# Improving Inclusiveness of Learning Environment in Introductory Physics and Content Understanding in Quantum Physics


by

**Yangqiuting Li**

B.S., Nanjing University, 2017

M.S., University of Pittsburgh, 2021


Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy


University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

**Yangqiuting Li**

It was defended on

July 29, 2022

and approved by

Dr. Robert Devaty, Associate Professor, Department of Physics and Astronomy

Dr. Russell Clark, Senior Lecturer, Department of Physics Astronomy

Dr. Roger Mong, Associate Professor, Department of Physics and Astronomy

Dr. Joseph Grabowski, Associate Professor, Department of Chemistry

Thesis Advisor/Dissertation Director: Dr. Chandralekha Singh, Distinguished Professor, Department of Physics and Astronomy

**Improving Inclusiveness of Learning Environment in Introductory Physics and Content Understanding in Quantum Physics**

Yangqiuting Li, PhD

University of Pittsburgh, 2022

Students' motivational beliefs about physics can influence their engagement and performance in physics as well as retention in their majors and careers. Students from underrepresented groups in physics such as women may not have enough encouragement and role models to help them develop strong motivational beliefs in physics. The societal stereotypes and biases in physics may further undermine their motivational beliefs and lead to withdrawal from physics courses or careers. In addition to motivational beliefs, students' academic performance at the end of a physics course is also an important course outcome. Prior studies have shown that factors such as students' prior preparation, quality of teaching, and sociocultural factors can influence students' motivational beliefs and academic performance. However, very few studies have investigated the effect of students' perception of the inclusiveness of the learning environment on their motivational beliefs and academic performance.

In this dissertation I address the question of what role is played by students' perception of the inclusiveness of the learning environment and its relationship to gender, motivational beliefs, and academic performance in introductory physics courses. I first investigated the relationship between different motivational beliefs using both quantitative and qualitative methods. I also examined how students' physics motivational beliefs predict their identity in physics-related disciplines such as engineering. Then I analyzed gender differences in students' motivational beliefs and academic performance at the end of introductory physics courses and how different

components of perception of the inclusiveness of the learning environment predict these course outcomes. These findings suggest that instructors play an important role in developing a more equitable and inclusive learning environment, in which all students can thrive. Lastly, I discuss my work on the development, validation and in-class implementation of a multiple-choice questions sequence (MQS), which was designed to help advanced undergraduate students improve their understanding of quantum measurement.

**Table of Contents**

**List of Tables**

**List of Figures**

## Preface

Firstly, I would like to express my sincere gratitude to my advisor Dr. Chandralekha Singh for her guidance and continuous support and encouragement during my doctoral studies. I would also like to thank my other committee members: Dr. Robert Devaty, Dr. Russell Clark, Dr. Roger Mong, and Dr. Joseph Grabowski for their insightful comments and advice.

Next, I would like to thank Dr. Yasemin Kalender and Dr. Kyle Whitcomb for their mentorship and kind help while I was first becoming acquainted with physics education research. I also want to thank all my research group members for being such good colleagues and friends.

Very special and hearty thanking goes to my friends: Muqing Yu, Sonja Cwik, Yasha Kaushal, Ly Malespina, Biprateep Dey, Elizabeth Burnette, Sam Pittman, and Emma Li. I am grateful for their continuous support and company. They made my life colorful and wonderful, and they made me feel at home in Pittsburgh.

Last but not the least, I would like to thank my parents for giving birth to me, for their unconditional love, and for supporting me spiritually throughout my life.

# 1.0 Introduction

Prior studies have shown that women are often underrepresented in many science, technology, engineering, and mathematics (STEM) courses and disciplines [1-12]. For example, even though women earn approximately 60% of all bachelor's degrees in the US, only 20% of the physics undergraduate degrees are earned by women [13]. In addition, several studies have reported a gender disparity in students' performance in physics [14,15]. Moreover, prior studies showed that female students leave physics at higher rates than male students [16,17]. These studies suggest that we are largely missing out on the talents of half of the population, which not only hinders the development of physics because of the loss of talent and diversity, but also hinders women from pursuing many great career opportunities. Therefore, efforts to promote participation, achievement, and continuation of women in physics are important for the development of both individuals and the society as a whole. Some prior studies suggest that individuals' performance in physics can be influenced by their motivational beliefs such as self-efficacy, interest and identity in that domain [18-23]. Students from underrepresented groups in physics such as women may not have enough encouragement and role models to help them develop strong motivational beliefs in physics. In addition, the societal stereotypes and biases in physics may further undermine their motivational beliefs and lead to withdrawal from physics courses or careers [11,24-32]. Thus, investigation of students' motivational beliefs is important for better understanding the underrepresentation of women and minority students in physics and can be useful for formulating guidelines for developing an inclusive learning environment and promoting diversity and equity in physics.

By inclusive learning environment, we refer to an environment in which all students feel welcome, valued, and supported. By equity in learning, we mean that not only should all students have adequate opportunities and access to resources, and have an inclusive learning environment with appropriate support and mentoring so that they can engage in learning in a meaningful and enjoyable manner, but the course outcomes should be equitable. Therefore, inclusiveness is necessary but not sufficient for equity since inclusiveness does not guarantee equitable course outcomes. By equitable course outcomes, we mean that students from all demographic groups (e.g., regardless of their gender identity or race/ethnicity) who have the prerequisites to enroll in the course, on average, have comparable outcomes, which is consistent with Rodriguez et al.'s equity of parity model [33]. The STEM course outcomes include student performance and their STEM motivational beliefs at the end of the courses because regardless of the performance, the motivational beliefs can influence students' short and long-term retention in STEM disciplines [34,35]. We note that adequate opportunity and access to resources, inclusive learning environment and equitable outcomes are strongly entangled with each other. For example, if the learning environment is not inclusive, the outcomes are unlikely to be equitable. In this study, we aim to understand how students' perception of the inclusiveness of the learning environment predicts their course outcomes including both academic performance and motivational beliefs.

### 1.1 Motivational Beliefs

Prior research suggests that students' self-efficacy is an important motivational belief for them to excel in a domain [20,36-38]. Self-efficacy is one's belief in one's ability to succeed in a specific area or accomplish a task [39,40]. Studies have shown than students' engagement and performance can be influenced by their self-efficacy [41-44]. For example, students who have high self-efficacy tend to see difficulties as challenges and believe that productive struggle can help them improve, so they often choose to take harder courses and ask to do more challenging problems than students with low self-efficacy, who usually see difficulties as threats and obstacles to success [45].

Another motivational characteristic is interest, which refers to students' curiosity, enjoyment and engagement in a specific area [46,47]. Studies have shown that interest can also influence students' learning [41,47-51]. For example, one study showed that students' performance can be improved by connecting physics courses to students' daily lives or using evidence-based curricula to make the courses more interesting [52].

According to Eccles's Expectancy-Value Theory (EVT) [53,54], interest and self-efficacy are correlated with each other and together influence students' learning outcomes and career choices. In the EVT, expectancy refers to students' belief in their ability to succeed in a given task [54], which is closely related to self-efficacy. Value refers to the subjective task value for students, which can be differentiated into four components: intrinsic value, attainment value, utility value, and cost [54]. Intrinsic value refers to students' interest in the task and the enjoyment they experience from doing the task. Attainment value reflects how important students themselves feel it is for them to develop mastery and do a good job in the field [54]. Utility value pertains to students' perception of whether the task can help them achieve some other goals [54]. The last value component is cost, which refers to the assessments of how much effort and time will be taken to engage in the task as well as the amount of opportunity cost and stress caused by the task [54]. In the EVT, students' learning goals, academic engagement and performance, and persistence in a field are impacted by their expectancy of success and the four components of value [54].

In addition, students' identity in a specific field such as physics is another important motivational belief that influences their career decisions and outcome expectations [21,22,55-62]. Students' physics identity is related to whether they see themselves as a physics person [21,22,55,58,63]. Some studies have found that female students often report lower physics identity than male students [63-65]. This gender difference in physics identity has been shown to be related to societal biases and stereotypes about who belongs in and can succeed in physics [66-68]. In physics, these stereotypes can negatively influence women's experiences, which may lower their sense of belonging and identity, and lead to withdrawal from physics [24,69,70]. Therefore, investigating students' physics identity may help us understand the gender difference in participation in physics.

## 1.2 Overview

In chapters 2 and 3, we focus on the physics identity framework in which introductory students' physics identity is predicted by their perceived recognition, self-efficacy and interest. We investigate the relationships among these motivational beliefs by using structural equation modeling (SEM) with many statistically equivalent models and discuss how our theoretical framework guides us to select good models from amongst the equivalent SEM models. Our theoretical framework suggests that in addition to statistical information about the model fit, researchers should use other evidence such as interview data and findings from prior studies as well as instructional implications of each equivalent model to select good models. According to prior interviews with students, we find that the perception of not feeling positively recognized by the instructors/TAs as a person who can excel in physics has the potential to deteriorate students' physics self-efficacy and interest. In addition, SEM models in which recognition predicts self-efficacy and interest have the potential to empower instructors and help them understand their role in creating an inclusive and equitable learning environment in which students are recognized and affirmed, compared to other equivalent models which may mask the important role instructors play in supporting their students. Thus, even though all the equivalent models are equally good from a statistical point of view, it is better to select the models with perceived recognition predicting self-efficacy and interest, consistent with our theoretical framework.

In chapter 4, we analyze data from individual interviews with 38 female students to investigate their learning experiences in physics courses in order to obtain a qualitative understanding of the factors that shape their beliefs. We find that female students' negative and positive perceived recognition from instructors and TAs greatly influenced their self-efficacy and interest and even impacted their desire to persist in STEM majors. We categorize different types of perceived recognition that women reported in our interviews and how they influenced them. These findings can help physics educators reflect on their interactions with students in order to contemplate ways to provide positive recognition and validation to their students. Our research suggests that it is important for instructors and TAs to realize that they have responsibility to intentionally develop an inclusive and equitable learning environment in which all students feel appropriately recognized and feel safe to express themselves and learn from each other.

In chapter 5, we investigate 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 non-inclusive 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 strongest 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.

In chapter 6, we discuss a study in a first year college introductory physics course that shows that women, on average, feel less recognized by their physics instructors than men as students who can excel in physics. We also discuss how this lack of perceived positive recognition pertaining to physics can adversely affect their self-efficacy and academic performance in the course. We recommend that physics instructors not be parsimonious in their praise of students and make a conscious effort to positively recognize their students for their effort and progress whenever an opportunity arises. Our study also suggests that instructors should be careful not to give unintended messages to students, e.g., by praising some students for brilliance or intelligence as opposed to their effort because praising a student for brilliance can convey to other students that they do not have what is required to excel in physics.

In chapter 7, we investigate how students' sense of belonging predicts their academic performance including both conceptual understanding and quantitative problem solving skills in a college introductory physics course, controlling for high school performance measured by high school grade point average (GPA) and Scholastic Aptitude Test (SAT) math scores. We found signatures of inequitable and non-inclusive learning environment; e.g., compared to male students, female students' sense of belonging was lower on average. Using structural equation modeling (SEM), we found that students' sense of belonging plays an important role in predicting their academic performance even after controlling for high school performance. These results indicate that the gender difference in students' sense of belonging at least partially explains the finding that while female students had a similar average SAT math score and a higher average high school GPA compared with male students, they had a lower average score on the conceptual test (Force Concept Inventory) than male students. Our findings suggest that physics instructors must be intentional about creating an inclusive and equitable learning environment in which students from all demographic groups have a high sense of belonging and can excel.

In chapter 8, we adapt a prior identity framework to investigate how students' perception of the inclusiveness of the learning environment (including their sense of belonging, perceived effectiveness of peer interaction and perceived recognition) predicts students' physics self-efficacy, interest and identity by controlling for their self-efficacy and interest at the beginning of a calculus-based introductory physics course. We surveyed 1203 students, 35% of whom identified as women. We found signatures of inequitable and non-inclusive learning environment in that not only were female students' physics self-efficacy and interest lower than male students' at the beginning of the course, but the gender gaps in these motivational constructs became even larger by the end of the course. Analysis revealed that the decreases in students' physics self-efficacy and interest were mediated by the learning environment and predicted students' physics identity. We find that perceived recognition played a major role in predicting students' physics identity, and students' sense of belonging in physics played an important role in explaining the change in students' physics self-efficacy.

In chapter 9, we focus on students' motivational beliefs in a two-term college calculus-based introductory physics sequence (physics 1 and physics 2) at a large public university in the US. We investigated how students' perception of the inclusiveness of the learning environment (including their sense of belonging, perceived effectiveness of peer interaction and perceived recognition) predicts their physics course grades and physics motivational beliefs including self-efficacy, interest and physics identity. Our results show that students' perception of the inclusiveness of the learning environment statistically significantly predicts their physics grades and motivational beliefs. Students' perceived recognition (e.g., by instructors) was an important predictor of their physics identity, and their sense of belonging was central to their self-efficacy. In addition, we found possible signatures of non-inclusive learning environment in that female students' mean scores in sense of belonging, perceived effectiveness of peer interaction and perceived recognition were all lower than male students' in both courses, and the gender gap in perceived recognition increased from physics 1 to physics 2. Even though there was no gender difference in students' SAT math scores and female students even had a higher average high school GPA than male students, female students' average grades were significantly lower than male students in both physics 1 and physics 2. Using structural equation modeling, we find that the gender differences in students' physics self-efficacy, interest, identity and grades were partially mediated by the different components of students' perception of the inclusiveness of the learning environment.

In chapters 10, we investigate how students' perception of the inclusiveness of the learning environment, including sense of belonging, perceived peer interaction and perceived recognition (e.g., by instructors), predicts students' scores on Force Concept Inventory (FCI) and physics motivational beliefs including self-efficacy, interest and identity at the end of the course after controlling for students' high school performance including SAT math scores and high school GPA as well as their FCI scores and motivational beliefs at the beginning of the course. We find that female students' mean scores in physics motivational beliefs and perception of the inclusiveness of the learning environment were all lower than male students', and the gender gap in students' self-efficacy increased from the beginning to the end of the course. Using structural equation modeling, we find that the gender differences in students' motivational beliefs and FCI scores were mediated by the different components of students' perception of the inclusiveness of the learning environment. In particular, students' perceived recognition, e.g., by instructors, was an important predictor of their physics identity, and their sense of belonging predicted their self-efficacy. Our findings can be valuable for contemplating guidelines for creating an inclusive and equitable learning environment in which all students can excel.

