition								
2	26%	55%	18%	1%	26%	53%	19%	2%
3	25%	55%	18%	2%	26%	49%	23%	2%
4	17%	49%	31%	3%	19%	47%	30%	4%
Physics self-efficacy								
5	5%	30%	64%	1%	7%	30%	57%	6%
6	2%	19%	75%	4%	5%	27%	64%	4%
7	2%	15%	70%	13%	4%	20%	62%	14%
8	1%	18%	73%	8%	2%	26%	61%	11%
Physics interest								
9	22%	48%	25%	5%	23%	45%	26%	6%
10	10%	40%	48%	2%	14%	40%	42%	4%
11	9%	59%	30%	2%	15%	59%	24%	2%
12	11%	37%	48%	4%	12%	42%	40%	6%

TABLE XI. Percentages of 148 men in physics 2 who answered each question in physics 2 by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO!, no, yes, YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

	Men							
	Presurvey				Postsurvey			
Question	1	2	3	4	1	2	3	4
Physics identity								
1	13%	61%	23%	3%	16%	53%	26%	5%
Physics perceived recognition								
2	16%	57%	23%	4%	18%	53%	25%	4%
3	16%	57%	23%	4%	18%	48%	26%	8%
4	10%	52%	35%	3%	12%	43%	40%	5%
Physics self-efficacy								
5	2%	21%	68%	9%	7%	20%	63%	10%
6	1%	7%	82%	10%	3%	12%	72%	13%
7	1%	6%	66%	27%	1%	11%	65%	23%
8	1%	11%	71%	17%	3%	16%	67%	14%
Physics interest								
9	7%	30%	48%	15%	6%	26%	42%	26%
10	3%	19%	61%	17%	3%	23%	56%	18%
11	2%	30%	59%	9%	5%	41%	45%	9%
12	4%	22%	62%	12%	3%	30%	53%	14%

---

[1] T. J. Nokes-Malach, Z. Y. Kalender, E. Marshman, C. D. Schunn, and C. Singh, Prior preparation and motivational characteristics mediate relations between gender and learning outcomes in introductory physics, in *Proceedings of PER Conf. 2018, Washington, DC* (2018), 10.1119/perc.2018.pr.Nokes-Malach.

[2] E. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?, Phys. Rev. Phys. Educ. Res. **14**, 020123 (2018).

[3] J. L. Smith, C. Sansone, and P. H. White, The stereotyped task engagement process: The role of interest and achievement motivation, J. Educ. Psychol **99**, 99 (2007).

[4] J. S. Eccles, Understanding women's educational, and occupational choices: Applying the Eccles, *et al.* model of achievement-related choices, Psychol. Women Q. **18**, 585 (1994).

[5] S. J. Correll, Gender and the career choice process: The role of biased self-assessments, Am. J. Soc. **106**, 1691 (2001).

[6] N. M. Hewitt and E. Seymour, A long, discouraging climb, ASEE Prism **1**, 24 (1992).

[7] R. Ivie and K. Stowe, Women in Physics, 2000, AIP Report (2000).

[8] G. C. Marchand and G. Taasoobshirazi, Stereotype threat and women's performance in physics, Int. J. Sci. Educ. **35**, 3050 (2013).

[9] S. L. Beilock, R. J. Rydell, and A. R. McConnell, Stereotype threat and working memory: Mechanisms, alleviation, and spillover, J. Exp. Psychol. Gen. **136**, 256 (2007).

[10] K. R. Christy and J. Fox, Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance, Comput. Educ. **78**, 66 (2014).

[11] R. J. Rydell, R. M. Shiffrin, K. L. Boucher, K. Van Loo, and M. T. Rydell, Stereotype threat prevents perceptual

learning, Proc. Natl. Acad. Sci. U.S.A. **107**, 14042 (2010).

[12] S. Tobias, *They're Not Dumb, They're Different* (Research Corporation Tucson, Arizona, 1990).

[13] A. Maries, N. I. Karim, and C. Singh, Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics?, Phys. Rev. Phys. Educ. Res. **14**, 020119 (2018).

[14] M.-T. Wang and J. Degol, Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields, Dev. Rev. **33**, 304 (2013).

[15] S. L. Britner, Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes, J. Res. Sci. Teach. **45**, 955 (2008).

[16] S. J. Correll, Constraints into preferences: Gender, status, and emerging career aspirations, Am. Soc. Rev. **69**, 93 (2004).

[17] C. P. Benbow and L. L. Minor, Mathematically talented males and females and achievement in the high school sciences, Am. Educ. Res. J. **23**, 425 (1986).

[18] J. E. Stets, P. S. Brenner, P. J. Burke, and R. T. Serpe, The science identity and entering a science occupation, Soc. Sci. Res. **64**, 1 (2017).

