Do female and male students' physics motivational beliefs change in a two-semester introductory physics course sequence?

Yangqiuting Li and Chandralekha Singh
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

(Received 7 July 2021; revised 7 February 2022; accepted 2 May 2022; published 3 June 2022)

We investigated students' physics motivational beliefs including their physics self-efficacy, interest, perceived recognition, and identity in a traditionally taught two-semester college calculus-based introductory physics sequence (referred to as physics 1 and physics 2). We studied whether and how these motivational beliefs evolve in this course sequence in terms of the average scores and the predictive relationships among them. The results show that both female and male students' physics self-efficacy and interest decreased from physics 1 to physics 2, while there was no statistically significant change in students' perceived recognition and identity. We found signatures of an inequitable and noninclusive learning environment in that not only was there a gender difference in students' motivational beliefs disadvantaging women, but the gender difference in perceived recognition increased from physics 1 to physics 2. We used structural equation modeling (SEM) to investigate the predictive relationships among students' motivational beliefs in physics 1 and physics 2. Analysis revealed that perceived recognition from others, e.g., instructors and teaching assistants, was the largest predictor of physics identity in both courses, and the role played by perceived recognition was even more important in physics 2 for predicting identity and mediating the gender difference in self-efficacy. Our findings suggest that perceived recognition is very important for the development of students' physics identity in both physics 1 and 2. However, female students feel less recognized in the current learning environment and this gender difference grows from physics 1 to physics 2. Instructors should be trained to create an equitable and inclusive learning environment, in which all students feel recognized and supported appropriately and develop a stronger physics self-efficacy, interest, and identity.

DOI: 10.1103/PhysRevPhysEducRes.18.010142

I. INTRODUCTION AND THEORETICAL FRAMEWORK

In the disciplines of science, technology, engineering, and mathematics (STEM), there have been efforts to enhance the participation and advancement of underrepresented groups such as women [1–11]. Prior research suggests that individuals' course enrollment, degree attainment, and achievement in STEM can be influenced by their motivational beliefs such as self-efficacy, interest, and identity in that domain [2,6,11–15]. For students from underrepresented groups, these motivational beliefs might be undermined, e.g., by negative societal stereotypes and biases about who belongs and can excel in STEM as well as lack of role models and encouragement from others, which can lead to withdrawal from STEM courses, majors, or careers [16–21]. Hence, investigating students' motivational beliefs is critical to understanding and addressing diversity, equity, and inclusion issues in STEM disciplines.

For explaining participation in STEM careers, identity has been argued to be a particularly important motivational construct [1,2,22,23]. Students' identity in an academic domain, such as physics, is students' views about whether they see themselves as a "physics person" [1,22]. Prior studies have shown that students' identity in STEM can be influenced by other motivational beliefs. For example, the well-known science identity framework by Carlone and Johnson includes three dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others") [1]. Hazari *et al.* adapted this framework to physics and added a new factor "interest." Also, they focused on students' beliefs about their competence and performance rather than how students can practice and exhibit them in class [24,25]. Moreover, they found that beliefs about performance and competence are actually not distinct and predict students' physics identity as a single construct [22,26]. Kalender *et al.* adapted Hazari *et al.*'s physics identity framework such that perceived recognition is a predictor of competency belief and interest while all of these predict identity similar to Hazari *et al.* [27]. Based on the studies discussed above, other studies have been conducted to examine the impact of physics identity on students' career intentions [28,29] as well as other possible factors that could affect physics identity such as out-of-class science and engineering activities [29] and students' sense of belonging [30,31]. In addition, the identity framework has also been adapted to studies of math and engineering identities [32–35].

Competency belief is defined as the extent to which a person feels they have the necessary attributes to succeed [36]. A related concept is self-efficacy, which refers to an individual's belief in their capacity to execute behaviors necessary to produce specific performance attainments [37–39]. In a particular academic domain, self-efficacy is defined as students' beliefs in their capability to succeed in specific situations or in accomplishing a task [40]. Since the definitions of competency belief and self-efficacy are very similar for the purposes of this research which uses validated survey data, and our survey items were adapted from prior studies that use the term self-efficacy [41,42], we continue using this term here. In this study, we will use the physics identity model in which physics identity is predicted by self-efficacy, interest, and perceived recognition to study progression of students' motivational beliefs in a two-semester college introductory physics course sequence.

Prior research suggests that self-efficacy is an important motivational belief of students in order for them to excel in a domain [5,6,11]. In particular, self-efficacy has been shown to influence students' engagement and performance in a given domain [12,14]. Students with high self-efficacy in a domain often enroll in more challenging courses in that domain than those with low self-efficacy because they perceive difficult tasks as challenges rather than threats [13]. In addition, studies show that students' self-efficacy predicts their career choices and persistence toward their short-term and long-term career goals [12].

Another component of identity is interest, which is defined by positive emotions accompanied by curiosity and engagement in a particular topic [43]. Interest has also been shown to influence students' learning [12,15,43]. For example, one study showed that making science courses more relevant to students' lives and transforming curricula to promote interest in learning can improve students' achievement [44]. In addition, studies have shown that students' interest is not completely independent of self-efficacy [12,45]. According to Eccles's expectancy-value theory (EVT) [45,46], interest is paired well with self-efficacy as connected constructs that predict students' academic outcome expectations and career aspirations.

The identity frameworks discussed earlier reveal that individuals' internal identity (how a person perceives themselves [47]) in a domain is also predicted by their perceived recognition from others (also called external identity). Perceived recognition in a domain, such as physics, refers to students' perception about whether other people see them as a physics person [25]. Some quantitative studies focusing on the relation between students' motivational beliefs and identity in a field show that perceived recognition is actually the strongest predictor of identity as compared to interest and self-efficacy [23,27]. For example, Godwin *et al.* found that students' physics and math recognition beliefs each have the largest direct effect on physics and math identity, respectively [23].

However, many studies have shown that female students did not feel that they were recognized appropriately in science even before they entered college [48,49]. For example, a report of the National Science Foundation [50] indicated that elementary and high school boys and girls interested in science felt that they were treated differently by parents, teachers, and friends with regard to their interest in science. While boys received admiration and encouragement for their interests, responses to girls were often characterized by ambivalence, lack of encouragement, or suggestions that their goals were inappropriate [50]. Studies show that these stereotypes and biases also exist in the university context [27,51]. For example, one study showed that science faculty members in biological and physical sciences exhibit biases against female students and rate male students as more competent even though only the names were different in the hypothetical information they were provided [51]. These experiences of being belittled or not being recognized as a science person not only have the potential to deteriorate female students' motivational beliefs and performance but may also dissuade them from pursuing study in science altogether. In addition, many disciplines in science such as physics have stereotypes about requiring a natural ability to excel in them, which may further impact female students' motivational beliefs because, due to societal stereotypes, being a genius or exceptionally smart is always associated with boys [48,52,53]. These stereotypes may cause female students to have a different perception of what it means to excel in these disciplines than male students, and they may even assume that they have to make extra efforts to succeed in these fields relative to male students. For example, one study shows that in introductory physics courses, female students had significantly lower physics self-efficacy than their equally performing male peers [54]. In addition, female students' interest in science can also be influenced by these negative stereotypes. For example, in a ten-year longitudinal study, researchers found that science-keen girls or young women who named physics as their favorite subject slowly lose their interest due to alienation, discrimination, and gender-biased beliefs about physics [48].

These studies show that students' motivational beliefs are not static; rather, they are dynamic and evolve in response to interactions with other people and the type of learning environment they are in. Moreover, in a field with biases and stereotypes about who belongs and can excel in it, motivational outcomes of underrepresented students such as women in physics are unlikely to show growth if the learning environment is not equitable and inclusive. Even though prior studies have shown that students' motivational beliefs can be influenced by many factors, there are very few studies focusing on how female and male students' motivational beliefs evolve in a year-long physics course sequence (e.g., college introductory physics sequence) in terms of the average scores and the predictive relationships among them.

Here we discuss an investigation focusing on female and male students' motivational beliefs in a two-semester college calculus-based introductory physics sequence (including physics 1 and physics 2) at a large public research university in the U.S. Each semester in the U.S. is generally 15–17 weeks long, and at the university discussed in this study, it is 15 weeks long. All full time students take classes in fall and spring semesters, and summer semesters are optional for students. Students who enrolled in the calculus-based introductory physics sequence are mostly majoring in engineering, physical sciences, and mathematics, and the survey shows that almost all students in this course sequence at our university had already learned at least some physics 1 topics in high school. Even though both courses are traditionally taught lecture-based courses with similar assessment style, there are several factors that may lead to progression in students' motivational beliefs. For example, students in this course sequence usually take physics 1 in the first semester and physics 2 in the second semester of their first year of undergraduate studies. Therefore, in physics 2, after students have been on campus for a semester, they may feel more comfortable and familiar with the way college physics courses are taught and how to interact with their instructors and classmates than when they were in physics 1, and thus the uncertainty and anxiety during the transition to college in their first semester may decrease, which can potentially impact their physics motivational beliefs [39,40,55]. In addition, physics 1 includes topics such as kinematics, forces, energy and work, while physics 2 includes topics such as electricity and magnetism, electromagnetic waves, and interference and diffraction of light. Also, not only are the physics 1 concepts more familiar from everyday life and less abstract, students were more familiar with the topics in physics 1 than those in physics 2 because, as noted earlier, most students in calculus-based introductory courses at our university had already learned at least some physics 1 topics in high school. Thus, we conducted a study focusing on how female and male students' physics motivational beliefs evolve in this two-semester introductory course sequence in terms of the average scores and predictive relationships among them.