Physics courses are important for engineering students because not only are they the foundation for many engineering courses, but students' physics motivational beliefs such as self-efficacy and identity may also influence their engineering identity as well as their choice of careers. Therefore, in chapters 11, we investigate first-year undergraduate engineering students' engineering identity and how it is predicted by their physics motivational beliefs (including physics self-efficacy, interest, perceived recognition and identity) in a calculus-based introductory physics course at a large research university in the US. We first investigated how these motivational beliefs change from the beginning to the end of the course (i.e., from pre to post) using descriptive statistics. Then, we investigated the predictive relationships among these motivational constructs using structural equation modeling (SEM). The SEM analysis revealed that students' engineering identity is predicted by their physics self-efficacy and identity. However, the descriptive statistics results showed that both male and female students' physics self-efficacy and identity decreased from pre to post, and female students' physics self-efficacy dropped even more than male students' did. Although students' average score on engineering identity also decreased from pre to post, this change was only statistically significant for male students. Our results show that students' physics perceived recognition is the strongest predictor of physics identity, and it also predicts students' engineering identity through physics identity and self-efficacy. We note that even though there were significant gender differences disadvantaging women in all motivational constructs studied, gender does not directly predict engineering and physics identities, which means that the gender differences in both identities are mediated through physics self-efficacy, interest and perceived recognition. Thus, in order to boost students' engineering identity, it is important to create an equitable and inclusive environment for learning physics, in which all students feel recognized and supported appropriately and develop a stronger physics and engineering identity.

In chapter 12, we discuss the development, validation, and implementation of a multiple-choice questions sequence (MQS) on the topic of quantum measurement. This MQS was developed using students' common difficulties with quantum measurements as a guide and was implemented in a junior-/senior-level quantum mechanics course at a large research university in the US. We compare student performance on assessment tasks focusing on quantum measurement before and after the implementation of the MQS and discuss how different difficulties were reduced and how to further improve students' conceptual understanding of quantum measurement.

Lastly, in the final chapter, we discuss future directions for this work.

# 2.0 How to Select Suitable Models from Many Statistically Equivalent Models: An Example from Physics Identity

## 2.1 Introduction

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 [1,2,4,6,7,13,20-22,36,37,63,71,72]. 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 [20,37,41,44,45,48,63,73]. For students from underrepresented groups, these motivational beliefs might be undermined, e.g., by negative societal stereotypes and biases about who belongs in STEM 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 [24,28,74-77]. Hence, investigating students' motivational characteristics is critical to understanding and addressing diversity, equity, and inclusion issues in STEM disciplines.

In explaining participation in STEM careers, identity has been argued to be a particularly important motivational construct [21,22,58,63]. Students' identity in an academic domain, such as physics, is students' views about whether they see themselves as a "physics person" [21,22]. 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") [22]. Hazari et al. modified the framework by adapting it to physics specifically and focused on students' beliefs about their competence and performance rather than how students can practice and exhibit them in class [78,79]. In addition, a new factor "interest" was added [78,79]. Moreover, they found that beliefs about performance and competence are actually not distinct and predict students' physics identity as a single construct [21,57]. Kalender et al. adapted Hazari et al.'s physics identity framework and used self-efficacy in place of performance/competence beliefs in their model [80].

Self-efficacy, which is very closely tied to performance/competence beliefs, is defined as students' beliefs in their capability to succeed in a certain situation, task, or particular domain [39,81,82]. It has been shown to influence students' engagement and performance in a given domain [41,44]. Students with high self-efficacy in a domain often enroll in more difficult courses in that domain than those with low self-efficacy because they perceive difficult tasks as challenges rather than threats [45]. In addition, studies suggest that students' self-efficacy predicts their career choices and persistence toward their short- and long-term career goals [41,73].

Another motivational characteristic is interest, which is defined by positive emotions accompanied by curiosity and engagement in a particular content [47]. Interest has also been shown to influence students' learning [41,47,48]. 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 [52]. In addition, studies have shown that students' interest is not independent of self-efficacy [41,54]. According to Eccles's Expectancy-Value Theory (EVT) [53,54], interest is paired well with self-efficacy as connected constructs that predict students' academic outcome expectations and career aspirations.

The identity framework discussed earlier reveals that individuals' internal identity in a domain is not only impacted by their self-efficacy and interest but also 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 [79]. Many studies have shown that female students did not feel that they were recognized appropriately in science even before they entered college [66,83]. For example, a report of the National Science Foundation [84] 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 [84]. Studies show that these stereotypes and biases also exist in the university context [81,85]. 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 [85]. These experiences of not being recognized as a science person may increase stereotype threat for female students, and all of these gender-based biases may accumulate over time and become a detriment to female students' science identity.

Some quantitative studies focusing on the relationship between students' science identity and other motivational beliefs show that perceived recognition is the strongest predictor of science identity as compared to interest and self-efficacy [58,81]. 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, which are critically important for engineering career choice [58]. In addition, Kalender et al.'s physics identity framework shows that college students' perceived recognition not only predicts their physics identity directly, but also indirectly predicts their physics identity by predicting their self-efficacy and interest in physics [81]. All of these quantitative studies also show that even though there are gender differences in students' identity, gender does not directly predict physics identity.

Moreover, studies have shown that the motivational characteristics that comprise physics identity (self-efficacy, interest, and perceived recognition) correlate to and interact with each other in meaningful ways [39,47,86]. Thus, one may wonder which of these constructs generally act as predictors for others. For example, does interest primarily drive self-efficacy or does self-efficacy primarily drive interest? It is difficult to answer this question from a statistical point of view because the multivariate statistical techniques such as structural equation modeling (SEM) [87] only examine the correlations among variables but cannot establish causal effects [88]. The model with interest predicting self-efficacy is actually statistically equivalent to the model with self-efficacy predicting interest, and they have identical model fit indices. The fact that there are many statistically equivalent models tells us that using SEM, we never test one statistical model but rather a whole class of equivalent models from which the hypothesized model cannot be distinguished by statistical means alone [89]. Thus, no matter how well a hypothesized model fits the data in comparison with competing models, no possibility exists of unequivocally confirming the model if there are models equivalent to it without using other sources of data [89]. In particular, model equivalence is methodologically significant because it shows that other input is required to prefer one statistically equivalent SEM model over another [89]. In this study, we propose a theoretical framework which could be helpful for researchers to select good models from several statistically equivalent ones. In particular, we focus on the statistically equivalent models showing different mediation mechanisms between gender, physics self-efficacy, interest, and perceived recognition predicting physics identity. We argue that even though these statistically equivalent models have the same model fit indices, our theoretical framework guides the selection of good models based upon interview data and their potential to empower the instructors to make positive changes.

## 2.2 Theoretical Framework

Prior studies of physics identity suggest several possible relationships among the motivational characteristics that comprise physics identity (i.e., self-efficacy, interest, and perceived recognition) [58,81,90]. However, researchers have not explicitly acknowledged the fact that there are several statistically equivalent models to the model they proposed. In a recent paper, physics education researchers have been asked to explain why they select a specific model out of several statistically equivalent ones while using SEM or multiple-regression as the analysis tool. In other words, researchers must explain why the proposed model is preferable to the other equivalent ones.

Here, we propose a theoretical framework, which can be helpful for researchers to select good models from several statistically equivalent ones. Our framework includes three criteria. First, a good model should be supported by robust quantitative analysis with statistical tools. For example, researchers can use several fit parameters to examine the consistency between their models and data, the model fit is good if the fit parameters satisfy specific threshold criteria. However, a model that fits the data well is not necessarily a good model. Another important criterion for selecting a good model from several statistically equivalent ones is that it should be supported by additional evidence. This evidence may include but is not limited to researchers' own interview data or findings from prior studies. Furthermore, in addition to quantitative and qualitative support, a good model should also have positive influence on pedagogical approaches.

One primary goal of physics education research is to empower instructors and inspire them to adopt and adapt evidence-based pedagogical approaches to help students learn physics effectively in an equitable and inclusive learning environment. Thus, researchers should consider the potential instructional implications of different statistically equivalent models and select the models that have high instructional benefits.

In this paper, we take physics identity as an example to show how the proposed theoretical framework is used to select good models from many statistically equivalent ones for how self-efficacy, interest, and perceived recognition mediate physics identity. In particular, inspired by prior studies and the framework of physics identity [21,58,81], here we consider 27 possible statistically equivalent models for how the relationship between gender and physics identity is mediated by the three motivational constructs (self-efficacy, interest, and perceived recognition) in college students' first calculus-based introductory physics course. We discuss affordances of different statistically equivalent models with different predictive relations between the three mediating constructs in order to elucidate how our theoretical framework guided us to select a few good models and highlight instructional implications of some of the models. For example, as shown in Figure 1, we first considered a model that only has covariance among the three motivational constructs that mediate the relation between gender and physics identity (note that although there can be a direct path from gender to identity, we never find that path to be statistically significant in any of the 27 statistically equivalent models so we have omitted it for clarity). Then, we considered other statistically equivalent models to the one shown in Figure 1 by making one or two of the mediating constructs as predictors of the other mediating constructs (discussed later in greater detail). We argue that even though the 27 SEM models are equivalent, our interviews with students suggest that some of these models are better than others and they also have the potential to inspire instructors to make positive transformations in their courses. Therefore, we prefer some SEM models over other equivalent ones based on our theoretical framework guided by interviews and instructional value driven considerations. Even though, in this study, we consider all of 27 possible statistically equivalent models, we are not asking researchers to list all possible statistically equivalent models for the situation they have, since it is unrealistic especially when the proposed model is complex. However, our theoretical framework posits that researchers should generate at least a few substantively meaningful different equivalent versions and deliberate based upon other data and instructional implications why the proposed model is better than the others [91].

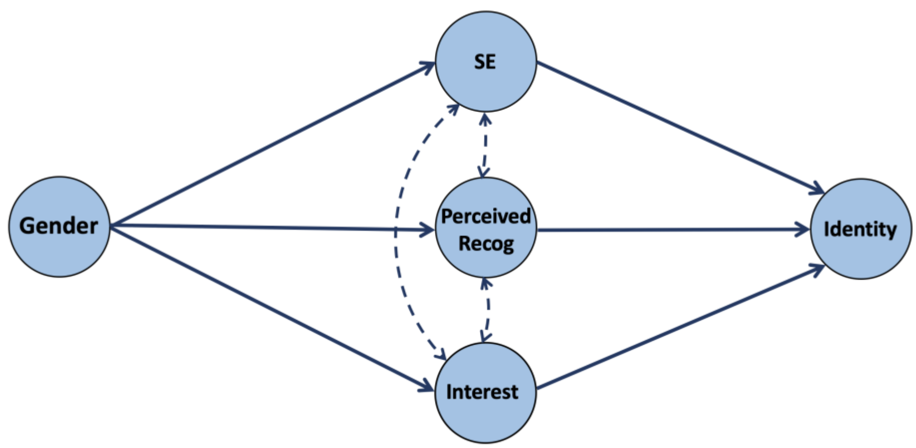



**Figure 1 Schematic representation of the path analysis that shows how the relationship between gender and physics identity is mediated by perceived recognition (Recog), self-efficacy (SE) and interest in one of the 27 statistically equivalent models. The covariances between the three mediating constructs are shown with dashed lines. The direct path from gender to identity is not shown (it is not statistically significant in our model).**

## 2.3 Research Questions

Our research questions pertaining to physics identity in a quantitative study involving structural equation modeling (SEM) [87] at the end of a calculus-based introductory physics course taken by engineering, physical science, and mathematics majors in the first semester of their first year at a large state-related university are as follows (please note that RQ1 focuses on descriptive statistics, which is necessary to draw meaningful inferences from the predictive statistics in RQ2):

**RQ1.** What are the gender differences in a set of physics motivational beliefs, i.e., self-efficacy, interest, perceived recognition, and identity, for students at the end of the course?

**RQ2.** How can we narrow down and select good models from a large number of statistically equivalent SEM models (in which the relation between gender and physics identity is mediated via perceived recognition, self-efficacy and interest) based upon our theoretical framework using other evidence and potential instructional implications of the models?

## 2.4 Methodology

### 2.4.1 Participants

In this study, we collected motivational survey data at the end of the semester from students who took the introductory calculus-based physics 1 course in two consecutive fall semesters. This course is taken mostly by students majoring in engineering, physical sciences, and mathematics. The paper surveys were handed out and collected by TAs in the last recitation class of a semester. Finally, we combined the two semesters' data. The demographic data of students—such as gender—were provided by the university. Students' names and IDs were de-identified by an honest broker who provided each student with a unique new ID (which connected students' survey responses with their demographic information). Thus, researchers could analyze students' data without having access to students' identifying information.

There were 1219 students participating in this survey including both semesters. In our final data analysis, we kept 1203 students (including 427 female students and 776 male students) because the other 16 students did not provide their gender information. We recognize that gender is a social construct and is not binary. However, because students' gender information was collected by the university, which offered binary options, we did the analysis with the binary gender data in this study. 1.3% of the students who did not provide this information were not included in this analysis.

### 2.4.2 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 1. The survey items were adapted from the existing motivational research [21,92-94] and have been re-validated in our prior work [37,95,96]. The validation and refinement of the survey involved use of one-on-one student interviews with both introductory and advanced students, exploratory and confirmatory factor analyses (EFA and CFA) [97], Pearson correlation between different constructs and Cronbach alpha (which is a measure of the internal consistency of each construct with several items) [98-100].

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 (Cronbach's alpha = 0.8) and these items had the response scale "NO!, no, yes, YES!". We also had four items for interest (Cronbach's alpha = 0.82). 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 [21]. Perceived recognition corresponds to whether a student thinks other people see them as a physics person [21,63,101], and it includes three items which correspond to family, friends and TA/instructor (Cronbach's alpha = 0.86). 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 [102].