[19] K. A. Robinson, T. Perez, J. H. Carmel, and L. Linnenbrink-Garcia, Science identity development trajectories in a gateway college chemistry course: Predictors and relations to achievement and STEM pursuit, Contemp. Educ. Psychol. **56**, 180 (2019).

[20] M. M. Chemers, E. L. Zurbriggen, M. Syed, B. K. Goza, and S. Bearman, The role of efficacy and identity in science career commitment among underrepresented minority students, J. Soc. Issues **67**, 469 (2011).

[21] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Damage caused by women's lower self-efficacy on physics learning, Phys. Rev. Phys. Educ. Res. **16**, 010118 (2020).

[22] J. Stewart, R. Henderson, L. Michaluk, J. Deshler, E. Fuller, and K. Rambo-Hernandez, Using the Social Cognitive Theory Framework to Chart Gender Differences in the Developmental Trajectory of STEM Self-Efficacy in Science and Engineering Students, J. Sci. Educ. Technol. **29**, 758 (2020).

[23] E. Marshman, Z. Y. Kalender, C. Schunn, T. Nokes-Malach, and C. Singh, A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences, Can. J. Phys. **96**, 391 (2018).

[24] R. A. Louis and J. M. Mistele, The differences in scores and self-efficacy by student gender in mathematics and science, Int. J. Sci. Math. Educ. **10**, 1163 (2012).

[25] J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 020105 (2016).

[26] D. Doucette and C. Singh, Why are there so few women in physics? Reflections on the experiences of two women, Phys. Teach. **58**, 297 (2020).

[27] S. Cwik and C. Singh, Students' sense of belonging in introductory physics course for bioscience majors predicts their grade, Phys. Rev. Phys. Educ. Res. **18**, 010139 (2022).

[28] S. L. Eddy, S. E. Brownell, and M. P. Wenderoth, Gender gaps in achievement and participation in multiple introductory biology classrooms, CBE Life Sci. Educ. **13**, 361 (2014).

[29] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).

[30] H. B. Carlone and A. Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, J. Res. Sci. Teach. **44**, 1187 (2007).

[31] J. P. Gee, Chapter 3: Identity as an analytic lens for research in education, Rev. Res. Educ. **25**, 99 (2000).

[32] K. L. Tonso, Student engineers and engineer identity: Campus engineer identities as figured world, Cult. Stud. Sci. Educ. **1**, 273 (2006).

[33] P. Vincent-Ruz and C. D. Schunn, The nature of science identity and its role as the driver of student choices, Int. J. STEM Educ. **5**, 48 (2018).

[34] 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).

[35] 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).

[36] 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).

[37] C. Monsalve, Z. Hazari, D. McPadden, G. Sonnert, and P. M. Sadler, Examining the relationship between career outcome expectations and physics identity, in *Proceedings of PER Conf. 2016, Sacramento, CA* (2016), 10.1119/perc.2016.pr.052.

[38] L. Archer, J. Moote, B. Francis, J. DeWitt, and L. Yeomans, The "exceptional" physics girl: A sociological analysis of multimethod data from young women aged 10–16 to explore gendered patterns of post-16 participation, Am. Educ. Res. J. **54**, 88 (2017).

[39] R. M. Lock, Z. Hazari, and G. Potvin, Physics career intentions: The effect of physics identity, math identity, and gender, AIP Conf. Proc. **1513**, 262 (2013).

[40] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice, J. Eng. Educ. **105**, 312 (2016).

[41] 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).

[42] A. Bandura, in *Encyclopedia of Psychology*, edited by R. J. Corsini (Wiley, New York, 1994), pp. 368–9.

[43] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. **84**, 191 (1977).

[44] V. Sawtelle, E. Brewe, and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, J. Res. Sci. Teach. **49**, 1096 (2012).

[45] C. Lindstrøm and M. D. Sharma, Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter?, Int. J. Innov. Sci. Math Educ. **19**, 1 (2011).

[46] R. M. Felder, G. N. Felder, M. Mauney, and C. E. Hamrin Jr, and E. J. Dietz, A longitudinal study of engineering student performance and retention. III. Gender differences in student performance and attitudes, J. Eng. Educ. **84**, 151 (1995).

[47] A. M. L. Cavallo, W. H. Potter, and M. Rozman, Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors, School Sci. Math. **104**, 288 (2004).

[48] B. J. Zimmerman, Self-efficacy: An essential motive to learn, Contemp. Educ. Psychol. **25**, 82 (2000).

[49] H. Fencl and K. Scheel, Research and Teaching: Engaging students—An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course, J. Coll. Sci. Teach. **35**, 20 (2005).

[50] D. H. Schunk and F. Pajares, *Development of Achievement Motivation* (Elsevier, London, 2002), pp. 15–31.