To study the predictive relationships among students' physics motivational beliefs quantitatively, we adapt the physics identity model from prior work [22,27], in which the relationship between gender and identity is mediated through self-efficacy, interest, and perceived recognition. As shown in Fig. 1(a), we first considered a model (model 1) in which there are only covariances between perceived recognition (recog), self-efficacy (SE) and interest, so this model does not make assumptions about predictive relationships between these three mediating constructs. Then, we considered another model (model 2) in which perceived recognition is the predictor of both self-efficacy and interest [Fig. 1(b)], which is similar to the model in Kalender *et al.*.'s prior work [27]. Because of societal stereotypes about physics, interest may be thought to be fixed by many people. However, prior studies suggest that ones' interest can be shaped by their self-efficacy [56–58]. Therefore, in model 2, self-efficacy is the predictor of interest, which is different from the model used in the prior work [27], in which interest is the predictor of self-efficacy. We fit both models with the data collected at the end of physics 1 and physics 2, and then we compared the predictive relationships among the constructs in the two courses.

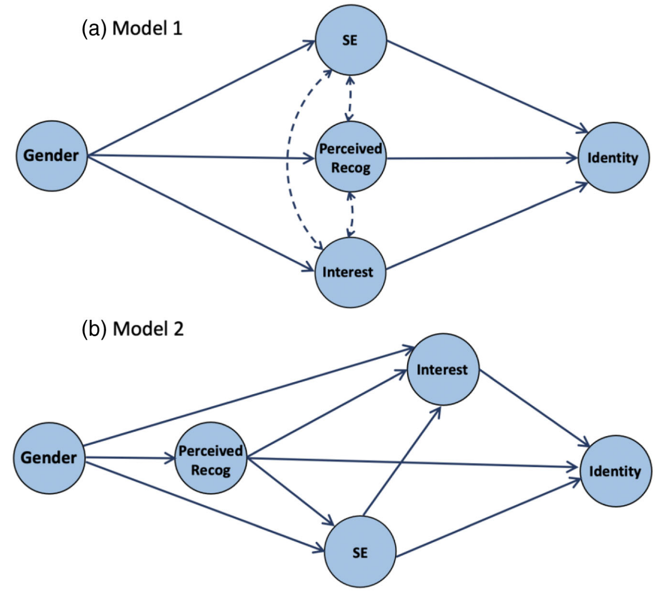

FIG. 1. Schematic representation of the path analysis part of the SEM models that shows how the relationship between gender and physics identity is mediated by perceived recognition (recog), self-efficacy (SE), and interest. (a) In model 1, all the three predictors are correlated with each other. (b) In model 2, perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest. The direct path from gender to identity is not shown because it is not statistically significant in both models for both courses.

II. RESEARCH QUESTIONS

Our research questions regarding matched students in a two-semester calculus-based introductory physics course sequence (in the recommended sequence with physics 1 in the fall semester and physics 2 in the spring semester) at a large public research university in the U.S. are as follows. These courses are required by engineering, physical science, and mathematics majors in the first year of their undergraduate studies.

RQ1. How do male and female students' physics motivational beliefs (including physics self-efficacy, interest, perceived recognition, and identity) change from physics 1 to physics 2?

RQ2. Are there gender differences in students' motivational beliefs and do they change from physics 1 to physics 2?

RQ3. How do perceived recognition, self-efficacy, and interest mediate the relation between gender and identity in physics 1 and physics 2?

RQ4. How do the predictive relationships among students' motivational beliefs change from physics 1 to physics 2?

III. METHODOLOGY

A. Participants

The motivational survey data used in this study were collected at the end of each course of a two-semester college calculus-based introductory physics sequence (including physics 1 and physics 2) in two consecutive school years at a large research university in the U.S. These courses are taken mostly by students majoring in engineering, physical sciences, and mathematics for whom they are mandatory. Because majority of students in these courses are required to take both physics courses, the populations of students in the two courses are almost the same. There were 1203 students in physics 1 and 921 students in physics 2 participating in the survey. In this study, we focused on 695 students (233 female students and 462 male students) who took the survey in both courses in recommended semesters, i.e., physics 1 in fall semester and physics 2 in spring semester because we wanted to track the same group of students' motivational beliefs in the two courses in the recommended sequence. Some possible reasons that some students took these courses in the off semesters (not recommended semesters) include students taking Advanced Placement (AP) physics in high school with scores that exempted them from college physics 1 and allowed them to directly enroll in physics 2 in their first semester, students repeating physics 1 in the off semester if they did not perform well the first time, and students putting off taking at least one of these courses in the summer semester due to their heavy course load in fall and spring semesters. Physics 1 mainly includes mechanics, while the main content of physics 2 is electricity and magnetism and optics. Both physics 1 and physics 2 are traditionally taught lecture-based courses (4 h per week) with recitations (1 h per week) and both courses were taught in person in the years studied. In the recitations, teaching assistants (TAs) answer students' homework questions and students typically work collaboratively on physics problems in part of the available time. Physics 1 was taught by six instructors and physics 2 was taught by seven instructors. Two of these instructors taught both physics 1 and physics 2. There were 8 classes for each course, and some instructors taught more than one class. We performed hierarchical linear modeling (HLM) to test the instructor level effects on students' motivational beliefs and the results show that interclass correlation coefficient (ICC) is around 0.01 for all motivational beliefs studied [59]. Since the ICC values are significantly smaller than 0.1, the instructor level effects can be ignored [59]. In physics 1, there were 52 recitation sessions, and 3 sessions (6%) were taught by female TAs. In physics 2, there were 49 recitation sessions, and 1 session (2%) was taught by female TA. Therefore, the gender balance of TAs is similar in physics 1 and physics 2. In addition, the assessment styles are also similar in physics 1 and physics 2, which are largely based on students' performance on midterm and final exams, which mainly focus on quantitative problem solving. Moreover, there was very little focus on using evidence-based pedagogies or intentional efforts to promote equity and inclusion in both courses.

This research protocol was approved and carried out in accordance with the principles outlined in the university institutional review board ethical policy. The paper surveys were handed out and collected by TAs during the last recitation class of a semester. Course instructors were encouraged to give students course credit or extra credit for completing the survey. Since this survey is also part of the department assessment survey, students' responses were first sent to an honest broker who was an expert in merging such data with demographics from provost's office as well as in de-identifying data before providing it to us for research. The honest broker generated a unique new ID for each student (which connected students' survey responses with their demographic information), so researchers could analyze students' data without having access to students' identifying information. We recognize that gender is fluid and on a spectrum rather than a binary construct. However, because students' gender information was obtained from the university, which offered binary options, we did the analysis with the binary gender data.

B. Survey instruments

In this study, we considered four motivational constructs—physics self-efficacy, interest, perceived recognition, and identity. The survey items for each construct are listed in Table I. The survey items were adapted from the existing motivational research [42,60–64] and have been revalidated in our prior work [6,65–68]. The validation and refinement of the survey involved use of one-on-one interviews with students using a think-aloud protocol, exploratory and confirmatory factor analyses (EFA and CFA) [69], Pearson correlation between different constructs and Cronbach's alpha (which is a measure of the internal consistency of each construct with several items) [70–72]. As shown in Table I, factor loadings (lambda) of each construct in physics 1 and physics 2 are very similar.

TABLE I. Survey questions for each of the motivational constructs, along with CFA factor loadings for physics 1 and physics 2. Lambda (factor loading) represents the correlation between each item and its corresponding construct, and the square of Lambda for each item gives the fraction of its variance explained by the construct. All Lambdas shown in this table are statistically significant with p value <0.001.

	Lambda	
Construct and item	Physics 1	Physics 2
Physics identity		
I see myself as a physics person.	1.000	1.000
Physics self-efficacy		
I am able to help my classmates with physics in the laboratory or in recitation.	0.700	0.684
I understand concepts I have studied in physics.	0.703	0.755
If I study, I will do well on a physics test.	0.717	0.734
If I encounter a setback in a physics exam, I can overcome it.	0.667	0.727
Physics interest		
I wonder about how physics works[a]	0.676	0.687
In general, I find physics[b]	0.757	0.809
I want to know everything I can about physics.	0.786	0.827
I am curious about recent physics discoveries.	0.693	0.721
Physics perceived recognition		
My parents see me as a physics person.	0.914	0.870
My friends see me as a physics person.	0.895	0.904
My physics TA and/or instructor sees me as a physics person.	0.672	0.721

[a]The response options for this question are never, once a month, once a week, every day.
[b]The response options for this question are very boring, boring, interesting, very interesting.