**Table 1 Survey questions for each of the motivational scales, along with CFA factor loadings. Lambda represents the factor loading of each item, which is the correlation between the item and the construct with *p* < 0.001 indicating the correlation is highly statistically significant. The square of Lambda for each item gives the fraction of its variance explained by the construct. †The response options for this question are "Never, Once a month, Once a week, Every day". ‡The response options for this question are "very boring, boring, interesting, very interesting".**

| Construct and Item | Lambda | *p* value |
|---|---|---|
| Physics Identity | | |
| I see myself as a physics person. | 1.000 | <0.001 |
| Physics Self-Efficacy (Cronbach's alpha = 0.8) | | |
| I am able to help my classmates with physics in the laboratory or in recitation. | 0.706 | <0.001 |
| I understand concepts I have studied in physics. | 0.728 | <0.001 |
| If I study, I will do well on a physics test. | 0.738 | <0.001 |
| If I encounter a setback in a physics exam, I can overcome it. | 0.675 | <0.001 |
| Physics Interest (Cronbach's alpha = 0.82) | | |
| I wonder about how physics works † | 0.669 | <0.001 |
| In general, I find physics ‡ | 0.793 | <0.001 |
| I want to know everything I can about physics. | 0.795 | <0.001 |
| I am curious about recent physics discoveries. | 0.692 | <0.001 |
| Physics Perceived Recognition (Cronbach's alpha = 0.86) | | |
| My family sees me as a physics person. | 0.903 | <0.001 |
| My friends see me as a physics person. | 0.900 | <0.001 |
| My physics TA and/or instructor sees me as a physics person. | 0.691 | <0.001 |

### 2.4.3 Quantitative Analysis of Survey Data

First, we calculated the mean score for each construct for each student. Then, we used a *t*-test [103,104] to compare the mean scores by gender. In our previous study [81], we have checked the response option distances for our survey constructs by using item response theory (IRT) to support the use of means across ratings [105]. Here, 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 [106,107]. 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 differences between the location parameters were 1.3 and 1.4. For the second group — “NO!, no, yes, YES!” — the differences between the location parameters were 1.4 and 2.0. These results show that the numerical values for the location differences for item responses are comparable, which suggests that calculating the traditional mean score for items is reasonable [105,107]. 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 coefficient are > 0.98 for all constructs), which indicates that the use of mean scores is reasonable [105].

To validate the items on our survey, we performed the confirmatory factor analysis (CFA) for each construct. The model fit is good if the fit parameters are above threshold. 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 good fit

[98]. In our study, CFI = 0.973, TLI = 0.964, RMSEA = 0.060 and SRMR = 0.033, which represents a good fit. Thus, there is additional quantitative support for dividing the constructs as proposed. Besides, as shown in Table 1, all of the factor loadings are higher than 0.5, which is considered as acceptable, 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 structural equation modeling (SEM) to analyze the predictive relationships among students' self-efficacy, interest, perceived recognition, and identity using our survey data. [108].

Before performing the SEM, we calculated the Pearson correlation coefficients pairwise between the constructs. As shown in Table 2, 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 separately examined in the SEM [109]. It is noteworthy that the correlation coefficient between physics identity and perceived recognition is 0.84. This is consistent with Godwin et al. and Kalender et al.'s prior work finding that perceived recognition (external identity) is the largest predictor of physics identity (internal identity) [58,81].

**Table 2 Pearson correlation coefficients of the constructs in the mediation model.**

| Observed Variable | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Physics identity | -- | -- | -- | -- |
| 2. Self-efficacy | 0.68 | -- | -- | -- |
| 3. Interest | 0.70 | 0.63 | -- | -- |
| 4. Perceived Recognition | 0.84 | 0.70 | 0.68 | -- |

In this study, we used the R software package "lavaan" to perform structural equation modeling (SEM) [110]. The SEM includes two parts: CFA and path analysis. Apart from CFA, the path analysis part of SEM gives regression coefficients $\beta$ for paths between each pairs of constructs. The value of each $\beta$ is a measure of the strength of that relationship. Compared with a multiple regression model, the advantage of SEM is that we can estimate all the regression links for multiple outcomes and factor loadings for items through CFA simultaneously [87]. The level of the SEM model fit is also given by CFI, TLI, RMSEA and SRMR [87,98].

Before performing gender mediation analysis (which tests how the relation between gender and physics identity is mediated through self-efficacy, interest, and perceived recognition), we first tested the gender moderation relations between each pair of constructs (i.e., does the strength of relationship between any two constructs in the model differ for women and men?), which includes testing of factor loadings, indicator intercepts and residual variances. Then, we performed multi-group SEM analysis to investigate whether the regression pathways were different across gender. Results showed that in all of 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 [87].

To answer RQ2, we explicitly illustrate only four equivalent SEM models out of 27 possible equivalent models for brevity (see Supplemental Material [111] for results of some other statistically equivalent SEM models if one is interested in comparing some other SEM models). We first delineate the model fit indices (e.g., CFI, TLI, RMSEA and SRMR) of these statistically equivalent models to make sure these models are robust from a statistical perspective. A comparison of the coefficients of determination (R squared) of constructs in the different models is presented to estimate how much variance of each construct that mediates the relation between gender and physics identity is explained by each model (of course, the total amount of variance explained in the physics identity is the same across all 27 models). We argue that some of these statistically equivalent models in which physics self-efficacy, interest, and perceived recognition mediate the relation between gender and physics identity are better than others based upon our theoretical framework using evidence gathered from individual interviews and instructional implications.

## 2.5 Results

### 2.5.1 Gender Differences in Motivational Characteristics

Pertaining to RQ1, Table 3 shows that women have significantly lower average scores in all of the four motivational constructs, and the effect sizes given by Cohen's *d* are all in the medium range [104]. In particular, female students' average scores on identity and perceived recognition were below the neutral score of 2.5. Thus, many female students did not think others see them as a physics person, and they did not see themselves as a physics person either. Table 3 also shows that male students' average scores on all four constructs were above the neutral score.

**Table 3 Descriptive statistics for female and male students in which M stands for construct mean value, SD is the standard deviation and N is the number of students. Effect sizes and *p*-values are presented in the right most column with $p < 0.001$ indicating highly statistically significant gender differences. A minus sign indicates that male students have higher scores than female students.**

| Constructs | Female Students N = 427 | | Male Students N = 776 | | Statistics | |
|---|---|---|---|---|---|---|
| | M | SD | M | SD | Cohen's d | *p* value |
| Physics Identity | 2.17 | 0.83 | 2.63 | 0.83 | -0.56 | < 0.001 |
| Perceived Recognition | 2.24 | 0.72 | 2.60 | 0.73 | -0.49 | < 0.001 |
| Self-efficacy | 2.71 | 0.57 | 2.99 | 0.50 | -0.53 | < 0.001 |
| Interest | 2.72 | 0.64 | 3.10 | 0.58 | -0.64 | < 0.001 |

### 2.5.2 Using the Theoretical Framework to Select Good Models from Several Equivalent SEM Models

In order to understand why there are 27 statistically equivalent models in which the relation between gender and physics identity is mediated by self-efficacy, interest, and perceived recognition, we consider the associations among these three mediating constructs. There are three possible associations between each pair. These associations are covariance, direct effect via regression from one to the other, or direct effect via regression in the reverse direction. For example, there can be a direct regression path from self-efficacy to interest or from interest to self-efficacy, or there may only be a covariance between self-efficacy and interest. Similarly, there are three possible types of associations between self-efficacy and perceived recognition, and between interest and perceived recognition. Thus, with the constraints that no regression arrows point to gender and arrows can only point to physics identity since it is the outcome variable, there are $3 \times 3 \times 3 = 27$ statistically equivalent SEM models in total. As discussed earlier, the model fit indices are the same for all 27 models and suggest a good fit to the data CFI = 0.972 (>0.90), TLI = 0.962 (>0.90), RMSEA = 0.056 (<0.08) and SRMR = 0.032 (<0.08). In addition, we also calculated the coefficient of determination $R^2$ for each construct in each statistically equivalent SEM model, which represents the proportion of each construct's variance explained by each model. The results show that $R^2$ of identity is 0.74 in all of the models, which means that each equivalent model explains 74% of the variance in physics identity, i.e., these models explain students' physics identity equally well (see Appendix A for results of $R^2$ values for other constructs). Thus, these statistically equivalent SEM models are robust from a statistical point of view, i.e., they all satisfy the first criterion of our framework (good statistical model fit).

Next, we focus on the next two criteria for selecting good models from many statistically equivalent ones. In particular, we explicitly discuss 4 of these 27 equivalent models to illustrate how to use the second and third criteria of our framework to select good models (see Supplemental Material [111] for results of some other statistically equivalent SEM models). In particular, we consider the data collected from our prior hour-long individual interviews with 52 student volunteers (42 women and 10 men) in the introductory physics courses using a semi-structured protocol [71,76,112] and the affordance of each model for its instructional implications.

First, we consider a model in which there are only covariances between each pair of perceived recognition, self-efficacy and interest, i.e., this model does not make assumptions about predictive relationships between these three mediating constructs. The results of the path analysis of the SEM for this model are presented visually in Figure 2. The numbers on the lines represent the standardized regression coefficients $\beta$ [87] between variables in which the standardization ensures that different $\beta$ in the model can be directly compared with each other. Figure 2 shows that there is a statistically significant regression line from gender to each of the three mediating constructs (self-efficacy, perceived recognition, and interest), consistent with Table 3 showing that there are significant gender differences in all of these three motivational constructs. However, the direct effect of gender on physics identity is statistically insignificant ($\beta = 0.02$, $p = 0.36$) even though female students' identity is significantly lower than that of male students, as shown in Table 3. 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, Figure 2 shows that even though there is a strong covariance among self-efficacy, interest, and perceived recognition, students' perceived recognition is the strongest predictor ($\beta = 0.59$) of their physics identity. This result is consistent with Hazari et al.'s and Kalender et al.'s prior work [21,81].

Model 1 in Figure 2 shows that there are positive correlations between students' physics self-efficacy, perceived recognition, and interest, which is consistent with our interview data showing that positive encouragement and recognition are likely to help students develop self-efficacy and interest in physics [71,76,112]. For example, some students reported that they feel better when the instructors know what students tend to struggle with and acknowledge that it's okay to not completely understand the content and they just need more practice, and eventually they will get over the struggle with that topic and move on to the next topic. In addition, the interviewed students also mentioned that they felt really encouraged when their physics problem solving, posters and talks were recognized by their instructors. On the other hand, students' self-efficacy and interest can deteriorate due to lack of appropriate recognition from instructors/TAs. For example, some students reported that when they went to the course instructor or TA to ask for help on physics problems, sometimes they were explicitly told that the problems were "easy", "obvious" or "trivial", which they perceived as disparaging or belittling in that they felt they were being told that they are not smart enough to do physics if they could not do such easy problems on their own. Thus, our interview data not only show that students' positive or negative perceived recognition is correlated with their self-efficacy and interest, but they also indicate that perceived recognition can affect students' physics self-efficacy and interest in the course and in physics overall. However, the important role played by perceived recognition on self-efficacy and interest is not reflected in Model 1, in which there are only covariances among self-efficacy, perceived recognition, and interest.

Furthermore, with regard to the instructional implications, the path analysis in Model 1 indicates that the gender difference in physics identity is mediated through self-efficacy, perceived recognition and interest. However, in Model 1, gender is the only predictor of self-efficacy and interest, which may be interpreted to imply that female students have lower physics identity because they have lower self-efficacy and interest in physics. Considering self-efficacy and interest may be perceived as fixed by many people, Model 1 does not provide suggestion with regard to instructors having the potential to improve students' self-efficacy and interest. Thus, this model in Figure 2 may not be the most instructionally beneficial model.

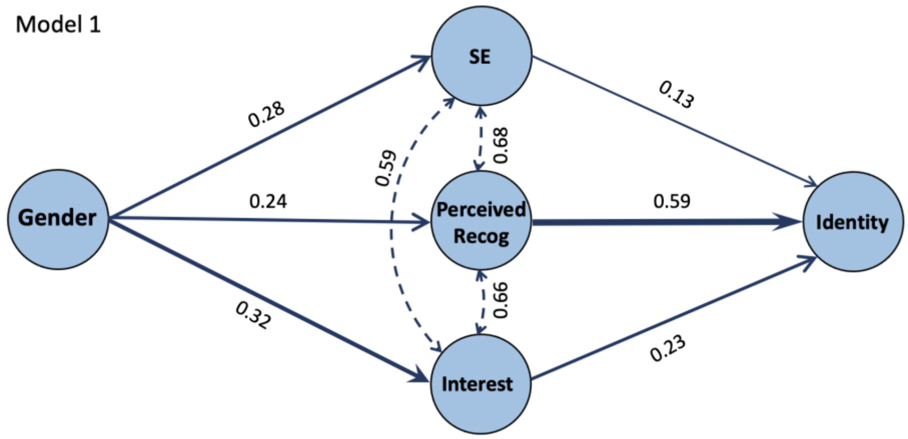



**Figure 2 Results of the path analysis part of SEM Model 1, in which there are only covariances between each pair of constructs: self-efficacy (SE), perceived recognition (Recog), and interest. Each regression line thickness qualitatively corresponds to the magnitude of $\beta$ values (standardized regression coefficient). All $\beta$ values shown are significant with $p < 0.001$. For clarity, we have removed the statistically insignificant regression path from gender to identity.**

Next, we consider two models in which self-efficacy and interest each act as predictors of the other two mediating constructs (see Figure 3). In particular, in Model 2, self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition. In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition. As shown in Figure 3, in Model 2, self-efficacy is only predicted by gender and the regression coefficient is 0.28, and in Model 3, interest is only predicted by gender with regression coefficient $\beta = 0.32$. These regression coefficients are the same as those in Model 1. However, in Model 2, the regression coefficients from gender to perceived recognition and interest are smaller or statistically insignificant compared to those in Model 1. This is because in Model 2, perceived recognition and interest are predicted not only by gender but also by self-efficacy, and thus there is more correlated effect being controlled for when estimating the regression coefficients from the gender to perceived recognition and interest. Similarly, in Model 3, the regression coefficients from gender to perceived recognition and self-efficacy are smaller or statistically insignificant compared to those in Model 1.

As noted, Models 2 and 3 in Figure 3 are statistically equivalent to Model 1 in Figure 2. However, these two models suggest that students' perceived recognition is predicted by their self-efficacy and interest, which is not consistent with our interview findings showing that perceived recognition may be an important driver of the other two constructs. Moreover, our interview also show that women were less likely than men to feel positively recognized by physics instructors/TAs, and this lack of recognition or discouraging feedback from instructors/TAs deteriorated their self-efficacy, and lowered perceived recognition and self-efficacy further lowered their interest [71,76]. For example, some interviewed women noted that their instructors/TAs sometimes showed more interest in male students' questions and answered male students' questions with more attention than when they answered their questions. This experience made the female students wonder whether it was because their questions were not good or too easy, and thus they started doubting their ability to do well in this course. Thus, Models 2 and 3 showing that students' perceived recognition is predicted by their interest and self-efficacy are not consistent with our interview data.