[51] T. Bouffard-Bouchard, S. Parent, and S. Larivee, Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students, Int. J. Behav. Dev. **14**, 153 (1991).

[52] S. L. Britner and F. Pajares, Sources of science self-efficacy beliefs of middle school students, J. Res. Sci. Teach. **43**, 485 (2006).

[53] J. M. Bailey, D. Lombardi, J. R. Cordova, and G. M. Sinatra, Meeting students halfway: Increasing self-efficacy and promoting knowledge change in astronomy, Phys. Rev. Phys. Educ. Res. **13**, 020140 (2017).

[54] P. Vincent-Ruz and C. D. Schunn, The increasingly important role of science competency beliefs for science learning in girls, J. Res. Sci. Teach. **54**, 790 (2017).

[55] S. Cheryan, S. A. Ziegler, A. K. Montoya, and L. Jiang, Why are some STEM fields more gender balanced than others?, Psychol. Bull. **143**, 1 (2017).

[56] A. Malespina and C. Singh, Gender differences in test anxiety and self-efficacy: Why instructors should emphasize low-stakes formative assessments in physics courses, Eur. J. Phys. **43**, 035701 (2022).

[57] S. Hidi and K. A. Renninger, The four-phase model of interest development, Educ. Psychol. **41**, 111 (2006).

[58] E. Lichtenberger and C. George-Jackson, Predicting high school students' interest in majoring in a STEM field: Insight into high school students' postsecondary plans, J. Career Tech. Educ. **28**, 19 (2013).

[59] A. C. Strenta, R. Elliott, R. Adair, M. Matier, and J. Scott, Choosing and leaving science in highly selective institutions, Res. High. Educ. **35**, 513 (1994).

[60] J. M. Harackiewicz, K. E. Barron, J. M. Tauer, and A. J. Elliot, Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation, J. Educ. Psychol **94**, 562 (2002).

[61] S. Hidi, Interest: A unique motivational variable, Educ. Res. Rev. **1**, 69 (2006).

[62] P. Häussler and L. Hoffmann, An intervention study to enhance girls' interest, self-concept, and achievement in physics classes, J. Res. Sci. Teach. **39**, 870 (2002).

[63] A. Wigfield and J. S. Eccles, The development of achievement task values: A theoretical analysis, Dev. Rev. **12**, 265 (1992).

[64] C. Goodenow, Classroom belonging among early adolescent students: Relationships to motivation and achievement, J. Early Adolesc. **13**, 21 (1993).

[65] S. Cwik and C. Singh, Not feeling recognized as a physics person by instructors and teaching assistants is correlated with female students' lower grades, Phys. Rev. Phys. Educ. Res. **18**, 010138 (2022).

[66] D. Doucette, R. Clark, and C. Singh, Hermione and the Secretary: How gendered task division in introductory physics labs can disrupt equitable learning, Eur. J. Phys. **41**, 035702 (2020).

[67] Y. Li, K. Whitcomb, and C. Singh, How perception of being recognized or not recognized by instructors as a "physics person" impacts male and female students' self-efficacy and performance, Phys. Teach. **58**, 484 (2020).

[68] D. Z. Grunspan, S. L. Eddy, S. E. Brownell, B. L. Wiggins, A. J. Crowe, and S. M. Goodreau, Males underestimate academic performance of their female peers in undergraduate biology classrooms, PLoS One **11**, e0148405 (2016).

[69] S. Knobloch-Westerwick, C. J. Glynn, and M. Huge, The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest, Sci. Com. **35**, 603 (2013).

[70] A. E. Lincoln, S. Pincus, J. B. Koster, and P. S. Leboy, The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s, Social Studies Sci. **42**, 307 (2012).

[71] Z. Hazari, C. Cass, and C. Beattie, Obscuring power structures in the physics classroom: Linking teacher positioning, student engagement, and physics identity development, J. Res. Sci. Teach. **52**, 735 (2015).

[72] C. M. Mueller and C. S. Dweck, Praise for intelligence can undermine children's motivation and performance, J. Pers. Soc. Psychol. **75**, 33 (1998).

[73] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Gendered patterns in the construction of physics identity from motivational factors, Phys. Rev. Phys. Educ. Res. **15**, 020119 (2019).

[74] L. Bian, S.-J. Leslie, and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, Science **355**, 389 (2017).

[75] S. Upson and L. F. Friedman, Where are all the female geniuses?, Sci. Am. Mind **23**, 63 (2012).

[76] Z. Hazari and C. Cass, Towards meaningful physics recognition: What does this recognition actually look like?, Phys. Teach. **56**, 442 (2018).

[77] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Why female science, technology, engineering, and mathematics majors do not identify with

physics: They do not think others see them that way, Phys. Rev. Phys. Educ. Res. **15**, 020148 (2019).