In our survey, each item was scored on a 4-point Likert scale (1–4). Students were given a score from 1 to 4 with higher scores indicating greater levels of interest, self-efficacy, perceived recognition and identity. Physics self-efficacy represents students' belief about whether they can excel in physics. We had four items for self-efficacy and these items had the response scale "NO!, no, yes, YES!" (Cronbach's $\alpha = 0.79$ for self-efficacy in physics 1 and $\alpha = 0.81$ for self-efficacy in physics 2), which have been shown to have good psychometric properties and a low cognitive load while reading [61,73]. We also had four items for interest (Cronbach's $\alpha = 0.82$ for interest in physics 1, $\alpha = 0.84$ for interest in physics 2). The question "I wonder about how physics works" had temporal response options "never, once a month, once a week, every day", whereas the question "In general, I find physics" had response options "very boring, boring, interesting, very interesting." The remaining two items were answered on the "NO!, no, yes, YES!" scale. Physics identity corresponds to students' belief about whether they designate themselves as a physics person [22]. Perceived recognition corresponds to whether a student thinks other people see them as a physics person [2,22], and it includes three items which correspond to family, friends and TA or instructor (Cronbach's $\alpha = 0.86$ for perceived recognition in both physics 1 and physics 2). These items involved a four-point Likert response on the scale "strongly disagree, disagree, agree, and strongly agree" and they correspond to 1 to 4 points [74].

C. Quantitative analysis of survey data

We calculated the mean score for each motivational construct for each student. Then, we used a t-test to compare students' mean scores for each motivational construct in physics 1 and physics 2 as well as conducted an analysis of gender differences using descriptive statistics. We note that in our previous study [27], we checked the response option distances for our survey constructs by using item response theory (IRT) to support the use of means across ratings [75,76]. Even for this study, we performed IRT with the new data set to verify the validity of using means across ratings. The parametric grades response model (GRM) by using the R software package "mirt" was used to test the measurement precision of our response scale [77,78]. Some of the items have response scales of strongly disagree, disagree, agree, and strongly agree while other items had response scale "NO!, no, yes, YES!." GRM calculates the location parameter for each response and calculates the difference between the locations. For the first group—strongly disagree, disagree, agree, and strongly agree—the difference between the location parameters were 1.3 and 1.5 for physics 1 and 1.3 and 1.4 for physics 2. For the second group—"NO!, no, yes, YES!"—the difference between the location parameters were 1.5 and 2.1 for

TABLE II. Zeroth order correlation coefficients of the constructs in the mediation model.

	Physics 1				Physics 2			
Observed variable	1	2	3	4	1	2	3	4
1. Physics identity	...	...	...	...	...	...	...	...
2. Self-efficacy	0.66	...	...	...	0.71	...	...	...
3. Interest	0.73	0.59	...	...	0.67	0.64	...	...
4. Perceived recognition	0.81	0.67	0.70	...	0.83	0.71	0.66	...

physics 1 and 1.4 and 1.9 for physics 2. These results show that the numerical values for the location differences for item responses are comparable, which suggests that calculating the traditional mean score of items is reasonable [75,78]. Furthermore, we estimated the IRT-based scores with expected *a posteriori* (EAP) computation method for each construct, and the results are highly correlated with the mean scores (the correlation coefficients are > 0.97 for all constructs), which indicates that the use of mean scores is reasonable [75].

Next, we calculated the Pearson correlation coefficients pairwise between the constructs for physics 1 and physics 2 separately. As shown in Table II, the correlation coefficients between the constructs are very similar in physics 1 and physics 2. In particular, the correlation coefficients of all constructs are above 0.2, and most of them are below 0.8, which means even though they have strong correlations with each other, the correlations are not so high that the constructs could not be examined as separate constructs [79]. We note that the correlation coefficient between physics identity and perceived recognition is 0.81 for physics 1 and 0.83 for physics 2. This is consistent with the prior findings of Godwin *et al.* [23] and Kalender *et al.* [27] that perceived recognition (external identity) is the largest predictor of physics identity (internal identity).

Finally, we used structural equation modeling to analyze the predictive relationships among the constructs [80]. Compared with a multiple regression model, the advantage of SEM is that we can estimate all of the regression links for multiple outcomes and factor loadings for items simultaneously, which improves the statistical power [80]. The SEM includes two parts: confirmatory factor analysis (CFA) and path analysis. CFA is used to test how well the measured variables represent the constructs studied, so it is also called measurement model. In CFA, comparative fit index (CFI) > 0.9, Tucker-Lewis index (TLI) > 0.9, root mean square error of approximation (RMSEA) < 0.08, and standardized root mean square residual (SRMR) < 0.08 are considered as acceptable and RMSEA < 0.06 and SRMA < 0.06 are considered as a good fit [70]. In our study, CFI $= 0.970$, TLI $= 0.960$, RMSEA $= 0.062$, and SRMR $= 0.035$ for physics 1 and CFI $= 0.969$, TLI $= 0.958$, RMSEA $= 0.066$, and SRMR $= 0.035$ for physics 2, all of which represent good fits. Thus, there is additional quantitative support for dividing the constructs as proposed. Apart from CFA, the path analysis in SEM gives regression coefficients β for paths between each pair of constructs and the value of each β is a measure of the strength of that relationship.

The major assumptions associated with SEM include: correct model specification, sufficiently large sample size and no systematic missing data [81–83]. Our study is based on the identity model in which students' physics identity is predicted by their perceived recognition, self-efficacy, and interest. This model has been examined by many prior studies [22,23,27,30]. According to Kline, a typical sample size in studies where SEM is used is about 200 [81], so the sample size of our study ($N = 695$) is sufficiently large for SEM. Moreover, since we focus on students who were in both physics 1 and physics 2 (matched students from physics 1 to physics 2), there were no missing data in our study except a couple of students forgetting to respond to one survey item. In addition, a well fitted measurement model (CFA) is also very important for performing full SEM [84]. As noted, our data fit the measurement model very well. Moreover, Table I shows that all factor loadings are higher than 0.5, which is considered as acceptable [84], and most of them are higher than 0.7. This means that the constructs extract sufficient variance from the observed variables, which allows us to perform the SEM [85]. In this study, we used full information maximum likelihood (FIML) to estimate parameters. FIML estimation is often the default in SEM software (e.g., Mplus and lavaan) and it has been shown to produce unbiased parameters estimates with great statistical power [86]. In addition, we also performed the same SEM models using diagonally weighted least square estimation and the results are very similar to those of FIML. In this study, we reported the results of FIML.

We first analyzed the saturated SEM model that includes all possible links between different constructs, and then we used the modification indices to improve the model fit. We kept path links which were statistically significant in SEM path analysis. We fit the two SEM models (model 1 and model 2) shown in Fig. 1 with the data from the end of physics1 and physics 2 separately, and then compared the SEM path analysis results (predictive relationships among the constructs) for physics 1 and physics 2.

IV. RESULTS

A. Gender difference in motivational beliefs

Table III shows the descriptive statistics for students' motivational beliefs at the end of physics 1 and physics 2. As shown in Table III, female students had significantly lower scores in all of the four motivational constructs in both courses, and the effect sizes are all in the medium range [87]. In particular, female students' average scores pertaining to perceived recognition and physics identity

TABLE III. Descriptive statistics of female and male students' motivational beliefs (matched students in physics 1 and physics 2). The sample size is 695 (462 male students and 233 female students). Cohen suggested that a typical value $d \sim 0.2$ be considered a small effect size, $d \sim 0.5$ a medium effect size and $d \sim 0.8$ a large effect size.