Furthermore, Models 2 and 3 also do not satisfy the third criterion of our theoretical framework for selecting a good model. In particular, Model 2 with self-efficacy in the first place may be interpreted to imply that students' self-efficacy about physics predicts their interest in physics and their views about whether other people see them as a physics person. Similarly, Model 3 with interest in the first place emphasizes that students' interest predicts their self-efficacy and their perceived recognition. Moreover, in Models 2 and 3, gender is the only predictor of interest or self-efficacy which together with the data in Table 3 (which shows gender differences in all of these motivational constructs) can be interpreted as deficit models. In particular, these equivalent models can be interpreted to imply, e.g., that women are not feeling positively recognized by their instructors and teaching assistants (TAs) as much as men because they have lower interest in physics and lower self-efficacy than men. Thus, Models 2 and 3 can be misinterpreted and instead of empowering instructors to recognize their students whenever an opportunity arises, they may lead to instructors becoming complacent about their own role in helping students develop identity as someone who can excel in physics. Thus, according to our interview data and the possible instructional implications of Models 2 and 3, these two models are not aligned with our framework.

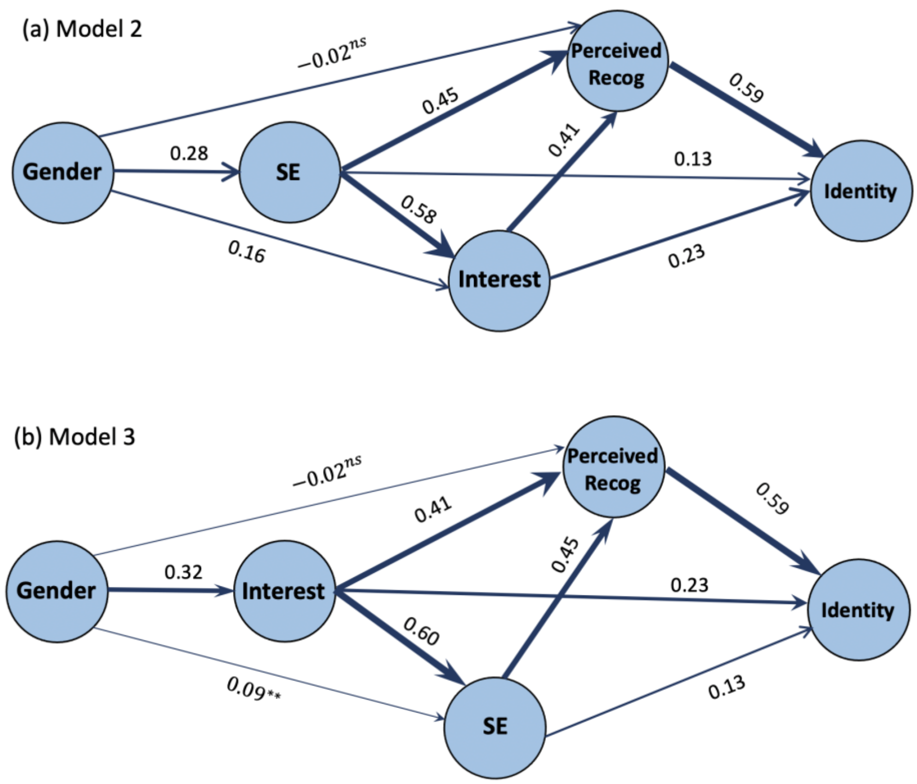



**Figure 3 Results of the path analysis part of the SEM models that show how the relationship between gender and physics identity is mediated through self-efficacy (SE), interest, and perceived recognition (Recog). (a) In Model 2, self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition. (b) In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition. Each regression line thickness qualitatively corresponds to the magnitude of $\beta$ with $0.001 \leqslant p < 0.1$ indicated by ** and $p > 0.05$ indicated by ns. All the other regression lines show relations with $p < 0.001$. For clarity, we have removed the statistically insignificant regression path from gender to identity.**

Finally, we consider a model in which perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest (see Figure 4). In Model 4, perceived recognition is only predicted by gender with regression coefficient $\beta = 0.24$, which is the same as that in Model 1, while the regression coefficients from gender to self-efficacy and interest are smaller than those in Model 1. Similar to Models 1-3, there is no statistically significant direct regression line from gender to physics identity in Model 4 ($\beta = 0.02$, $p = 0.36$). In addition, the regression coefficient from self-efficacy to identity is 0.13, from interest to identity is 0.23, and from perceived recognition to identity is 0.59, which are the same as those in Models 1-3.

As noted, according to our interview data, perceived recognition plays an important role in the development of students' self-efficacy and interest. Also, our interviews show that female students feel less recognized by their instructors/TAs than their male peers, which indicates that the current learning environment is not inclusive and equitable. For example, some interviewed women reported that men in their physics courses were generally praised more by the instructors/TAs than women, and sometimes instructors/TAs called men who answered the questions "brilliant", which made them feel as though they were not brilliant. In the interviews, some women reported that because of the negative experiences in their physics courses, they started questioning "Why am I here in the first place? Am I really interested in this?", and some had contemplated switching out of their major (either engineering or physics) while men never expressed similar concerns [76]. Thus, Model 4 with perceived recognition predicting self-efficacy and interest model is more theoretically grounded than the other three models based on our interview data.

Moreover, Model 4 with perceived recognition as the predictor of interest and self-efficacy is also likely to empower the instructors and give them the message that students' interest and self-efficacy in physics can be influenced by the recognition they receive from instructors. In addition, in Model 4, gender is the direct predictor of perceived recognition, which is consistent with Table 3 showing that women feel significantly less recognized as a physics person. Thus, Model 4 is more likely to inspire instructors to recognize their role in creating more inclusive and equitable learning environments in which all students including those from the underrepresented groups in physics such as women feel more positively recognized and affirmed. Moreover, Model 4 may be even more instructionally beneficial than other models, e.g., in which gender predicts perceived recognition but interest predicts self-efficacy unlike Model 4 because interest may be perceived as fixed by some instructors. We note that although Model 1 does not mask instructors' role in supporting their students by recognizing them, it may not empower instructors as much as Model 4. Thus, while physics self-efficacy, interest and perceived recognition are correlated constructs and predict each other differently in the statistically equivalent SEM models, by emphasizing different predictor-outcome relationships between these motivational constructs and between gender and these constructs, these statistically equivalent models can be interpreted very differently by physics instructors and therefore have different instructional implications. Thus, Model 4 with perceived recognition predicting self-efficacy and interest is theoretically more grounded consistent with our framework and more instructionally beneficial than the other models discussed.

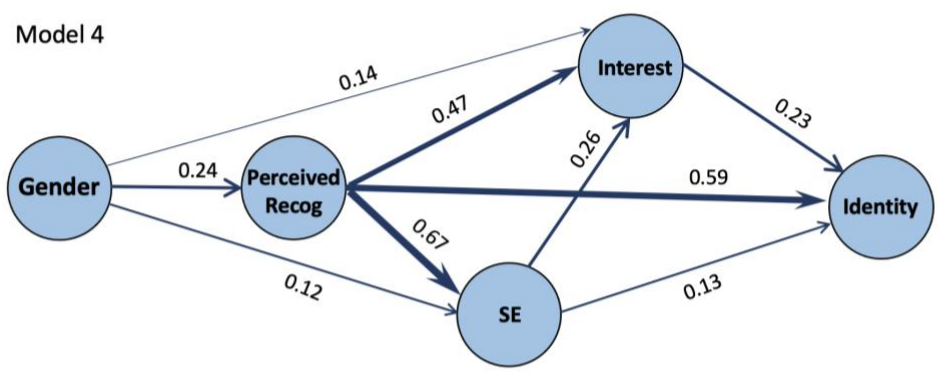



**Figure 4 Results of the path analysis part of SEM Model 4, in which perceived recognition predicts interest and self-efficacy, and self-efficacy predicts interest. Each regression line thickness qualitatively corresponds to the magnitude of $\beta$ values. All $\beta$ values shown are significant with $p$ < 0.001. For clarity, we have removed the statistically insignificant regression path from gender to identity.**

## 2.6 Summary and Discussion

Building on prior research on the gendered nature of physics identity, we investigated how the relationship between gender and physics identity is mediated by the three motivational constructs (perceived recognition, self-efficacy and interest) in different statistically equivalent SEM models. In particular, we show that there are 27 statistically equivalent SEM models (with identical model fit parameters) with different predictive relations between the three mediating constructs. We discuss how some equivalent models are preferable than others based on our theoretical framework using evidence gathered from our prior interviews with students and the affordance of each model for instructional implications, e.g., the likely interpretations of these statistically equivalent models by the physics instructors as well as educational policy makers. We emphasize the importance of selecting good models that are theoretically grounded from statistically equivalent models. These models are supported by additional evidence and have the potential to empower stakeholders such as instructors and TAs to understand the value of recognizing and affirming their students. This issue is especially important for underrepresented students in physics classes, e.g., women, because we want these students to develop a strong physics identity and see themselves as people who can excel in physics despite all the societal stereotypes and biases pertaining to who can excel in physics.

Moreover, our study shows that in current classes, there are statistically significant gender differences disadvantaging female students in all motivational constructs. Thus, to make physics learning equitable and inclusive, we highlight the importance of framing physics education research intentionally in an instructionally beneficial manner, e.g., even if the 27 models of identity with gender mediation are statistically equivalent, they can convey very different messages to physics educators about their role in supporting their students and creating an equitable and inclusive classroom. Some of these equivalent models emphasize the importance of the critical roles played by educators in supporting students, while others may inadvertently even relieve the educators of their responsibility and strengthen their fixed mindset [113,114]. Studies have indeed shown that instructors' mindset about whether all or only some of their students can excel in their courses can influence students' motivation and achievement, and underrepresented students are more likely to be demotivated and have negative experiences in classes taught by fixed mindset instructors [114].

For example, if we put interest as the predictor of both self-efficacy and perceived recognition as in Model 3 of Figure 3 (b), the important roles played by instructors and TAs in recognizing students and developing their interest and self-efficacy may not be clear to them. In fact, prior research suggests that instructors with fixed mindset believe that students come to their classes with fixed interest and there is nothing they can do if students do not have interest in the subject [113,114]. On the other hand, prior studies suggest that students' interest can be triggered and sustained by their learning environment [47,86]. For example, inclusive learning environments can trigger situational interest [115]. Thus, both prior interviews and instructional implications suggest that Model 4 in Figure 4 is a good model to select (as opposed to Model 3 of Figure 3 (b)) and can empower instructors to take responsibility to create a learning environment in which students are appropriately recognized and in turn have the opportunity to grow their physics self-efficacy and interest in a supportive environment. On the other hand, even though Model 3 fits the data well with the same model fit parameters as Model 4, it has the potential to inadvertently strengthen the notion that interest is immutable and weaken instructors' role in creating an equitable and inclusive learning environment.

In summary, according to our theoretical framework, the statistically equivalent models in which perceived recognition is the predictor of both self-efficacy and interest are good models to select, because they are supported by interview findings and can also inspire instructors to embrace their role and recognize their students appropriately since these models suggest that students' physics self-efficacy, interest and identity can be influenced by perceived recognition. Furthermore, Model 4 signifies that students' interest is not a fixed thing but rather can be influenced by both perceived recognition and self-efficacy, which instructors can improve by creating an equitable and inclusive learning environment and affirming their students for their progress. Thus, this model places the most emphasis on the importance of recognizing students and is most productive for empowering instructors and helping them develop a growth mindset about their students' potential and their role in supporting learning [113,114].

# 3.0 Selecting Pedagogically Beneficial Models That Are Theoretically Grounded from Statistically Equivalent Models of Physics Identity

## 3.1 Introduction and Theoretical Framework

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 [116-120]. In explaining participation in STEM careers, identity has been argued to be a particularly important motivational construct [21,22,58,121]. Students' identity in a domain, e.g., physics, is defined by whether they see themselves as a 'physics person' [21,22]. Prior studies have found that students' physics identity can be predicted by their other motivational beliefs such as self-efficacy, interest, and perceived recognition [21,22,57,80]. These motivational beliefs have also been shown to influence students' engagement, performance and retention in physics [18,20,21,23,45,48,122,123]. Therefore, the investigation of students' motivational beliefs and the relationships among them can help us understand students' performance and retention in physics and can provide guidelines for developing an inclusive and equitable learning environment and promoting diversity in physics.

Self-efficacy is defined as one's belief in one's ability to do well in a specific situation, task or domain such as physics [39]. Interest refers to students' curiosity, enjoyment and engagement in physics [47]. Physics perceived recognition refers to students' perception about whether others see them as a physics person [21]. Prior studies have shown that these three motivational constructs correlate to and interact with each other in meaningful ways [39,86]. Thus, one may wonder which of these constructs generally acts as a predictor for the others. For example, does interest primarily drive self-efficacy or does self-efficacy primarily drive interest? Prior studies have shown many possible predictive relationships among students' physics self-efficacy, interest, and perceived recognition [58,81,90,124,125]. For example, some prior studies used the model in which self-efficacy is the predictor of both interest and perceived recognition [58,125], while another study used the model in which interest is the predictor of both self-efficacy and perceived recognition [90]. However, these studies have not explicitly acknowledged the fact that statistical techniques used by them alone cannot be used to draw conclusions about the directionality of the associations between the constructs [88], so the model with interest predicting self-efficacy is actually statistically equivalent to the model with self-efficacy predicting interest, and they have identical model fit indices. In a recent paper, physics education researchers have been asked to explain why they select a specific model out of several statistically equivalent ones while using SEM or multiple-regression as the analysis tool . In other words, model equivalence shows that other input is required to prefer one statistically equivalent model over another [89].

Here, we propose a theoretical framework, which provides one way to select good models from several statistically equivalent ones. Our framework suggests researchers consider two aspects of the statistically equivalent models when selecting good models from them. First, what potential instructional implications each model has, i.e., whether these instructional implications have positive influence on pedagogical approaches. Second, whether the instructionally beneficial models are also supported by additional evidence in addition to model fit indices. This additional evidence may include but is not limited to researchers' own interview data or findings from prior studies such as intervention with experiment and control groups.