[78] G. Potvin and Z. Hazari, The development and measurement of identity across the physical sciences, *Proceedings of PER Conf. 2013, Portland, OR*, 10.1119/perc.2013.pr.058.

[79] A. M. Flowers III and R. Banda, Cultivating science identity through sources of self-efficacy, J. Multicultural Educ. **10**, 405 (2016).

[80] 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).

[81] A. S. Bryk and S. W. Raudenbush, *Hierarchical linear models: Applications and data analysis methods* (Sage Publications, Inc., Newbury Park, CA, 1992).

[82] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey, Phys. Rev. Phys. Educ. Res. **2**, 010101 (2006).

[83] S. M. Glynn, P. Brickman, N. Armstrong, and G. Taasoobshirazi, Science motivation questionnaire II: Validation with science majors and nonscience majors, J. Res. Sci. Teach. **48**, 1159 (2011).

[84] PERTS Academic Mindsets Assessment, https://www.perts.net/orientation/ascend.

[85] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge, London, 2013).

[86] L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika **16**, 297 (1951).

[87] Activation lab tools: Measures and data collection instruments, http://www.activationlab.org/tools/.

[88] J. Schell and B. Lukoff, Peer Instruction self-efficacy instrument [Developed at Harvard University] (unpublished).

[89] M.-C. Shanahan and M. Nieswandt, Creative activities and their influence on identification in science: Three case studies, J. Elem. Sci. Educ. **21**, 63 (2009).

[90] A. Godwin, P. Geoff, and H. Zahra, The development of critical engineering agency, identity, and the impact on engineering career choices, in *Proceedings of the 2013 ASEE Annual Conference & Exposition* (2021)..

[91] Z. Hazari, P. M. Sadler, and G. Sonnert, The science identity of college students: Exploring the intersection of gender, race, and ethnicity, J. Coll. Sci. Teach. **42**, 82 (2013).

[92] R. Likert, A technique for the measurement of attitudes, Arch. Psychol. **22**, 55 (1932).

[93] R. C. Team, *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, 2013).

[94] R. C. MacCallum, M. W. Browne, and H. M. Sugawara, Power analysis and determination of sample size for covariance structure modeling, Psychological Methods **1**, 130 (1996).

[95] A. J. Tomarken and N. G. Waller, Structural equation modeling: Strengths, limitations, and misconceptions, Annu. Rev. Clin. Psychol. **1**, 31 (2005).

[96] G. Cumming and S. Finch, Inference by eye: Confidence intervals and how to read pictures of data, Am. Psychol. **60**, 170 (2005).

[97] Y. Li and C. Singh, Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses, Phys. Rev. Phys. Educ. Res. **17**, 010143 (2021).

[98] R. Lock and Z. Hazari, Discussing underrepresentation as a means to facilitating female students' physics identity development, Phys. Rev. Phys. Educ. Res. **12**, 020101 (2016).

[99] D. Solomon, V. Battistich, D.-i. Kim, and M. Watson, Teacher practices associated with students' sense of the classroom as a community, Social Psychol. Educ. **1**, 235 (1996).

[100] S. M. Aguillon, G.-F. Siegmund, R. H. Petipas, A. G. Drake, S. Cotner, and C. J. Ballen, Gender differences in student participation in an active-learning classroom, CBE Life Sci. Educ. **19**, ar12 (2020).

[101] A. Rockinson-Szapkiw, J. L. Wendt, and J. S. Stephen, The efficacy of a blended peer mentoring experience for racial and ethnic minority women in STEM pilot study: Academic, professional, and psychosocial outcomes for mentors and mentees, J. STEM Educ. Res. **4**, 173 (2021).

[102] D. S. Yeager and G. M. Walton, Social-psychological interventions in education: They're not magic, Rev. Educ. Res. **81**, 267 (2011).

[103] J. M. Harackiewicz, E. A. Canning, Y. Tibbetts, S. J. Priniski, and J. S. Hyde, Closing achievement gaps with a utility-value intervention: Disentangling race and social class, J. Pers. Soc. Psychol. **111**, 745 (2016).

[104] G. M. Walton, C. Logel, J. M. Peach, S. J. Spencer, and M. P. Zanna, Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering, J. Educ. Psychol **107**, 468 (2015).

[105] K. R. Binning, N. Kaufmann, E. M. McGreevy, O. Fotuhi, S. Chen, E. Marshman, Z. Yasemin Kalender, L. Limeri, L. Betancur, and C. Singh, Changing social norms to foster the benefits of collaboration in diverse workgroups, Psychol. Sci. **31**, 1059 (2020).

[106] Z. Hazari, R. Dou, G. Sonnert, and P. M. Sadler, Examining the relationship between informal science experiences and physics identity: Unrealized possibilities, Phys. Rev. Phys. Educ. Res. **18**, 010107 (2022).