	Self-efficacy		Statistics		Interest		Statistics	
Gender	Physics 1	Physics 2	p value	Cohen's d	Physics 1	Physics 2	p value	Cohen's d
Male	3.06	2.91	<0.001	0.29	3.14	3.00	<0.001	0.24
Female	2.83	2.65	<0.001	0.34	2.76	2.61	0.011	0.23
p value	<0.001	<0.001			<0.001	<0.001		
Cohen's d	0.48	0.47			0.65	0.64		

	Perceived recognition		Statistics		Identity		Statistics	
Gender	Physics 1	Physics 2	p value	Cohen's d	Physics 1	Physics 2	p value	Cohen's d
Male	2.74	2.72	0.654	0.03	2.75	2.68	0.190	0.09
Female	2.37	2.25	0.065	0.17	2.29	2.16	0.100	0.16
p value	<0.001	<0.001			<0.001	<0.001		
Cohen's d	0.52	0.68			0.57	0.63		

show that on average, female students did not think others see them as a physics person, and they did not see themselves as a physics person either. Furthermore, the effect sizes of gender differences in students' perceived recognition and identity increased from physics 1 to physics 2. The confidence interval for the gender difference in perceived recognition is (0.36, 0.68) for physics 1, and (0.51, 0.84) for physics 2. A prior study [88] shows that $p \leq 0.05$ when the overlap of the 95% confidence intervals is no more than about half the average margin of error, that is when proportion overlap is about 0.5 or less. In our case, the midpoint of the first confidence interval (0.36, 0.68) is 0.52, which is comparable to the lower bound of the second confidence interval 0.51. Thus, the widening gender gap in perceived recognition is on the boundary of statistical significance of $p \leq 0.05$. On the other hand, the confidence interval for gender difference in physics identity is (0.41, 0.73) for physics 1, and (0.47, 0.79) for physics 2. Thus, the widening gender gap in physics identity is not statistically significant.

When we compared students' motivational beliefs in the two courses, we found that from physics 1 to physics 2, both male and female students' self-efficacy and interest in physics significantly decreased, while there was no statistically significant change in students' perceived recognition and identity. We note that even though male students' self-efficacy and interest dropped in physics 2, they were still higher than female students' in physics 1. Moreover, although we focus on students' motivational beliefs at the end of the courses in the main text, readers who are also interested in motivational beliefs at the beginning of these courses (pre) for the same students for whom we have discussed the data here can see Appendix A for the descriptive statistics. Appendix A shows similar results as the results shown here in the sense that students' motivational beliefs either decreased or were unchanged from pre to post, and the gender differences disadvantaging women increased at the end of both courses. In addition, we report the percentages of students who selected each choice for each survey item in Appendix B, which show consistent results with the descriptive statistics shown in Table III.

B. SEM models mediated by motivational factors

Before performing gender mediation analysis, we first tested the gender moderation relations between each pair of the constructs using multigroup SEM (to investigate any interaction effects with gender), which includes testing of factor loadings, indicator intercepts, residual variances and regression coefficients. Results showed that in all our models, strong measurement invariance holds and there is no difference in any regression coefficients by gender, which allowed us to perform the gender mediation analysis using SEM. In addition, we calculated the standard deviation for each motivational construct in physics 1 and physics 2. Results showed that the standard deviations of all motivational constructs are roughly the same in the two courses. This means that if the predictive relationships among the constructs changed from physics 1 to physics 2, it is not because of changes in the standard deviations; instead, it means that the relationships themselves have changed.

For gender mediation, we first consider a model (model 1) in which there are only covariances between each pair of constructs: perceived recognition, self-efficacy, and interest. Thus, this model does not make assumptions about the predictive relationships between these three mediating constructs. We fit this model with our motivational survey data collected in physics 1 and physics 2 separately, and the results of the path analysis of the SEM model for each course are visually presented in Fig. 2. A summary of all direct and indirect effects can be found in Table IV (for physics 1) and Table V (for physics 2). The model fit indices suggest a good

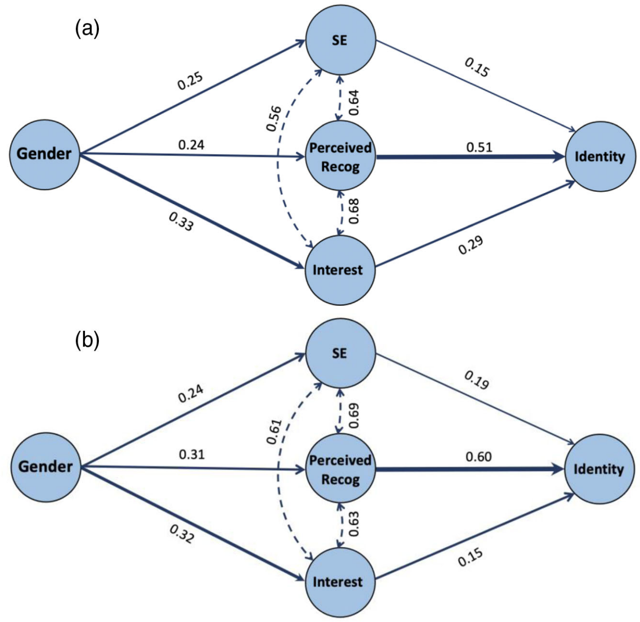

FIG. 2. Results of the path analysis part of SEM model 1, in which there are only covariances between each pair of constructs: perceived recognition (recog), self-efficacy (SE) and interest. (a) Shows the results of using physics 1 data, and (b) shows the results for physics 2 data. The dashed lines represent residual covariances between constructs. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients (β values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the β value. All β values shown are significant with $p < 0.001$.

fit to the data: For physics 1, CFI = 0.970 (>0.90), TLI = 0.959 (>0.90), RMSEA = 0.057 (<0.08), and SRMR = 0.033 (<0.08). For physics 2, CFI = 0.969 (>0.90),TLI = 0.959 (>0.90), RMSEA = 0.061 (<0.08), and SRMR = 0.034 (<0.08).

As shown in Fig. 2 or Tables IV and V, in both physics 1 and physics 2, there is a statistically significant regression line from gender to self-efficacy, perceived recognition, and interest, consistent with Table III showing that there are significant gender differences in all these three motivational constructs. However, the direct effect of gender on physics identity is statistically insignificant ($p = 0.76$ for physics 1, and $p = 0.90$ for physics 2) even though female students' identity is significantly lower than that of male students in both courses as shown in Table III. This result indicates that the relation between gender and physics identity is mediated by the other three motivational constructs (self-efficacy, interest, and perceived recognition). In addition, Fig. 2 shows that even though there is a strong covariance among self-efficacy, interest, and perceived recognition, students' perceived recognition is the strongest predictor of their physics identity ($\beta = 0.51$ for physics 1 and $\beta = 0.60$ for physics 2). This result is consistent with the prior work of Hazari *et al.* [22] and Kalender *et al.* [27].

TABLE IV. Results of the path analysis part of SEM model 1 for physics 1. Recog represents perceived recognition. SE represents self-efficacy. Single-headed arrows represent direct or indirect regression relationships. Double-headed arrows represent covariances.

Relationships	Unstandardized estimates	Standardized estimates	Standard error	p value
Direct effects				
Gender → SE	0.24	0.25	0.04	<0.001
Gender → interest	0.38	0.33	0.05	<0.001
Gender → recog	0.40	0.24	0.06	<0.001
Gender → identity	0.01	0.01	0.04	0.758
SE → identity	0.27	0.15	0.07	<0.001
Interest → identity	0.44	0.29	0.06	<0.001
Recog → identity	0.56	0.51	0.05	<0.001
Indirect effects				
Gender → identity	0.46	0.26	0.06	<0.001
Covariances				
SE ↔ recog	0.21	0.64	0.02	<0.001
SE ↔ interest	0.13	0.56	0.01	<0.001
Recog ↔ interest	0.26	0.68	0.02	<0.001

By comparing the SEM path analysis results of physics 1 [Fig. 2(a) or Table IV] and physics 2 [Fig. 2(b) or Table V], we found that the effect of gender on perceived recognition is larger in physics 2 ($\beta = 0.31$) than in physics 1 ($\beta = 0.24$). This result indicates that the gender difference in perceived recognition increased in physics 2, which is consistent with the descriptive statistics shown in Table III. In addition, we

TABLE V. Results of the path analysis part of SEM model 1 for physics 2. Recog represents perceived recognition. SE represents self-efficacy. Single-headed arrows represent direct or indirect regression relationships. Double-headed arrows represent covariances.