In this paper, we take physics identity model as an example to show how the proposed theoretical framework is used to select good models from several statistically equivalent ones for how the relationship between gender and physics identity is mediated by the three motivational constructs (self-efficacy, interest, and perceived recognition) in a college calculus-based introductory physics course. According to the Lee–Hershberger replacing rules [126], there are 27 statistically equivalent models with different predictive relations between the three mediating constructs. In this paper, we discuss 4 of these models, which are representative of the models in prior research, to elucidate how our theoretical framework guided us to select a few good models and highlight instructional implications of some of the models. As shown in Figure 5, we first considered a model that only has covariances among self-efficacy, interest, and perceived recognition. Then, we considered other statistically equivalent models to the one shown in Figure 5 by making one of the mediating constructs as the predictor of the other constructs (discussed later in greater detail).

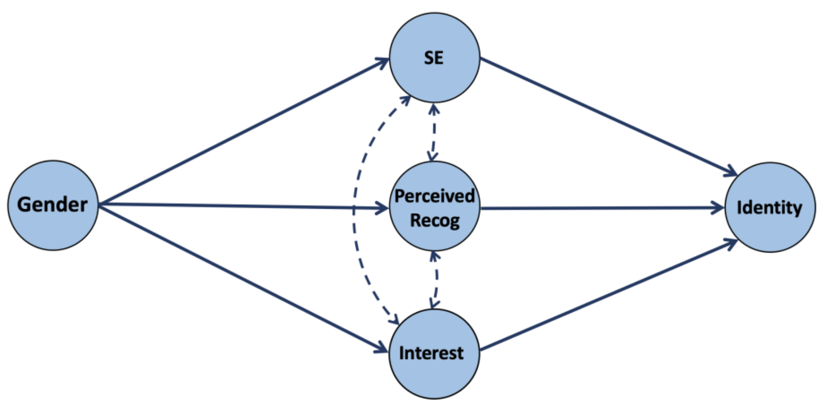



**Figure 5 Schematic representation of the path analysis that shows how the relationship between gender and physics identity is mediated by self-efficacy (SE), interest, and perceived recognition (Recog) in one of the 27 statistically equivalent models. The covariances between the three mediating constructs are shown with dashed lines. The direct path from gender to identity is not shown (since it is not statistically significant in our model).**

## 3.2 Research Questions

Our research questions are as follows (please note that RQ1 focuses on descriptive statistics, which is necessary to draw meaningful inferences from the predictive statistics in RQ2):

**RQ1.** What are the gender differences in a set of physics motivational beliefs, i.e., self-efficacy, interest, perceived recognition, and identity, for students at the end of the course?

**RQ2.** How can we narrow down from a large number of statistically equivalent SEM models (in which the relation between gender and physics identity is mediated via self-efficacy, interest, and perceived recognition) based upon our theoretical framework using potential instructional implications of the models as well as other data?

## 3.3 Methodology

### 3.3.1 Participants

The motivational survey data used in this study were collected at the end of the second course of a two-term college calculus-based introductory physics sequence (physics 2) in two consecutive spring semesters. Physics 2 includes topics such as electricity and magnetism, electromagnetic waves, images, interference, and diffraction. This course is a traditional lecture-based course (4 hours per week) with recitations (1 hour per week) in which students typically work collaboratively on physics problems. This course is generally mandatory and taken by those intending to major in engineering, physical science, and mathematics in the second semester of the first year of their undergraduate studies. In physics 2, after students have been on campus for a semester, the feeling of uncertainty and anxiety during the transition to college may decrease. As students gradually adapt to the new environment, their motivational beliefs may also become more stable. In addition, students in the introductory physics course are admitted to engineering school and school of arts and sciences as undecided majors, and they usually declare their majors in their second year, which is after physics 2. Thus, their physics motivational beliefs at the end of physics 2 are important in influencing their major decisions in engineering and physical science disciplines. Furthermore, the introductory physics sequence is generally mandatory for these students because it is the foundation of many disciplines and contributes directly to engineering, and most scientific fields, so their motivational beliefs at the end of this sequence may also have a long-term influence on their future studies and career. Therefore, in this study, we focus on students' motivational beliefs and the predictive relationships among them at the end of physics 2.

This research protocol was approved and carried out in accordance with the principles outlined in the university institutional review board (IRB) ethical policy. The paper surveys were handed out and collected by TAs in the last recitation class of a semester. We encouraged the instructors to give students small amount of course credit or extra credit for completing the survey. The demographic data of students—such as gender—were provided by the university. Students' names and IDs were de-identified by an honest broker, so researchers could analyze students' data without having access to students' identifying information. There were 915 students participating in this survey, including both semesters. In our final data analysis, we focused on 907 students (including 299 female students and 608 male students) who provided their gender information. Less than 1% of the participants who did not provide this information were not included in this analysis.

### 3.3.2 Instrument Validity

In this study, our analysis includes four motivational constructs—physics self-efficacy, interest, perceived recognition, and identity. Table 4 shows the survey questions for each motivational construct. We adapted these questions from existing motivational research [21,92-94] and re-validated them in our prior work [37,65,95,127,128]. The validation and refinement of the survey involved use of individual interviews with students, exploratory and confirmatory factor analyses (EFA and CFA) [97], Pearson correlation between different constructs and Cronbach's alpha [98-100].

In our survey, each item was scored on a 4-point Likert scale (1-4) [102] with higher scores indicating greater levels of self-efficacy, interest, perceived recognition and identity. Most of the self-efficacy and interest questions had response options 'NO!, no, yes, YES!', which have been shown to have good psychometric properties and a low cognitive load while reading [82,92]. The questions under physics identity and perceived recognition all had response options 'strongly disagree, disagree, agree, and strongly agree'.

**Table 4 Survey questions for each motivational construct, along with CFA item loadings. Lambda represents the factor loading of each item, which is the correlation between the item and the construct. 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. †The response options for this question are 'Never, Once a month, Once a week, Every day'. ‡The response options for this question are 'very boring, boring, interesting, very interesting'.**

| Construct and Item | Lambda |
|---|---|
| Physics Identity | |
| I see myself as a physics person. | 1.000 |
| Physics Self-Efficacy (Cronbach's alpha = 0.81) | |
| I am able to help my classmates with physics in the laboratory or in recitation. | 0.698 |
| I understand concepts I have studied in physics. | 0.742 |
| If I study, I will do well on a physics test. | 0.735 |
| If I encounter a setback in a physics exam, I can overcome it. | 0.715 |
| Physics Interest (Cronbach's alpha = 0.85) | |
| I wonder about how physics works † | 0.701 |
| In general, I find physics ‡ | 0.811 |
| I want to know everything I can about physics. | 0.829 |
| I am curious about recent physics discoveries. | 0.720 |
| Physics Perceived Recognition (Cronbach's alpha = 0.87) | |
| My family sees me as a physics person. | 0.885 |
| My friends see me as a physics person. | 0.908 |
| My physics TA and/or instructor sees me as a physics person. | 0.730 |

We then calculated the Pearson correlation coefficients pairwise between the constructs. As shown in Table 5, these constructs have strong correlations with each other, but the correlations are not so high that these constructs cannot be examined separately [109]. Table 5 shows that the highest correlation coefficient is between physics identity and perceived recognition, which is consistent with Godwin et al.'s and Authors' prior work finding that perceived recognition (external identity) is the strongest predictor of physics identity (internal identity) [58,81].

**Table 5 Pearson correlation coefficients of the motivational constructs.**

| Variable | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. Physics identity | -- | -- | -- | -- |
| 2. Self-efficacy | 0.71 | -- | -- | -- |
| 3. Interest | 0.69 | 0.63 | -- | -- |
| 4. Perceived Recognition | 0.83 | 0.70 | 0.68 | -- |

### 3.3.3 Quantitative Analysis of Survey Data

First, we calculated the mean score for each construct for each student. Then, we used a t-test [103,104] to compare the mean scores by gender. Next, we performed the structural equation modeling (SEM) [87] to analyze the predictive relationships among students' gender, self-efficacy, interest, perceived recognition, and identity using our survey data. Initially, we tested different levels of measurement invariance model. In each step, we fixed different elements of the model to equality across gender and compared the results to the previous step using the Likelihood Ratio Test [87]. Results showed that in all of our models, measurement invariance holds and regression pathways among the constructs do not have differences across gender. These results allowed us to perform the gender mediation SEM model [87].

As noted, there are 27 possible statistically equivalent models showing different predictive relationships among the three mediating constructs (self-efficacy, interest, and perceived recognition). For brevity, here we explicitly illustrate four of them, which are representative of the models in prior research. We first delineate the model fit indices of these models to make sure they are good models from a statistical perspective. Then, we argue that some of these statistically equivalent models are better than others based upon our theoretical framework focusing on model's instructional implications and supporting evidence from individual interviews with students. In particular, we take advantage of the data from our prior hour-long empathetic individual interviews with 52 student volunteers (42 women and 10 men) in the introductory physics course [71,76,112,129]. Even though these interviews were conducted in the context of our prior studies, the interview findings do provide some extra evidence that can be used to select good models. The interviews followed a semi-structured think aloud protocol established by the researchers prior to conducting the interviews [130]. We call these interviews empathetic interviews because the goal of the interviews is to understand the experiences of students in physics courses in order to improve equity and inclusion. In particular, inspired by standpoint theory [131,132], we centered the experiences of students from underrepresented groups, e.g., women, in order to understand how to support them and improve the learning environment. Therefore, there were more women who were interviewed than men. During the interviews, we asked students a set of questions about their experiences and asked them to think aloud as they answered the questions. We did not disturb the students when they thought aloud in order to not disrupt their thought processes and asked them for clarifications of the points they may not have made only after they had finished their thoughts. We coded the interviews using hybrid coding methods that involved both deductive and inductive coding [133]. One of the authors coded the interviews and both authors discussed and converged on the codes developed based on the student interviews. In those interviews, one theme that kept recurring was that recognition by instructors and teaching assistants is very important in shaping students' self-efficacy and interest. We make use of this finding to confirm that the instructionally beneficial models are also supported by the interview data.

## 3.4 Results and Discussion

### 3.4.1 Gender Differences in Motivational Beliefs and Grades

Pertaining to RQ1, Table 6 shows that women have significantly lower average scores in all of the four motivational constructs than male students, and the effect sizes given by Cohen's *d* are all in the medium range [104].

**Table 6 Descriptive statistics for female and male students' motivational beliefs in which M stands for construct mean value, SD is the standard deviation and N is the number of students. All effect sizes (Cohen's *d* values) are statistically significant with $p < 0.001$ and a minus sign indicates that male students have a higher score than female students.**

| Constructs | Females N = 299 | | Males N = 608 | | Cohen's d |
|---|---|---|---|---|---|
| | M | SD | M | SD | |
| Physics Identity | 2.13 | 0.83 | 2.67 | 0.86 | -0.63 |
| Perceived Recognition | 2.24 | 0.71 | 2.71 | 0.70 | -0.66 |
| Self-efficacy | 2.64 | 0.57 | 2.92 | 0.56 | -0.48 |
| Interest | 2.61 | 0.66 | 3.02 | 0.62 | -0.65 |

### 3.4.2 Equivalent SEM Models and Path Analysis

In order to understand why there are 27 statistically equivalent models in which the relation between gender and physics identity is mediated by self-efficacy, interest, and perceived recognition, we consider the associations among these three mediating constructs. According to the replacing rules [126], there are three possible associations between two endogenous variables. These associations are covariance, direct effect via regression from one to the other, or direct effect via regression in the reverse direction. Therefore, with the constraints that no regression arrows point to gender and arrows can only point to physics identity since it is the outcome variable, there are $3 \times 3 \times 3 = 27$ statistically equivalent SEM models in total. The level of SEM model fit can be represented by Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). In our study, the model fit indices are the same for all 27 models and suggest a good fit to the data: CFI = 0.968 (> 0.90), TLI = 0.956 (> 0.90), RMSEA = 0.063 (< 0.08) and SRMR = 0.034 (< 0.08) [98]. Thus, these statistically equivalent SEM models are all robust from a statistical point of view. In this section, for brevity, we explicitly discuss only 4 of these 27 equivalent models, which are representative of the models in prior research, to illustrate how to use our theoretical framework to select good models from many statistically equivalent ones.

First, we consider a model (Model 1) in which there is no predictive relationship among self-efficacy, interest and perceived recognition. Instead, there are covariances between each pair. Figure 6 shows the path analysis results of this SEM model. We note that even though female students' physics identity is significantly lower than that of male students, as shown in Table 6, Model 1 indicates that gender does not directly predict physics identity, which means that the gender difference in physics identity is mediated through self-efficacy, perceived recognition and interest, which is consistent with prior study [81]. However, in Model 1, gender is the only predictor of self-efficacy and interest, which may be interpreted to imply that female students have lower physics identity because they have lower self-efficacy and interest in physics. Considering self-efficacy and interest may be perceived as fixed by many people, Model 1 does not provide suggestion with regard to instructors having the potential to improve students' self-efficacy and interest. Thus, this model in Figure 6 is not the most instructionally beneficial model.

In addition, Model 1 also does not fully reflect our finding from prior interviews. In particular, our interviews data show that recognition by instructors and teaching assistants is very important in shaping students' self-efficacy and interest [71,76,112,129]. For example, 31% of the interviewed students suggested that it would be helpful if instructors could realize that physics is challenging and acknowledge that it's okay to not completely understand the content right away and they just need more practice in order to develop a good grasp of concepts. In addition, 19% of the interviewed students mentioned that they felt really encouraged and became more interested in physics when their opinions and success in physics problem solving were recognized by their instructors. On the other hand, students' self-efficacy and interest deteriorated when there was lack of appropriate recognition from instructors/TAs. For example, 38% of the interviewed students reported that when they went to the course instructor or TA to ask for help on physics problems, sometimes they were explicitly told that the problems were "*easy*", "*obvious*" or "*trivial*", which they perceived as disparaging or belittling in that they felt they were being told that they are not smart enough to do physics if they could not do such easy problems on their own. Thus, our interview data indicate that students' positive or negative perceived recognition can affect their physics self-efficacy and interest in the course and in physics overall. However, the important role played by perceived recognition on self-efficacy and interest is not reflected in Model 1, in which there are only covariances among self-efficacy, perceived recognition, and interest.