Relationships	Unstandardized estimates	Standardized estimates	Standard error	p value
Direct effects				
Gender → SE	0.26	0.24	0.05	<0.001
Gender → interest	0.38	0.32	0.05	<0.001
Gender → recog	0.49	0.31	0.06	<0.001
Gender → identity	0.01	0.00	0.04	0.901
SE → identity	0.33	0.19	0.07	<0.001
Interest → identity	0.23	0.15	0.06	<0.001
Recog → identity	0.70	0.60	0.05	<0.001
Indirect effects				
Gender → identity	0.52	0.28	0.06	<0.001
Covariances				
SE ↔ recog	0.24	0.69	0.02	<0.001
SE ↔ interest	0.16	0.61	0.02	<0.001
Recog ↔ interest	0.24	0.63	0.02	<0.001

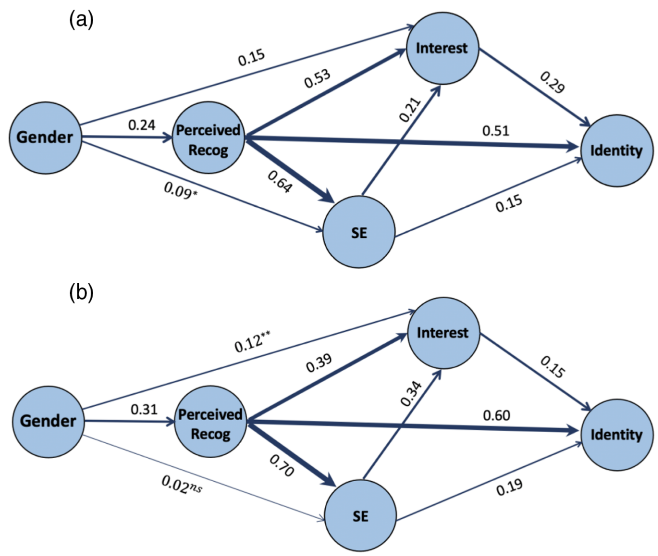

FIG. 3. Results of the path analysis part of SEM model 2, in which perceived recognition (Recog) predicts self-efficacy (SE) and interest, and self-efficacy predicts interest. (a) Shows the results of using physics 1 data, and (b) shows the results for physics 2 data. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients (β values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the β with $0.001 \leq p < 0.01$ indicated by **, $0.01 \leq p < 0.05$ indicated by *, and $p \geq 0.05$ indicated by ns (not significant). All the other regression lines show relations with $p < 0.001$.

note that the effect of perceived recognition on identity is also larger in physics 2 ($\beta = 0.60$) than that in physics 1 ($\beta = 0.51$), which means that, in physics 2, perceived recognition plays an even more important role in predicting identity. However, the effect of interest on identity is smaller in physics 2 ($\beta = 0.15$) than that in physics 1 ($\beta = 0.29$).

Next, we consider a model (model 2) in which perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest. We fit this model with our motivational survey data collected in physics 1 and physics 2 separately, and the results of the path analysis of the SEM model for each course are presented visually in Fig. 3. A summary of all direct and indirect effects can be found in Table VI (for physics 1) and Table VII (for physics 2). This model also fits the data very well: For physics 1, CFI $=$ 0.970 ($>$0.90), TLI $=$ 0.959 ($>$0.90), RMSEA $=$ 0.057 ($<$0.08), and SRMR $=$ 0.033 ($<$0.08). For physics 2, CFI $=$ 0.969 ($>$0.90), TLI $=$ 0.959 ($>$0.90), RMSEA $=$ 0.061 ($<$0.08), and SRMR $=$ 0.034 ($<$0.08). We note that, for both physics 1 and physics 2, the direct effect of gender on perceived recognition in model 2 is the same as that in model 1. This is because in both models, gender is the only predictor of perceived recognition. On the other hand, for both courses, the direct effects of gender on self-efficacy and interest are smaller or statistically insignificant in model 2 compared with those in model 1. This is because in model 2, self-efficacy and interest are predicted by more constructs than in model 1, and thus there is more correlated effect being controlled for when estimating the

TABLE VI. Results of the path analysis part of SEM model 2 for physics 1. Recog represents perceived recognition. SE represents self-efficacy. Single-headed arrows represent direct or indirect regression relationships.

Relationships	Unstandardized estimates	Standardized estimates	Standard error	p value
		Direct effects		
Gender → SE	0.09	0.09	0.04	0.011
Gender → interest	0.18	0.15	0.04	<0.001
Gender → recog	0.40	0.24	0.06	<0.001
Gender → identity	0.01	0.01	0.04	0.758
Recog → SE	0.38	0.64	0.03	<0.001
Recog → interest	0.37	0.53	0.04	<0.001
SE → interest	0.25	0.21	0.07	<0.001
SE → identity	0.27	0.15	0.07	<0.001
Interest → identity	0.44	0.29	0.06	<0.001
Recog → identity	0.56	0.51	0.05	<0.001
		Indirect effects		
Gender → SE	0.15	0.16	0.03	<0.001
Gender → interest	0.21	0.18	0.03	<0.001
Gender → identity	0.46	0.26	0.06	<0.001
Recog → interest	0.09	0.13	0.03	<0.001
Recog → identity	0.31	0.28	0.04	<0.001
SE → identity	0.11	0.06	0.03	0.001

TABLE VII. Results of the path analysis part of SEM model 2 for physics 2. Recog represents perceived recognition. SE represents self-efficacy. Single-headed arrows represent direct or indirect regression relationships.

Relationships	Unstandardized estimates	Standardized estimates	Standard error	p value
		Direct effects		
Gender → SE	0.02	0.02	0.04	0.530
Gender → interest	0.14	0.12	0.04	0.001
Gender → recog	0.49	0.31	0.06	<0.001
Gender → identity	0.01	0.00	0.04	0.901
Recog → SE	0.48	0.70	0.03	<0.001
Recog → interest	0.30	0.39	0.05	<0.001
SE → interest	0.38	0.34	0.07	<0.001
SE → identity	0.33	0.19	0.07	<0.001
Interest → identity	0.23	0.15	0.06	<0.001
Recog → identity	0.70	0.60	0.05	<0.001
		Indirect effects		
Gender → SE	0.24	0.22	0.03	<0.001
Gender → interest	0.25	0.20	0.04	<0.001
Gender → identity	0.52	0.28	0.06	<0.001
Recog → interest	0.18	0.24	0.03	<0.001
Recog → identity	0.27	0.23	0.04	<0.001
SE → identity	0.09	0.05	0.03	0.001

TABLE VIII. Coefficient of determination (R^2) for various constructs in different models. All R^2 values are significant and p values <0.001. In model 1, there are only covariances between each pair of constructs: physics self-efficacy, perceived recognition, and interest. In model 2, the arrows indicate the direction of the predictive relationships.

Models	Courses	Constructs	R^2
Model 1 SE+ recog + interest	Physics 1	Perceived recognition	0.06
		Self-efficacy	0.06
		Interest	0.11
		Identity	0.72
	Physics 2	Perceived recognition	0.10
		Self-efficacy	0.06
		Interest	0.10
		Identity	0.73
Model 2 Recog → SE → interest	Physics 1	Perceived recognition	0.06
		Self-efficacy	0.45
		Interest	0.54
		Identity	0.72
	Physics 2	Perceived recognition	0.10
		Self-efficacy	0.50
		Interest	0.51
		Identity	0.73

regression coefficients from gender to self-efficacy and interest.

By comparing the path analysis results of model 2 in physics 1 [Fig. 3(a) or Table VI] and physics 2 [Fig. 3(b) or Table VII], we found that the effect of gender on perceived recognition is larger in physics 2 than in physics 1. This result indicates that the gender difference in perceived recognition increased in physics 2, which is the same as what we found in model 1 (Fig. 2) as discussed earlier. We also found that the effect of perceived recognition on identity is larger in physics 2 than that in physics 1, while the effect of interest on identity became smaller in physics 2. In addition, Fig. 3 shows that gender directly predicts self-efficacy in physics 1, while this effect is not statistically significant in physics 2. Because the total effect of gender on self-efficacy is the same in physics 1 ($\beta = 0.24 \times 0.64 + 0.09 = 0.24$) and physics 2 ($\beta = 0.31 \times 0.70 + 0.02 = 0.24$), this result means that, in physics 2, more effect of gender on self-efficacy was mediated by perceived recognition. Likewise, we note that even though the total effect of perceived recognition on interest is similar in physics 1 and physics 2 ($\beta = 0.66$ in physics 1 and $\beta = 0.63$ in physics 2), the direct effect of perceived recognition on interest became smaller in physics 2 ($\beta = 0.53$ in physics 1 and $\beta = 0.39$ in physics 2), while the indirect effect mediated by self-efficacy became larger in physics 2 ($\beta = 0.64 \times 0.21 = 0.13$ in physics 1 and $\beta = 0.70 \times 0.34 = 0.24$ in physics 2). This means that more effect of perceived recognition on interest was mediated by self-efficacy in physics 2.

To further understand the relationships among the motivational constructs in different models and in different courses, we calculated the coefficients of determination R squared (fraction of variance explained) for each construct in each model using physics 1 and physics 2 data separately (Table VIII). We note that the R^2 of physics identity is 0.72 in physics 1 and 0.73 in physics 2 in both model 1 and model 2. This is because in both models, identity is predicted by all the other constructs even though the predictive relationships among these predictors are different. Similarly, since perceived recognition is only predicted by gender in both models, the R^2 of perceived recognition in each course is the same across models. On the other hand, for both physics 1 and physics 2, the R^2 of self-efficacy and interest are larger in model 2 than in model 1. This is because in model 2, self-efficacy and interest are predicted by more constructs than they are in model 1, and thus more variance in self-efficacy and interest is explained by model 2. Even though the R^2 values of self-efficacy and interest are different in model 1 and model 2 as discussed above, R^2 of each construct in each model is very similar in physics 1 and physics 2.

V. SUMMARY AND DISCUSSION

In this study, we investigated progression in female and male students' physics motivational beliefs in a traditionally taught two-semester college calculus-based introductory physics sequence. In particular, we focused on whether and how students' physics motivational beliefs evolve from physics 1 to physics 2 in terms of not only the average score on each motivational construct but also the predictive relationships among them. To quantitatively analyze the predictive relationships, we adapted a prior identity framework [22,27] and performed structural equation modeling for students' motivational beliefs in each course.

Although the first semester in physics 1 consisted of high school to college transition period and physics 1 also had different content than physics 2, our research suggests that the big picture trends are very similar in both physics 1 and physics 2. We found that female students had significantly lower scores in all of the motivational constructs than male students in both physics 1 and physics 2.

Even though the larger trends are similar in the two courses, there are some changes from physics 1 to physics 2. For example, gender differences in perceived recognition actually increased in physics 2 (Table III). In addition, we found that both female and male students' self-efficacy and interest significantly decreased in physics 2. These results indicate that the current learning environment in these traditionally taught courses is not helping to improve students' motivational beliefs and the gender differences became larger. In our prior interviews with students in this course sequence, some interviewed female students noted that their instructors or TAs sometimes showed more interest in male students' questions and answered male students' questions with more attention than when they answered their questions [9,20]. The interviewed female students also reported that men in their physics courses were generally praised more by the instructors or TAs than women, and sometimes instructors or TAs called men who answered the questions "brilliant," which made them feel as though they were not brilliant [9,20]. Because of societal stereotypes and biases about who belongs in physics and who can do well in physics, female students may not have received enough recognition and encouragement even before they entered the college physics courses, e.g., from their high school counselors, teachers and other stakeholders. Our study indicates that without intentional efforts to improve students' motivational beliefs and promote equity, the current learning environment may further impact students' motivational beliefs and be particularly detrimental to the underrepresented students such as women.

In terms of the predictive relationships among the motivational constructs, we found that the SEM models for physics 1 and physics 2 are qualitatively similar. In both courses, the relationship between gender and physics identity is mediated through self-efficacy, interest and perceived recognition. Moreover, students' physics identity is predicted by their self-efficacy, interest and perceived recognition, with perceived recognition being the strongest predictor, which confirms perceived recognition as a key factor of physics identity throughout. We note that in physics 2, perceived recognition plays an even more important role not only in predicting students' identity but also in mediating the gender difference in students' self-efficacy. One hypothesis that may at least partly explain these findings is that in a more abstract course that students have less prior exposure to, they may need even more encouragement and recognition to help them build self-efficacy and identity in physics, and thus instructors and TAs may play an even more important role in supporting and affirming their students as they learn. This is particularly important for students from underrepresented groups such as women, who have few role models and may also experience stereotype threat in these physics courses. Also, our results show that the gender difference in perceived recognition actually increased from physics 1 to physics 2 (and also increased from pre to post in both courses as shown in Appendix A), which means that the current learning environment disadvantages female students more than male students and is therefore not equitable and inclusive.

Our findings suggest that it is important to make intentional efforts to create an inclusive and equitable learning environment, in which not only should all students have equitable opportunities and access to resources, but they are also recognized and supported appropriately. A study found that the synergy between explicitly and implicitly recognizing strategies used by instructors is a critical feature of effective recognition that can be internalized effectively by the students [89]. For example, instructors can explicitly recognize students by directly acknowledging their work and expressing faith in their ability to excel. They can also implicitly recognize students by valuing students' opinions and assigning a leadership position or a challenging task to students in small groups that makes them feel excited [89]. In addition to positive recognition, instructors should be careful not to give unintended messages to students, e.g., praising some students for brilliance or intelligence as opposed to their effort since it may convey to other students that they do not have what is required to excel in physics. In addition, when students ask instructors for help on physics problems, if instructors inadvertently label those problems as "easy," "trivial," or "obvious," it can also make students feel disparaged [90]. What is important for instructors to internalize is that it is not their intentions that matter but the impact they are having on the students.

In this study, both physics 1 and physics 2 are traditional lecture-based courses. Prior studies show that in lecture-based courses, students are often passive, and they may not have enough opportunities to let instructors know their extent of learning [91,92]. Research in physics education has shown that traditional lectures, even when perceived as good lectures, have limited success in helping students learn physics [93–95]. Therefore, it could help if instructors incorporate more research-validated active engagement pedagogies in their class, but that is not enough to make physics courses equitable and inclusive. According to prior studies [19,96], active engagement in an inequitable learning environment actually can increase the gender gap in students' performance because the stereotyped group (e.g., women) may not feel safe to participate without feeling judged or anxious if the environment is not equitable and inclusive. Therefore, instructors need to keep in mind how the societal stereotypes and biases about who belongs in physics and who can excel in it impact the stereotyped groups. They should be mindful of students' motivational beliefs and have an explicit goal of improving these in their physics courses by supporting and recognizing their students appropriately. They should also have an explicit equity goal and strive to close the gap between underrepresented students in physics such as women and students from the dominant groups. There are some research-based classroom interventions that have been shown to enhance students' self-efficacy and reduce gender gaps in students' performance [97–100]. Instructors can also tailor these short interventions in their classes to help all students develop positive motivational beliefs and learn physics equitably.

VI. LIMITATIONS AND FUTURE DIRECTIONS

In this study, we investigated possible evolution in students' motivational beliefs in a traditionally taught college introductory physics course sequence. To better understand how students' experience in each course influence their motivational beliefs, we also compared students' motivational beliefs at the beginning and the end of each course (see Appendix A for detailed results). However, since perceived recognition and identity constructs were included in our survey at the end of physics 1 in the first year of study, we do not have the pre-perceived recognition and pre-identity data for that semester. Thus, for physics 1, we only present students' pre- and postmotivational beliefs in the second year studied in Appendix A with 291 students. Although the size of this sample is reasonable, in our future studies, we will further investigate students' motivational beliefs at the beginning and end of this course with a larger sample size. In addition, in this study, we did not track the data for which students were assigned to which TAs, so we could not test the effect of gender balance of TAs on students' motivational beliefs. In future studies, it would be helpful to track gender of the TAs and analyze data to test if the gender balance of TAs impacted students' motivational beliefs.

In this study, we found that even though physics 1 and physics 2 have different class content, the larger trends of students' motivational beliefs and the relationships among these beliefs in these two traditional taught courses are similar. This result is based on students' self-reported responses to the motivational survey. It would be helpful to interview more students to get a deeper qualitative understanding of what they experience during the learning process in physics 1 and physics 2, how their experiences shape their physics motivational beliefs in the two courses, and whether and how the class content impacts their motivational beliefs in a nuanced manner. Moreover, future studies can also investigate students' motivational beliefs in active learning classes and the classes in which there is an intentional focus on equity and inclusion to see if class content in those class settings impacts students' motivational beliefs.

Our study was conducted in a large public research university in the U.S., and the introductory physics courses discussed here are very similar in these large universities. Therefore, we believe that the results are likely to be generalizable to other similar universities in the U.S. More studies should be conducted in different types of institutions such as small colleges and universities in the U.S. and in other countries to see if similar results are obtained. In addition, as noted, this study was conducted in a traditionally taught introductory calculus-based physics course sequence. It would be interesting to investigate how different teaching approaches and class formats, such as studio physics class, affect students' motivational beliefs and the predictive relationships among them. It would also be interesting to investigate students' motivational beliefs in algebra-based physics course sequence for bioscience majors where women are usually overrepresented.

ACKNOWLEDGMENTS

This work was supported by Grant No. DUE-1524575 from the National Science Foundation. We would like to thank all students whose data were analyzed and Dr. Robert P. Devaty for his constructive feedback on the manuscript.

APPENDIX A: STUDENTS' PRE- AND POSTMOTIVATIONAL BELIEFS

In the main text, we focused on students' motivational beliefs at the end of physics 1 and physics 2. Here, we present the descriptive statistics of students' motivational beliefs at both the beginning and end of each course. The results presented here are from the same students for whom we have discussed the data in the main text. We matched students' responses from pre to post. Since not all of them took the survey at the beginning of the courses, the sample sizes presented here are somewhat smaller than that in the main text. In addition, because the perceived recognition and identity constructs were included to our survey at the end of physics 1 in the first year of study, we do not have the pre-perceived recognition and pre-identity data for that semester. Thus, for physics 1, we only present students' pre- and postmotivational beliefs in the second year studied (Table IX).

TABLE IX. Descriptive statistics of female and male students' motivational beliefs in physics 1 in the second year studied. The sample size is 291 (179 male students and 112 female students). Cohen suggested that a typical value $d \sim 0.2$ be considered a small effect size, $d \sim 0.5$ represents a medium effect size and $d \sim 0.8$ a large effect size.