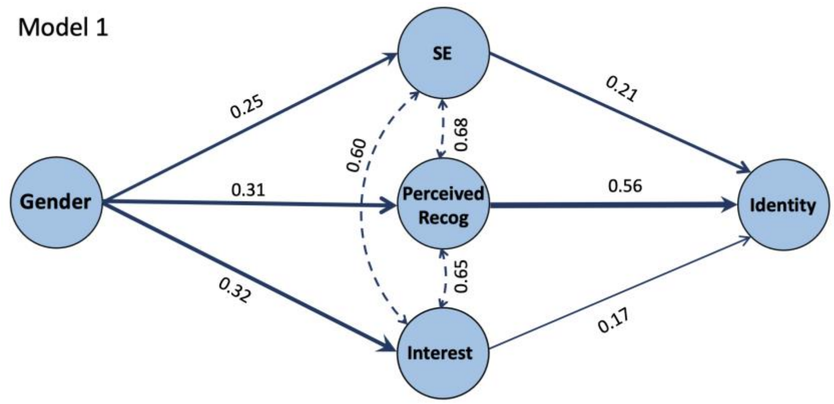



**Figure 6 Results of the path analysis part of SEM Model 1, in which there are only covariances between each pair of constructs: self-efficacy (SE), perceived recognition (Recog), and interest. The dashed lines represent residual covariances between constructs. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients ($\beta$ values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the $\beta$ value. All $\beta$ values shown are significant with $p < 0.001$. For clarity, we have removed the statistically insignificant regression path from gender to identity.**

Next, we consider two models in which self-efficacy and interest each act as predictors of the other two mediating constructs (see Figure 7). In particular, in Model 2, self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition. In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition.

With regard to potential instructional implications, Models 2 and 3 may also not be consistent with our theoretical framework for selecting good models. In Models 2 and 3, gender is the only predictor of interest or self-efficacy which together with the data in Table 6 (which shows gender differences in all of these motivational constructs) can be interpreted as deficit models. In addition, Models 2 and 3 emphasize that students' perceived recognition is predicted by their interest and self-efficacy. Thus, these models may be interpreted to imply, e.g., that women are not feeling positively recognized by their instructors and teaching assistants (TAs) as much as men because they have lower interest in physics and lower self-efficacy than men. Therefore, Models 2 and 3 can be misinterpreted and instead of empowering instructors to recognize their students whenever an opportunity arises, they may lead to instructors becoming complacent about their own role in helping students develop identity as someone who can excel in physics.

In addition, Models 2 and 3 suggest that students' perceived recognition is predicted by their self-efficacy and interest, which is also not consistent with our prior interview findings showing that perceived recognition may be an important driver of the other two constructs. Moreover, our interviews show that women were less likely than men to feel positively recognized by physics instructors/TAs, and this lack of recognition or discouraging feedback from instructors/TAs deteriorated their self-efficacy, and lowered perceived recognition and self-efficacy further lowered their interest [71,76,129]. For example, 21% of the interviewed women explicitly noted that their instructors/TAs sometimes showed more interest in male students' questions and answered male students' questions with more attention than when they answered their questions. This experience made the female students wonder whether it was because their questions were not good or too easy, and thus they started doubting their ability to do well in the course. Thus, Models 2 and 3 showing that students' perceived recognition is predicted by their interest and self-efficacy are not aligned with our framework in regard to their potential instructional implications and lack of support from our prior interview data.

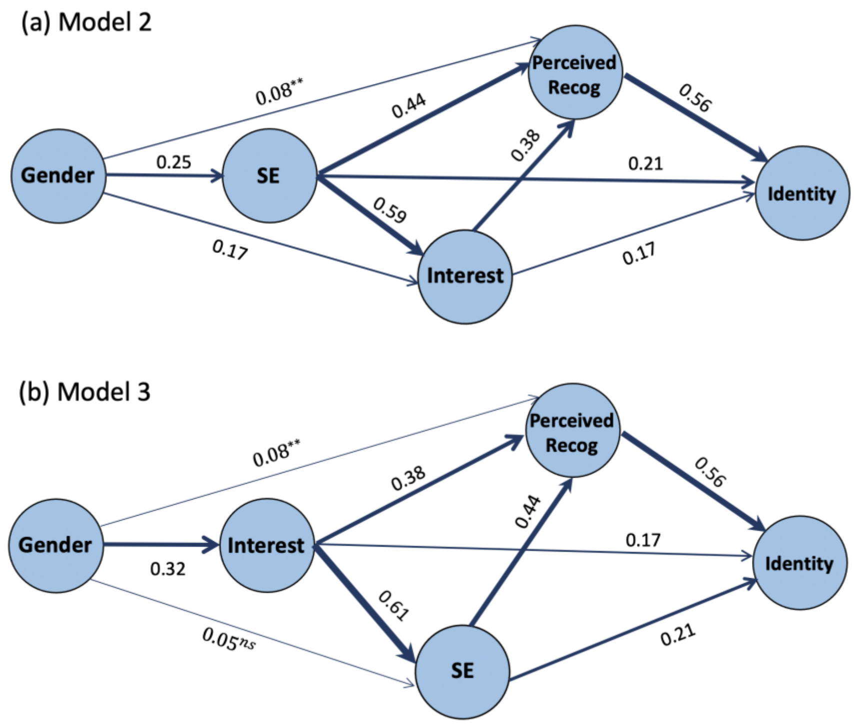



**Figure 7 Results of the path analysis part of the SEM models that show how the relationship between gender and physics identity is mediated through self-efficacy (SE), interest, and perceived recognition (Recog). (a) In Model 2, self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition. (b) In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition. Regression coefficients with $0.001 \leq p < 0.1$ are indicated by $^{**}$ and with $p > 0.05$ are indicated by ns. All the other regression lines show relations with $p < 0.001$. For clarity, we have removed the statistically insignificant regression path from gender to identity.**

Finally, we consider a model (Model 4) in which perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest (see Figure 8). Even though statistically equivalent to Models 1-3, Model 4 is more likely to empower the instructors and give them the message that students' interest and self-efficacy in physics can be influenced by the recognition they receive from instructors. In addition, in Model 4, gender is the direct predictor of perceived recognition, which is consistent with Table 6 showing that women feel significantly less recognized as a physics person. Thus, Model 4 is also more likely to inspire instructors to recognize their role in creating more inclusive and equitable learning environments in which all students including those from the underrepresented groups in physics such as women feel more positively recognized and affirmed. Moreover, Model 4 may be even more instructionally beneficial than other models, e.g., in which gender predicts perceived recognition but interest predicts self-efficacy unlike Model 4 because interest may be perceived as fixed by some instructors.

As noted, according to our prior interview data, perceived recognition plays an important role in the development of students' self-efficacy and interest. Also, our interviews show that female students feel less recognized by their instructors/TAs than their male peers, which indicates that the current learning environment is not inclusive and equitable. For instance, 31% of the interviewed female students mentioned that instructors/TAs treated men and women differently in class. For example, they noted that men in their physics courses were generally praised more by the instructors/TAs than women, and sometimes instructors/TAs called men who answered the questions "*brilliant*", which made them feel as though they were not brilliant. Furthermore, in the interviews, 24% of the interviewed women reported that because of the negative experiences in their physics courses, they had contemplated switching out of their major (either engineering or physics) while men never expressed similar concerns [76,129]. Thus, Model 4 with perceived recognition predicting self-efficacy and interest model is not only more instructionally beneficial than the other models discussed, but it is also supported by our prior interview data.

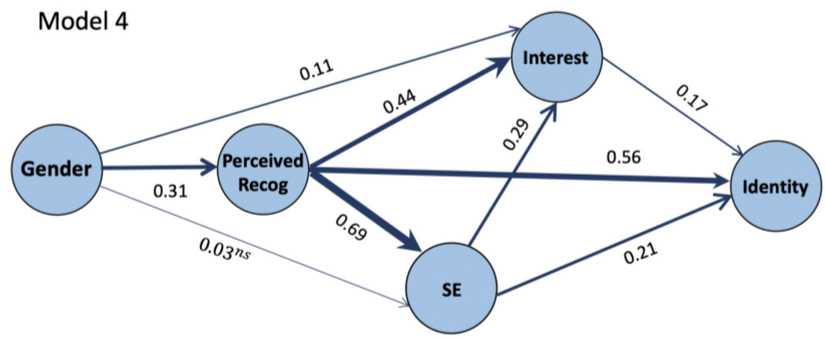



**Figure 8 Results of the path analysis part of SEM Model 4, in which perceived recognition predicts interest and self-efficacy, and self-efficacy predicts interest. Each regression line thickness qualitatively corresponds to the magnitude of $\beta$ values with $p$ > 0.05 are indicated by ns. All $\beta$ values shown are significant with $p$ < 0.001. For clarity, we have removed the statistically insignificant regression path from gender to identity.**

## 3.5 Summary

Building on prior research on the gendered nature of physics identity, we investigated how the relationship between gender and physics identity is mediated by the three motivational constructs: perceived recognition, self-efficacy and interest. In particular, we show that there are 27 statistically equivalent SEM models (with identical model fit parameters) with different predictive relations between the three mediating constructs. We discuss how the models with perceived recognition predicting self-efficacy and interest are more instructionally beneficial than the others, and they are also supported by our prior interview data. In particular, the models with perceived recognition predicting self-efficacy and interest emphasize the importance of the roles played by educators, while others may inadvertently even relieve the educators of their responsibility and strengthen their fixed mindset [113,114]. Therefore, as instructors, we should take the responsibility to create an inclusive and equitable learning environment in which all students are appropriately recognized and validated and have the opportunity to grow their physics motivational beliefs. As researchers, we should consider the instructional implications of our work and frame our research in a way which is instructionally beneficial.

### 3.6 Limitations and Future Directions

In this study, we emphasize the importance of considering statistically equivalent models and provide one way to select good models from the statistically equivalent ones using instructional implications and extra evidence. In particular, we discussed four physics identity models, which are representative of the models in prior research to elucidate how our theoretical framework guided us to select good models. There are several other statistically equivalent models not discussed in this paper, e.g., models involving self-reinforcing feedback loops, which would be worth investigating in future studies. In addition, even though the single physics identity item is commonly used in studies involving physics identity [58,90,124,125], it would be helpful in future studies to develop more survey items for this construct. In addition, similar to prior studies [58,90,124,125], the data used in this study were collected at one time point. Future studies using longitudinal data may provide further understanding of the predictive relationships among students' motivational beliefs. In this study, since the gender data were collected using only binary categories, we did not have the gender information of students who did not identify as male or female. This issue has been resolved very recently in the manner in which our university is now collecting data. In addition, since the sample size of these students is small (less that 1% of participants), we would not be able to analyze them as separate groups using multigroup analysis in SEM even if we knew their gender identity. Future studies can use other research methods to investigate motivational beliefs of students in other gender categories. In the future studies, we also intend to investigate motivational beliefs of students from other underrepresented groups such as ethnic/racial minority students. Even though this study is conducted in the introductory calculus-based physics course, it would be valuable to investigate the relationship among motivational beliefs of students in other courses, such as algebra-based physics courses, where women are the majority group, often making up 60% or more of the classroom. In addition, it would be also useful to investigate students' physics motivational beliefs in advanced physics courses, which are usually taken by physics major students beyond the first year. Similar studies in different types of institutions and in other countries would also be helpful for developing a deeper understanding of the relationships among students' motivational beliefs in different contexts.

# 4.0 The Impact of Perceived Recognition by Physics Instructors on Women's Self-Efficacy and Interest

## 4.1 Introduction

In the disciplines of science, technology, engineering, and mathematics (STEM), there have been efforts to enhance the participation and advancement of marginalized groups such as women [1,2,4,6,7,13,20-22,36,37,63,71,72]. Prior research suggests that individuals' course enrollment, degree attainment and achievement in STEM can be influenced by their domain specific motivational beliefs such as self-efficacy, interest and identity [20,37,41,44,45,48,63,73]. For students from marginalized groups, these motivational beliefs might be undermined, e.g., by negative societal stereotypes and biases about who belongs in STEM 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 [24,28,74-77]. Hence, investigating students' motivational beliefs is critical to understanding and addressing diversity, equity, and inclusion issues in STEM disciplines.

Self-efficacy, which is defined as students' beliefs in their capability to succeed in a certain situation, task, or particular domain [39,81,82], has been shown to influence students' engagement and performance in a given domain [41,44]. 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 [45]. In addition, studies suggest that students' self-efficacy predicts their career choices and persistence toward their career goals [41,73].

Another important motivational belief is interest, which is defined by positive emotions accompanied by curiosity and engagement in particular content [47]. Interest has also been shown to influence students' learning [41,47,48]. 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 [52].

According to Eccles's Expectancy-Value Theory (EVT) [53,54], interest is paired well with self-efficacy as connected constructs that predict students' academic outcome expectations and career aspirations. Moreover, prior studies [21,22,58,81] show that students' self-efficacy and interest are two important predictors of their identity in a given domain. Students' identity in an academic domain, such as physics, relates to their views about whether they see themselves as a "physics person" [21,22], and it has been shown to influence students' career decisions as well as short- and long-term academic goals.

Prior quantitative studies have shown that female students often report a lower level of self-efficacy, interest and identity than their male peers in many STEM fields [38,134-136]. Moreover, a prior study showed that female students with an A grade had similar physics self-efficacy as male students with a C grade by the end of a two-semester introductory calculus-based physics course (whereas these women with A grades had the same self-efficacy as men with B grades at the end of the first-semester physics course) [38]. In other words, the gender difference in self-efficacy increased over time. This study is consistent with another study showing that female students are more likely to drop STEM majors with significantly higher grade point averages than their male peers [137]. In order to shed light on these types of quantitative findings, qualitative studies probing the experiences of female students in physics courses can be invaluable. These qualitative studies can also shine light on whether women have differential experiences compared to men in physics courses and what instructors and TAs could do to create an equitable and inclusive learning environment and better support female students in these courses.