Physics 1	Self-efficacy		Statistics		Interest		Statistics	
Gender	pre	post	p value	Cohen's d	pre	post	p value	Cohen's d
Male	3.15	3.07	0.121	0.17	3.08	3.08	0.946	0.01
Female	3.06	2.90	0.006	0.37	2.89	2.78	0.154	0.19
p value	0.070	0.003			0.003	<0.001		
Cohen's d	0.22	0.37			0.36	0.49		
	Perceived recognition		Statistics		Identity		Statistics	
Gender	pre	post	p value	Cohen's d	pre	post	p value	Cohen's d
Male	2.74	2.73	0.898	0.01	2.84	2.73	0.195	0.14
Female	2.54	2.47	0.442	0.10	2.54	2.40	0.175	0.18
p value	0.012	0.002			0.001	0.001		
Cohen's d	0.31	0.38			0.39	0.40		

TABLE X. Descriptive statistics of female and male students' motivational beliefs in physics 2 in both years studied. The sample size is 626 (411 male students and 215 female students).

Physics 2	Self-efficacy		Statistics		Interest		Statistics	
Gender	pre	post	p value	Cohen's d	pre	post	p value	Cohen's d
Male	3.08	2.90	<0.001	0.37	3.09	3.01	0.038	0.14
Female	2.87	2.65	<0.001	0.43	2.73	2.62	0.069	0.18
p value	<0.001	<0.001			<0.001	<0.001		
Cohen's d	0.48	0.45			0.64	0.64		
	Perceived recognition		Statistics		Identity		Statistics	
Gender	pre	post	p value	Cohen's d	pre	post	p value	Cohen's d
Male	2.76	2.71	0.257	0.08	2.75	2.69	0.306	0.07
Female	2.41	2.24	0.011	0.24	2.28	2.15	0.096	0.16
p value	<0.001	<0.001			<0.001	<0.001		
Cohen's d	0.53	0.68			0.57	0.65		

As shown in Tables IX and X, female students had lower average motivational beliefs at the beginning of both physics 1 and physics 2, and the gender differences in most constructs increased at the end of the courses. In addition, both female and male students' motivational beliefs are lower in post than in pre (even though not all of which are statistically significant), and the effect sizes are larger for women.

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

In the main text, we investigated how students' motivational beliefs change from physics 1 to physics 2 by comparing their average scores on the motivational constructs in the two courses. Here, we present the percentages of female (Table XI) and male students (Table XII) who selected each answer choice from a 4-point Likert scale for each survey item. Students were given a score from 1 to 4, respectively, with higher scores indicating greater levels of interest, self-efficacy, perceived recognition, and identity.

As shown in Tables XI and XII, for both female and male students, the percentages of students who selected 3 or 4 decreased from physics 1 to physics 2 for most survey items, while the percentages of students who selected 1 or 2 increased. In particular, these shifts were larger in self-efficacy and interest items than in perceived recognition and identity items. These results are consistent with the

TABLE XI. Percentages of female students who selected each choice from a 4-point Likert scale for each survey item in physics 1 and physics 2. The self-efficacy and interest (Int) items have the response scale: 1 = NO!, 2 = no, 3 = yes, and 4 = YES!, while the perceived recognition and identity (Idt1) items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

	Physics 1				Physics 2			
Survey items	1	2	3	4	1	2	3	4
SE1	5%	24%	63%	8%	9%	37%	49%	5%
SE2	3%	13%	76%	9%	5%	25%	65%	5%
SE3	4%	22%	60%	14%	7%	30%	52%	12%
SE4	3%	22%	63%	12%	4%	27%	61%	9%
Int1	8%	19%	49%	24%	8%	29%	43%	20%
Int2	6%	16%	64%	14%	9%	26%	53%	12%
Int3	6%	40%	42%	11%	10%	45%	38%	7%
Int4	6%	34%	46%	14%	9%	33%	47%	11%
Recog1	15%	38%	38%	9%	18%	40%	35%	6%
Recog2	13%	36%	40%	10%	15%	45%	33%	6%
Recog3	17%	50%	31%	3%	17%	53%	28%	2%
Idt1	18%	43%	31%	8%	21%	46%	27%	5%

TABLE XII. Percentages of male students who selected each choice from a 4-point Likert scale for each survey item in physics 1 and physics 2. The self-efficacy and interest items have the response scale: 1 = NO!, 2 = no, 3 = yes, and 4 = YES!, while the perceived recognition and identity items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.

	Physics 1				Physics 2			
Survey items	1	2	3	4	1	2	3	4
SE1	2%	16%	67%	15%	5%	28%	55%	12%
SE2	1%	8%	70%	22%	2%	17%	67%	14%
SE3	2%	11%	58%	30%	3%	19%	51%	27%
SE4	1%	14%	68%	17%	2%	19%	59%	21%
Int1	3%	8%	43%	45%	3%	15%	48%	34%
Int2	1%	7%	61%	30%	4%	12%	62%	23%
Int3	2%	20%	52%	26%	3%	23%	55%	19%
Int4	3%	18%	52%	26%	3%	20%	55%	22%
Recog1	5%	25%	48%	21%	5%	27%	47%	20%
Recog2	5%	28%	47%	20%	5%	30%	47%	18%
Recog3	8%	37%	48%	7%	7%	41%	42%	10%
Idt1	6%	30%	48%	16%	8%	33%	42%	17%

descriptive statics shown in Table III, which show that both male and female students' self-efficacy and interest statistically significantly decreased from physics 1 to physics 2, while the decreases in perceived recognition and identity were not statically significant. In addition, by comparing Tables XI and XII, we found that the percentages of female students who selected 1 or 2 were larger than those of male students, while the percentages of female students who selected 4 were smaller than those of female students. These findings are also consistent with Table III showing that there were statistically significant gender differences in all motivational constructs studied.

[1] 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).

[2] 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).

[3] L. J. Sax, K. J. Lehman, R. S. Barthelemy, and G. Lim, Women in physics: A comparison to science, technology, engineering, and math education over four decades, Phys. Rev. Phys. Educ. Res. **12**, 020108 (2016).

[4] J. Blue, A. L. Traxler, and X. C. Cid, Gender matters, Phys. Today **71**, 40 (2018).

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

[6] E. M. 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).

[7] B. Van Dusen and J. Nissen, Equity in college physics student learning: A critical quantitative intersectionality investigation, J. Res. Sci. Teach. **57**, 33 (2020).

[8] M. Lorenzo, C. H. Crouch, and E. Mazur, Reducing the gender gap in the physics classroom, Am. J. Phys. **74**, 118 (2006).

[9] 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).

[10] R. Ivie, S. White, and R. Y. Chu, Women's and men's career choices in astronomy and astrophysics, Phys. Rev. Phys. Educ. Res. **12**, 020109 (2016).

[11] 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).

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

[13] H. M. Watt, The role of motivation in gendered educational and occupational trajectories related to maths, Educ. Res. Eval. **12**, 305 (2006).

[14] D. H. Schunk and F. Pajares, The development of academic self-efficacy, in *Development of Achievement Motivation: A Volume in the Educational Psychology Series*, edited by A. Wigfield and J. S. Eccles (Academic Press, San Diego, 2002), p. 15.

[15] 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).

[16] E. Seymour, Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology, Sci. Educ. **86**, 79 (2002).

[17] C. Hill, C. Corbett, and A. St. Rose, *Why so Few? Women in Science, Technology, Engineering, and Mathematics* (ERIC, Washington, DC, 2010).

[18] A. B. Diekman, E. K. Clark, A. M. Johnston, E. R. Brown, and M. Steinberg, Malleability in communal goals and beliefs influences attraction to STEM careers: Evidence for a goal congruity perspective, J. Pers. Soc. Psychol. **101**, 902 (2011).

[19] 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).

[20] 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).

[21] N. Abramzon, P. Benson, E. Bertschinger, S. Blessing, G. L. Cochran, A. Cox, B. Cunningham, J. Galbraith-Frew, J. Johnson, L. Kerby, E. Lalanne, C. O'Donnell, S. Petty, S. Sampath, S. Seestrom, C. Singh, C. Spencer, K. S. Woodle, and S. Yennello, Women in physics in the United States: Recruitment and retention, AIP Conf. Proc. **1697**, 060045 (2015).

[22] 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).

[23] 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. Engin. Educ. **105**, 312 (2016).

[24] 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).

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

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

[27] 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).
[28] 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).
[29] R. M. Lock, Z. Hazari, and G. Potvin, Impact of out-of-class science and engineering activities on physics identity and career intentions, Phys. Rev. Phys. Educ. Res. **15**, 020137 (2019).
[30] 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).
[31] C. Good, A. Rattan, and C. S. Dweck, Why do women opt out? Sense of belonging and women's representation in mathematics, J. Pers. Soc. Psychol. **102**, 700 (2012).
[32] J. D. Cribbs, Z. Hazari, G. Sonnert, and P. M. Sadler, Establishing an explanatory model for mathematics identity, Child Development **86**, 1048 (2015).
[33] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, Understanding engineering identity through structural equation modeling, in *Proceedings of the 2013 IEEE Frontiers in Education Conference (FIE)* (IEEE, Bellingham, WA, 2013), p. 50.
[34] A. Godwin and A. Kirn, Identity-based motivation: Connections between first-year students' engineering role identities and future-time perspectives, J. Engin. Educ. **109**, 362 (2020).
[35] D. Verdín, A. Godwin, A. Kirn, L. Benson, and G. Potvin, Understanding how engineering identity and belongingness predict grit for first-generation college students, in *Proceedings of the 2018 Collaborative Network for Engineering and Computing Diversity Conference, Crystal City, United States* (American Society for Engineering Education, Crystal, 2018).
[36] S. Harter, *The Construction of the Self: A Developmental Perspective* (Guilford Press, New York, NY, 1999).
[37] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, Psychol. Rev. **84**, 191 (1977).
[38] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice Hall, Englewood Cliffs, NJ, 1986).
[39] A. Bandura, W. Freeman, and R. Lightsey, Self-efficacy: The exercise of control, J. Cogn. Psychotherapy **13**, 158 (1999).
[40] A. Bandura, Self-efficacy, in *Encyclopedia of Psychology*, 2nd ed., edited by R. J. Corsini (Wiley, New York, 1994), Vol. 3, p. 368.
[41] K. Miller, J. Schell, A. Ho, B. Lukoff, and E. Mazur, Response switching and self-efficacy in Peer Instruction classrooms, Phys. Rev. ST Phys. Educ. Res. **11**, 010104 (2015).
[42] 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).
[43] S. Hidi, Interest: A unique motivational variable, Educ. Res. Rev. **1**, 69 (2006).
[44] 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).
[45] A. Wigfield and J. S. Eccles, Expectancy-value theory of achievement motivation, Contemp. Educ. Psychol. **25**, 68 (2000).
[46] A. Wigfield and J. S. Eccles, The development of achievement task values: A theoretical analysis, Dev. Policy Rev. **12**, 265 (1992).
[47] J. E. Côté and C. G. Levine, *Identity, Formation, Agency, and Culture: A Social Psychological Synthesis* (Psychology Press, Hove, UK, 2014).
[48] 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).
[49] L. Bian, S.-J. Leslie, and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, Science **355**, 389 (2017).
[50] K. Crowley, M. A. Callanan, H. R. Tenenbaum, and E. Allen, Parents explain more often to boys than to girls during shared scientific thinking, Psychol. Sci. **12**, 258 (2001).
[51] 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).
[52] S. Upson and L. F. Friedman, Where are all the female geniuses?, Scientific American Mind **23**, 63 (2012), https://www.jstor.org/stable/24942297.
[53] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).
[54] 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).
[55] A. Bandura, The evolution of social cognitive theory, in *Great Minds in Management*, edited by K. G. Smith and M. A. Hitt (Oxford University Press, Oxford, 2005), p. 9.
[56] P. J. Silvia, Interest and interests: The psychology of constructive capriciousness, Rev. Gen. Psychology **5**, 270 (2001).
[57] P. J. Silvia, Self-efficacy and interest: Experimental studies of optimal incompetence, J. Vocat. Behav. **62**, 237 (2003).
[58] R. W. Lent, S. D. Brown, and G. Hackett, Toward a unifying social cognitive theory of career and academic interest, choice, and performance, J. Vocat. Behav. **45**, 79 (1994).
[59] A. S. Bryk and S. W. Raudenbush, *Hierarchical Linear Models: Applications and Data Analysis Methods* (Sage Publications, Inc., Newbury Park, CA, 1992).
[60] D. Hammer, Epistemological beliefs in introductory physics, Cognit. Instr. **12**, 151 (1994).

[61] Learning Activation Lab, Activation lab tools: Measures, and data collection instruments (2017), http://www.activationlab.org/tools/.

[62] B. M. Zwickl, T. Hirokawa, N. Finkelstein, and H. J. Lewandowski, Epistemology and expectations survey about experimental physics: Development and initial results, Phys. Rev. ST Phys. Educ. Res. **10**, 010120 (2014).

[63] J. Schell and B. Lukoff, Peer instruction self-efficacy instrument [Developed at Harvard University] (unpublished) (2010).

[64] PERTS Academic Mindsets Assessment (2020), https://www.perts.net/orientation/ascend.

[65] Z. Y. Kalender, E. Marshman, T. J. Nokes-Malach, C. D. Schunn, and C. Singh, Motivational characteristics of underrepresented ethnic and racial minority students in introductory physics courses, presented at PER Conf. 2017, Cincinnati, OH, 10.1119/perc.2017.pr.046.

[66] T. Nokes-Malach, E. Marshman, Z. Y. Kalender, C. Schunn, and C. Singh, Investigation of male and female students' motivational characteristics throughout an introductory physics course sequence, presented at PER Conf. 2017, Cincinnati, OH, 10.1119/perc.2017.pr.064.

[67] 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, presented at PER Conf. 2018, Washington, DC, 10.1119/perc.2018.pr.Nokes-Malach.

[68] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Large gender differences in physics self-efficacy at equal performance levels: A warning sign?, presented at PER Conf. 2018, Washington, DC, 10.1119/perc.2018.pr.Kalender.

[69] B. Thompson, *Exploratory and Confirmatory Factor Analysis* (American Psychological Association, Washington, DC, 2004).

[70] D. Hooper, J. Coughlan, and M. Mullen, Structural equation modeling: Guidelines for determining model fit, Electronic J. Bus. Res. Methods **6**, 53 (2007).

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

[72] K. Pearson and F. Galton, VII. Note on regression and inheritance in the case of two parents, Proc. R. Soc. London **58**, 240 (1895).

[73] 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).

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

[75] S. E. Embretson and S. P. Reise, *Item Response Theory for Psychologists* (Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 2000).

[76] https://www.stata.com/meeting/australia15/abstracts/materials/oceania15_rosier.pdf.

[77] R. P. Chalmers, mirt: A multidimensional item response theory package for the R environment, J. Stat. Softw. **48**, 1 (2012).

[78] F. Samejima, Estimation of latent ability using a response pattern of graded scores, *Psychometrika Monograph* (Psychometric Society, Richmond, VA, 1969), p. 17.

[79] H. Akoglu, User's guide to correlation coefficients, Turkish J. Emerg. Med. **18**, 91 (2018).

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

[81] R. B. Kline, *Principles and Practice of Structural Equation Modeling* (Guilford Publications, New York, 2015).

[82] D. Kaplan, *Structural Equation Modeling: Foundations and Extensions* (Sage Publications, Newbury Park, CA, 2008), Vol. 10.

[83] D. Kaplan, Structural equation modeling, in *International Encyclopedia of the Social, and Behavioral Sciences*, edited by N. J. Smelser and P. B. Baltes (Pergamon, Oxford, 2001), p. 15215.

[84] Z. Awang, *SEM Made Simple: A Gentle Approach to Learning Structural Equation Modeling* (MPWS Rich Publication, Bangi, 2015).

[85] A. G. Yong and S. Pearce, A beginner's guide to factor analysis: Focusing on exploratory factor analysis, Tutorials Quant. Methods Psychol. **9**, 79 (2013).

[86] C. K. Enders and D. L. Bandalos, The relative performance of full information maximum likelihood estimation for missing data in structural equation models, Struct. Equ. Model. **8**, 430 (2001).

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

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

[89] J. Wang and Z. Hazari, Promoting high school students' physics identity through explicit and implicit recognition, Phys. Rev. Phys. Educ. Res. **14**, 020111 (2018).

[90] 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).

[91] T. E. Sutherland and C. C. Bonwell, *Using Active Learning in College Classes: A Range of Options for Faculty* (Jossey-Bass, New York, 1996).

[92] W. E. Cashin, Improving Lectures. Idea Paper No. 14 (1985).

[93] L. C. McDermott, Research on conceptual understanding in mechanics, Phys. Today **37**, 24 (1984).

[94] R. K. Thornton and D. R. Sokoloff, Learning motion concepts using real-time microcomputer-based laboratory tools, Am. J. Phys. **58**, 858 (1990).

[95] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, Sci. Educ. **66**, 211 (1982).

[96] A. Maries, N. Karim, and C. Singh, Active learning in an inequitable learning environment can increase the gender performance gap: The negative impact of stereotype threat, Phys. Teach. **58**, 430 (2020).

[97] K. Binning, N. Kaufmann, E. McGreevy, O. Fotuhi, S. Chen, E. Marshman, Z. Y. 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).

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

[99] 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).

[100] A. Miyake, L. E. Kost-Smith, N. D. Finkelstein, S. J. Pollock, G. L. Cohen, and T. A. Ito, Reducing the gender achievement gap in college science: A classroom study of values affirmation, Science **330**, 1234 (2010).