## 4.2 Theoretical Framework

In this study, we conducted individual hour-long interviews with 38 women in both introductory and advanced physics courses at a large research university in the US to investigate their experiences and how these experiences shape their motivational beliefs such as self-efficacy and interest. Our interviews are inspired by the framework of standpoint theory [131,132], which emphasizes that one should center the experiences of people from traditionally marginalized groups since they have experienced inequities and will be able to articulate how those inequities manifest in their everyday experiences and point to the issues within the system that must be fixed. Therefore, in these interviews, we centered the experiences of female students in physics courses from introductory to advanced levels in order to understand how to support them and improve the physics learning environment.

In addition to the theoretical framework of standpoint theory, our interviews are also inspired by the framework of physics identity [21,81]. According to the identity framework, in addition to self-efficacy and interest, individuals' identity pertaining to a domain such as physics is predicted by their perceived recognition from others. Perceived recognition in a domain, such as physics, refers to students' perception about whether other people see them as a physics person [79]. Prior studies have shown that female students did not feel that they were recognized appropriately in many science domains even before they entered college [66,83]. A report from the National Science Foundation [84] 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, interactions and responses from others were often characterized by girls as ambivalent, with lack of encouragement, or suggestions that their goals were inappropriate [84]. Prior studies show that similar biases and stereotypes also impact female students in the university context [81,85]. For example, one study showed that science faculty members in biological and physical sciences exhibited biases against female students and rated male students as more competent, and they were more likely to hire them and pay them more even though only the names were different in the hypothetical information they were provided about the student [85]. Although prior studies have shown that students' perceived recognition plays an important role in shaping their identity in a given domain, the relationships between perceived recognition and the other predictors of identity, i.e., self-efficacy and interest, are not as clear. Therefore, in this qualitative study, we focused on female students' perceived recognition from their instructors and TAs in physics courses to investigate the relationships between students' perceived recognition and their other physics beliefs such as self-efficacy and interest.

### 4.3 Research Questions

Consistent with our theoretical framework, we seek to address the following research questions:

**RQ1.** What are female students' perceptions of different types of recognition from instructors/TAs in physics courses?

**RQ2.** How do female students' perceived recognition from instructors/TAs shape their self-efficacy and interest?

## 4.4 Methodology

### 4.4.1 Participants

We conducted semi-structured, empathetic interviews with 38 female students in physics courses and majoring in Physics and Astronomy, Engineering/Computer science, or Chemistry (includes those on a bio track) at a large research university in the US. Although 22 of these women were White, 12 were Asian, 3 were Hispanic and 1 was Black, our focus here is on their experiences as women in physics. The self-reported year and discipline are shown in Table 7. We sent the interview advertisement to the department's undergraduate administrative assistant to share it with all of the undergraduate students in the physics department, and we also sent the advertisement to instructors of different physics courses and asked them to share it with their students. Each student received a $25 gift card for participating in an hour-long interview. Roughly half of these students were interviewed before the COVID-19 pandemic and half of them during the pandemic. The interviews followed a semi-structured think aloud protocol with interview questions that were agreed upon by the researchers prior to the interviews [130]. We call these interviews empathetic interviews because the goal of the interviews was to understand the experiences of female students in physics courses in order to improve equity and inclusion. In particular, before interviews, we communicated with students that the broader goal of the research is to make the physics learning environments equitable and inclusive. All participants agreed to be audio-recorded and quoted in academic publications. Students also had the opportunity to ask questions about the research before and after the interviews.

**Table 7 Self-reported year and discipline of the female students (N=38) interviewed.**

| | |
|---|---|
| Year | |
| First year | 20 |
| Second year | 8 |
| Third year | 3 |
| Fourth year | 6 |
| First year graduate student talking about undergraduate experience | 1 |
| Discipline | |
| Physics and Astronomy | 19 |
| Engineering/Computer science | 17 |
| Chemistry | 2 |

### 4.4.2 Semi-structured Interviews

To investigate female students' learning experience in physics courses and their perception of the learning environment, we assembled and refined a list of potential interview questions via an iterative process between the two researchers. These included questions about the student's background (e.g., how they got interested in STEM, early experiences in K-12 including high school experience); overall college experience so far; experience in physics courses (such as their interaction with instructors, TAs and peers) both inside and outside of the classroom; and perception of their learning in physics courses (including any challenges in learning, thoughts on underrepresentation of women and how to improve the physics learning environments). Also, inspired by the framework of physics identity [21,81], we focused on female students' perceived recognition from instructors and TAs to investigate the relationship between students' perceived recognition and their other motivational beliefs such as self-efficacy and interest. Therefore, many questions in the interview are used to elicit students' perceived recognition from instructors and TAs. Examples of these types of questions are shown in Table 8. We made these semi-structured hour-long individual interviews empathetic to give students the opportunity to express themselves freely, dig deeply on critical issues of equity and inclusion in physics courses, and make sure they felt comfortable expressing themselves. We asked students to think aloud as they answered the questions, and we did not disturb them when they thought aloud in order to not disrupt their thought processes and asked them for clarifications of the points they may not have made only after they had finished their thoughts. Most participants required little prompting and were very keen to share their thoughts with us openly. All interviews were audio-recorded and transcribed.

**Table 8 Examples of the interview questions that can elicit students' perceived recognition from instructors and TAs.**

| Example questions |
| --- |
| How do you think your experiences in your physics classes might have affected your identity as a physics person? |
| Do you feel respected as a physics person/someone who can excel in physics courses by your peers, TAs, and instructors? |
| Do you feel supported by your instructors and TAs/UTAs in your physics course(s)? |
| Do you feel that your contributions are valued by your peers, TAs, and instructors? |
| Do you think your experiences in physics have been different because of your gender? If so, how? |
| Do you think your gender has had an impact on your success as a physics student/a student in physics courses? |
| Have you witnessed any barriers to success that students in physics at this institution have experienced because of their gender? |
| Do you think that it is more difficult to succeed in physics/physics courses as a woman? Why/why not? |

### 4.4.3 Data Analysis

We coded the interviews using hybrid coding methods that involved both deductive and inductive coding [133]. Initially, deductive coding was used based on the interview protocols, but after reading through the interviews, we incorporated inductive coding to encompass different aspects of the interviewed women's experiences. The first author coded the interviews using Nvivo and both authors discussed and converged on the codes developed based on the student interviews. The codes themselves are inspired by the research questions and interviews themselves, in which we center the experiences of the female students in physics courses in order to understand how to better support them and improve the physics learning environments. Both authors discussed how the codes could be combined to form greater themes. In all, 43 codes resulted in 5 broad analytic themes (Perceived recognition from instructors and TAs, High school and other pre-college experiences, Interactions with peers, Research experiences, and Suggestions to improve physics learning environments). In this paper, we focus on the first theme, perceived recognition from instructors and TAs, and the relationships between students' perceived recognition and their self-efficacy and interest.

## 4.5 Results

In our interviews, 26 (68%) of the interviewed female students reported negative perceived recognition or lack of positive recognition from their physics instructors or TAs, while only 4 students (11%) reported positive perceived recognition. The other students did not provide answers to questions related to perceived recognition that were clearly categorized as positive or negative by researchers. We divide the interview findings into three broad sections: (i) Negative perceived recognition or lack of positive recognition from instructors/TAs; (ii) Impact of negative perceived recognition or lack of positive recognition from instructors/TAs on students' self-efficacy and interest; (iii) Positive recognition from instructors/TAs and its influence on students' self-efficacy.

### 4.5.1 Negative Perceived Recognition or Lack of Positive Recognition from Instructors/Tas

In our interviews, 68% of the interviewed female students reported negative perceived recognition or lack of positive recognition from their physics instructors/TAs. The negative perceived recognition or lack of positive recognition by female students from instructors/TAs can be categorized into four subthemes: (1) Feeling belittled for questions or efforts; (2) Feeling marginalized due to differential gender dynamics; (3) Feeling that the physics learning environment is unsafe; (4) Feeling negatively recognized about their abilities and potential. As shown in Table 9, each subtheme includes two codes, and each code is followed by a definition and an illustrative example from the participants in this study.

**Table 9 Subthemes for negative perceived recognition or lack of positive recognition from instructors/TAs. The percentages in parentheses represent the proportions of the interviewed women whose experiences were coded under each subtheme. In total, 68% of the interviewed female students reported perceived recognitions that were coded under at least one of the following subthemes.**

| Subtheme | Code | Definition | Example from participant data |
|---|---|---|---|
| Feeling belittled for questions or efforts (17/38 = 45%) | Belittling students' questions or struggles | Expressing that students' questions or some concepts that students are struggling with are easy, obvious or trivial | *"…I thought this was so easy… I'm disappointed you guys didn't get this."* |
| | Belittling students' efforts | Negative recognition or lack of positive recognition for students' efforts, improvement, and achievement | *"Once he finally comes over to help me, he doesn't actually acknowledge any of the work that I've done…he doesn't try to find a solution from what I started, he just does it his own way."* |
| Feeling marginalized due to differential gender dynamics (14/38=37%) | Differential treatment of female and male students | Responding to or treating men and women differently when interacting with students | *"I've interacted with, very sadly, a TA who has sort of brushed off my question and then answered a guy's question in my group…"* |
| | Letting men dominate the class | Letting men dominate the class so that the classroom dynamics dominated by men make women feel marginalized | *"They had men ask and answer questions and everyone else was sitting quietly"* *"Once I asked a question in office hour, he asked a guy to do it for me on the board."* |
| Feeling that the physics learning environment is unsafe (11/38=29%) | Condescending or intimidating behavior | Displaying a feeling of patronizing superiority or causing feelings of fear in students about asking questions | *"My professor was pretty condescending, just not really willing to help… I was personally very terrified of him. I never wanted to go to him with anything."* |
| | Cold calling | Cold calling students randomly to answer questions or show work in front of their peers | *"He would ask you to write stuff on the board in front of the whole group. It was very intimidating."* |
| Feeling negatively recognized about their abilities and potential (6/38=16%) | Underestimating students' abilities | Doubting some students' ability to do well in a task or having low expectations of some students | *"He was like leaning over my shoulder and like telling me one key at a time what to type instead of just trusting me to be able to spell the word."* |
| | Fixed mindset about students' potential | Emphasizing brilliance rather than efforts | *"He definitely thinks that some people cannot do physics."* |

#### 4.5.1.1 Feeling Belittled for Questions or Efforts

In our analysis, 45% of the interviewed female students reported experiences that were coded under the subtheme *feeling belittled for questions or efforts*, which includes two codes: *belittling students' questions or struggles* and *belittling students' efforts*.

**(a) Belittling students' questions or struggles**

This code includes instances in which physics instructors and/or TAs express that students' questions or some concepts that students are struggling with are easy, obvious or trivial. Often these comments were given by instructors and TAs explicitly using the words "easy", "trivial", and "obvious", which made these female students feel that their questions were not good and they did not want to ask more questions. Hailey shared an experience from her Electricity and Magnetism class:

*Hailey: He [the instructor] was also the kind of professor that would say like, "oh, all this information is trivial, so let me just like [go over it] really fast and assume that you know it" ... So I just felt like I didn't want to go and like ask a stupid question to him.*

Mary shared a similar experience in her physics class:

*Mary: My physics professor said, like, that, things were obvious a lot. And according to my TA, who's worked closely with him, he [the instructor] says that to make it [the question] not seem as overwhelming. But for me, at least, the effect was much more like, if you don't know this, you're dumb. So that made it difficult to like, ask questions in class.*

Thus, Hailey and Mary note that the comments from their physics instructors make them hesitate to ask questions to their instructors because they are worried that their questions will look "stupid". Mary also mentioned that even though her TA thinks that the instructor may want to make students feel at ease and encourage them by saying that the problems are easy or trivial, these comments have a negative impact on her and make her feel "dumb" rather than encouraged.

Sometimes, instructors may not explicitly use words such as "easy", "trivial", "obvious", etc. directly; however, the interviewed students reported that the way instructors communicate with students can also make them feel nervous about not knowing and asking questions. Fem shared her experience of interacting with her physics instructors:

*Fem: Sometimes when they [the physics instructors] are explaining something, they kind of make you feel dumb. It's never their intention, it [is] just clearly in their tone [and] in the way that they talk, it's like you should know this…I don't like going to my teachers, because sometimes they make me feel dumb.*

Elaine recalled a similar experience in a physics instructor's office hours:

*Elaine: I had some bad experiences [in office hours], where like, I would ask a question, and then I'd just feel the way he would explain it would make me feel like really dumb … because I had like a few and, what I think are embarrassing moments. I stopped going to office hours, to my professor's office hours in fall semester.*

As we can see from both Fem and Elaine's experiences, even without explicitly saying something is "easy" or "trivial", the tone or body language of instructors can also convey similar message that can discourage students. In addition, instructors may not even realize it when they are conveying this kind of message to students. Fem later repeated that "*It's never their intention*"; however, it still negatively influenced her. It is important for instructors to understand that it is really the effects of instructors on students that matter rather than instructors' intentions.

In addition to body language, when instructors respond to students' questions quickly and carelessly or they show lack of interest in students' questions, this can also convey the message that students' questions are not good, worth asking or important. Hailey recalled that in office hours, her physics instructor seemed to be more interested in hard physics questions (usually asked by male students), which made her nervous about asking homework questions.

*Hailey: Even though he's like, "Oh yeah, come ask me questions", he didn't really seem like he was interested in helping me… he seemed like he was interested in answering what he considered to be good questions about physics whereas like, I just wanted to know how to do homework, you know?*

Shreya narrated that the quick and careless response from her physics instructor was very discouraging for her.

*Shreya: When I'd asked him about like specific things, he'd just be like, "Oh, I think you should just review this topic" ...I think it definitely deterred me from asking more questions.*

As we can see from the experiences of the interviewed female students in physics courses, the discouraging and belittling comments from instructors made students feel less comfortable or outright uncomfortable to ask questions because they worried that their questions were not good enough and felt nervous about showing that they did not know something. Moreover, many female students reported that these discouraging behaviors of physics instructors are contagious. For example, Samantha mentioned that she noticed that after physics professors used "easy", "trivial", and "obvious" in their lectures, her male peers started to pick up these cues and used these terms.

*Samantha: "Oh, this is trivial, and you should know this right?" I did not encounter that phrase during high school. And then when I came to college, I heard professors start to use that. And then shortly after I heard the professors use that, I started hearing my [male] peers say that, and I feel like it's something that's almost like also taught along with the curriculum, [it] is like, this is how physics culture is.*

This experience of Samantha is consistent with our interviews with other female students, many of whom reported being also talked down and disparaged by their male peers during physics learning.

**(b) Belittle students' efforts**

This code refers to situations in which students described receiving negative recognition or lack of positive recognition for their efforts, improvement, and achievement. For example, Suzie shared the following example of her instructor giving negative recognition to students routinely:

*Suzie: The instructor usually begins the classroom by saying "You guys aren't serious about this class because you're not doing well on the work", and I'm just like you don't have to tell me that, you don't have to start every class with that, I don't want to hear.*

Suzie mentioned that these comments discouraged her and her peers from working hard in this course because their work was never recognized by the instructor. In addition to negative recognition, lack of positive recognition can also discourage students. For example, Maya noted that even small recognition or acknowledgement could encourage her; however, she never gets any when asking questions to her college physics instructors and TAs:

*Maya: [When asking questions to physics instructors or TAs], I usually clearly show that I'm reading the material and I understand what's going on, but there's a certain concept that I don't understand. However, I've never been told directly like, "Oh, you're doing a good job", or like, "No, you got it, you're just making a simple mistake" or something like that. So I guess I never really had the reinforcement, from a professor or TA of like, "You're being too hard on yourself".*

Lack of positive recognition or acknowledgement can make students worry about whether their work is good enough even though they might actually have done well, which may impact students' self-assessment and self-efficacy [25]. This is especially true for students who are underrepresented in physics, a field with strong stereotypes about who belongs and can excel in. Lack of positive recognition has negatively impacted the entry and retention in physics related disciplines for decades.

#### 4.5.1.2 Feeling Marginalized Due to Differential Gender Dynamics

In our analysis, 37% of the interviewed female students reported experiences that were coded under the subtheme *feeling marginalized due to differential gender dynamics*, which includes two codes: *differential treatment of female and male students* and *letting men dominate the class.*

**(a) Differential treatment of female and male students**

This code refers to situations in which instructors and/or TAs respond to or treat women and men differently when interacting with students. For example, Amy shared her experience in a college physics recitation:

*Amy: There have also been times, where I've interacted with, very sadly, a TA who has sort of brushed off my question and then answered a guy's question in my group or came into the room and only greeted the guys in my group and not greeted me.*

Amy further mentioned that even though she thinks this experience is not necessarily aggressive or outright offensive, it makes her feel that there is a difference in the treatment of men and women in this field. In our interviews, we found that the differential treatment of women and men also happens in other STEM courses and sometimes in STEM courses even before college. For example, Shally shared her experience of how her instructor in a coding course provides more detailed answers when answering men's questions:

*Shally: Sometimes when the girls mess up in the class, he's [the instructor] just like "okay, just share your screen, like I cannot understand what you're talking about" and then when it's the guys, he's like "oh we'll just do this, do this, do this, like..."*

Mila shared a similar experience in her high school chemistry class:

*Mila: When a guy is asking question to him [the instructor] he would be like "oh that's a really interesting question, let's look into it more". For mine, he'd be like "oh it's like we've been over this before" and it's like okay, just because my question is more simplistic doesn't mean that [it is not worth going over] ... you know what I mean?*

Mila also mentioned that this high school instructor "*sometimes would be too busy playing a chess game with one of the guys to like actually [bother to even] answer my questions"*. We can see from these examples from Amy, Shally and Mila that they felt that the instructors or TAs gave more detailed feedback and answers to male students than they gave to female students, which usually made the female students feel that their questions were not as important or good as male students'. In our interviews, we also found that sometimes female students felt that the instructors overexplained something to women but not to men during office hours, which was perceived by them as being due to the instructors' low expectations of them and thinking that they did not know much physics. This type of behavior can also discourage female students. For example, Lucy shared her experience in physics office hours:

*Lucy: I remember going to office hours with one of my female friends. And we were waiting outside. And then two of the guys that we knew from class came out, and then we went in and like, whenever we'd ask this professor question, he would just kind of like, work out the whole problem for us. Even though we just asked for like one little thing, we had done most of it ourselves already. When we asked for like one part of a five-part question, he went over the whole problem from the* *beginning. And then like we compared notes with the two guys who were there before, and they're like, "yeah, he didn't do that for us".*

Both Lucy and her female friends wondered whether it was because the instructor did not trust their answers to other parts of the question. The fact that the instructor did not answer male students' questions in this way made these female students feel that this instructor had lower expectations of women. In addition, the interviewed female students also reported that instructors sometimes used humiliating sexist language and explicitly showed biases toward women in class. For example, Evelyn shared her experience in a college physics class:

*Evelyn: There's like a group of girls that are also next to each other, they're really good friends. And he [the instructor] called on one of them. And she didn't know the answer. And he was like, "did you read the book?" And she said, "No, I haven't read yet" ... And then he was like, "what, so all of you are just in college for the social aspect?" And that was, everyone was just like, kind of like, visibly like, Oh, that was a really weird thing to say to a group of girls who were like friends…*

Evelyn also recalled that because it was only the second day of class, actually no one in the class had read the book yet by adding:

*Evelyn: [the instructor] was like suggesting that maybe like they [the group of female students] are only going to school because they want the image or that they have ulterior motives [for being in physics] or as though they're not really passionate, hardworking scientists, which they absolutely are…I think it's kind of just like ingrained sexism and like, misogyny… I think that there is some part of him that maybe thinks that women don't take it as seriously as men do. And he might not even be aware that he feels that way.*

Evelyn also mentioned that even though there had been times when she noticed that several physics professors' behaviors were not appropriate, she had never said anything to the professors or reported her experiences to superiors because she thought, *"what's it gonna matter?"*. In addition to Evelyn, other interviewed female students noted that they felt hesitant to communicate their feelings to the instructors or report their experiences because they worried that this will influence how the instructors think of them and even jeopardize their course grade. It is clear from the interviews that the lack of safe and effective ways for students to communicate their negative interactions with instructors and TAs and their overall learning experiences is itself very problematic.

**(b) Letting men dominate the class**

This code refers to situations in which the instructors and/or TAs let men dominate the class and the classroom dynamics dominated by men makes women feel marginalized. Elaine shared her experience in a physics professor's office hours, in which she felt being marginalized in this way.

*Elaine: I wasn't a very big fan of like, how he [the professor] did them [office hours]. Because if you'd like had a question, he gives you like, just like a tiny bit of information and wants you to just know what that means… or like, sometimes you'd ask a question, and he'd be like, aah, this person [a male student in the office hours] can explain that ... that was like, really embarrassing for me… it made me feel like lesser than people who were supposed to be my peers.*

Elaine mentioned that maybe the physics professor thought that having students explain physics to their peers can reinforce their learning, but to her, *"it felt kind of like, degrading, having a [male] peer explain something that I asked to a professor, you know? It felt like, my question wasn't important enough to be answered [by the professor]."* In addition, in the office hours, this professor usually gave *"just a tiny bit of information"* for Elaine's questions, while spending a lot of office hour time answering male students' questions even though many of their questions were not very relevant to that physics course:

*Elaine: I think another problem with the office hours is that there were also a lot of boys, they're trying to show off, and they asked like, very deep questions and stuff that wasn't like really important [for that physics course]… I showed up at office hours to get help with the homework that I needed help with. And they'd be asking totally random questions about like one very, very specific thing. I was just kind of annoyed that they were like talking about something so irrelevant when I needed help with something more tangible. But also, it kind of was annoying that the professor spent so much time answering their irrelevant questions when I had like, actual problems with the material we were learning in class.*

Elaine mentioned that this also happened in her physics TA's office hours and she sometimes ended up wasting a lot of time as in this situation described below:

*Elaine: [The TA's] office hours were like, two hours long. And we spent the first hour answering one guy's question that had nothing to do with what we were learning. So I just sat there for an hour, like trying to interject, but like, my TA, just kept going on and on about this, like, irrelevant subject. And I don't know, I just feel like it's, like, those questions shouldn't be prioritized [in office hours].*

As we can see from Elaine's experiences, the physics offices hours are sometimes dominated by male students. She felt that the professor and TA did not make efforts to balance the time they spent on each student and, on the contrary, they showed more interest in male students' questions, which made female students like her feel marginalized. Elaine noted that after these experiences, her interest in the course decreased and she stopped going to the office hours because "*It just felt like I wasn't supposed to be there. That wasn't the right course for me.*"

In addition to office hours, the interviewed female students also reported that they sometimes felt being talked down to or marginalized in group learning situations especially when they were the only female student in the group with several male students. Emma shared an example of her experience as the only woman in a physics learning group with three male students in an online collaboration over Zoom during the pandemic:

*Emma: We had this, like a weekly assignment, and I was the one like writing down the answers and turning in ... I was [the] only one that put my camera on and unmute. [When] I asked questions and then they wouldn't answer them, I would basically have to do it by myself. After two to three weeks, I was like "hey guys I've been writing down the answers for a while, I think it's someone else's turn to do it". So then, as soon as one of the guys started writing down the answers, the rest of the guys started participating…*

Emma further noted that she felt very sad and helpless, and attempted to find resolution by reporting these experiences to her male physics TA:

*Emma: I had like no patience left…it was bad and I started crying and I went to the TA, and I was like "I can't be in this group anymore, they don't listen to me, didn't even call me by my name, they called me 'the girl' to my face ... My name is on the [Zoom] screen, it's not that hard" …He [the TA] was just like "It doesn't matter, this is such a low stakes thing, it's just a participation grade. If you don't want to participate with them, just try to practice problems on your own, why are you worried about how your group was doing?".*

As we can see from Emma's experience, the male TA did not help her resolve this problem, and he did not even validate her feelings. Emma further noted that, for the rest of the semester, she just turned off her camera and microphone and worked on her own during the times assigned for group learning, in which students were supposed to co-construct knowledge in a collaborative environment. Emma's example is consistent with prior studies [75,138] showing that in a non-inclusive and inequitable learning environment, female students cannot fully benefit from interactive learning and gender performance gaps can grow.

In addition to being marginalized in physics learning groups, some female students also reported being marginalized in classes dominated by male students. Paola shared an example of her experience in a physics lab course, in which she was the only female student and all the other male students paired with each other and left her to do the experiments alone.

*Paola: "I was the only girl in that class, which is not ideal…people [the other students] kind of paired off, like, "Hey, do you want to do the photoelectric effect with me?" or "Do you want to do blackbody radiation with me?"… I ended up doing that lab by myself. The first few weeks were fine, because it was like programming and stuff, which I'm obviously comfortable with. And then for the first like, intensive lab that we had to do for that class, I didn't have a partner, so I was doing like the photoelectric experiments by myself, which was just a ton of data, a ton of lab write up. I knew, I can't do this anymore.*

In addition to having to do the lab alone, Paola also mentioned that in this lab class, "*people [male students] don't really listen to you. It's not necessarily intentional. But you suggest something, and people just ignore you…I didn't want to go to that class anymore…I didn't want to deal with the mental stress of that. Or, you know, pay money for a class I didn't want to go to*". Paola ended up withdrawing from the class after several weeks. Even though she told the lab instructor about her experiences before she withdrew, he did not try to improve the learning environment or make efforts to help her.

Michaela’s recollections of her physics classes, below, sum up much of what the other interviewed female student discussed in their narratives:

*Michaela: I never felt like they [her physics instructors] were encouraging women… The men in the class, were kind of like, already dominating conversations, they [instructors] would kind of answer those [men’s] questions and go on with whatever they [men] were thinking. They kind of focused on that as opposed to encouraging women to share their ideas and ask questions.*

#### 4.5.1.3 Feeling That the Physics Learning Environment Is Unsafe

In our analysis, 29% of the interviewed female students reported experiences that were coded under the subtheme *feeling that the physics learning environment is unsafe*, which includes two codes: *condescending or intimidating behavior* and *cold calling*.

**(a) Condescending or intimidating behavior**

This code refers to situations in which female students noted that the instructors and/or TAs displayed a feeling of patronizing superiority when answering student questions or caused them to feel fearful about not knowing, which can deter them from asking questions and communicating with instructors and TAs. Madalynn shared her fear of going to office hours for her physics courses:

*Madalynn: I feel like it's intimidating sometimes to go to professor's office hours. I don't want to go in there and look like a complete idiot… Sometimes, in office hours, you get a professor who's like “You're really asking me this question?” It's like, that's just what I want to avoid.*

Angelina shared an experience about feeling intimidated in a synchronous online class.

*Angelina: The teacher would be harsh when people asked “not good” questions… He’d be like “You don't understand what you're asking!”, then he might call you out for it, so quite a few* *people were like a little scared to ask questions, because if you don't ask the right question, you might get him annoyed.*

As we can see from Madalynn and Angelina's experiences, they were not only discouraged but also intimidated by their instructors. Angelina added that the instructor disabled the chat function, so students could not type their questions in chat. In addition to implicitly being harsh on students, the way instructors ask or answer questions can also be intimidating to students. For example, Mary shared how intimidated she was when she saw how her physics instructor answers students' questions in his office hours:

*Mary: He [the professor] had been talking about like, you should answer questions until you don't know…and then you should look for ways to figure out how. If you want, I can help you with that...And while I was there [in his office hours], this other kid in the class had come. [The professor] would basically ask him questions until he didn't know stuff… And I was like, intimidated by this other kid in there doing that, well, I had a really dumb question to ask...*

Even though Mary's instructor may want to use this Socratic method to encourage students to find out the answers by themselves, the way he did it during office hours might make students even more nervous about not knowing and feel unsafe to expose their struggles. This is particularly true for students from the marginalized groups in physics such as women.

In addition to feeling intimidated by instructors, the interviewed female students also reported that sometimes physics instructors may create a stressful learning environment in which top performance is over emphasized. For example, Evelyn shared an experience in one of her physics courses.

*Evelyn: He [the professor] would announce names and grades of high scorers [in front of the class], which I thought was kind of weird because it was like if you weren't in the top and then* *everyone knew what your score was almost… I don't know what the purpose of that was, maybe like to chase the glory of being, having your name called in class, which is just kind of weird to me…I think this is like a weird environment… [my score] shouldn't be anyone's business…I think everyone [other women she talks to] thought it was really weird.*

Evelyn further mentioned that announcing names of top students may promote an unhealthy competition. This may foster a fixed mindset in the classroom, especially when most of the top students are men, because due to societal stereotypes and biases, the words "brilliance" and "intelligence" are associated with men.

In addition, the interviewed female students also reported that some of their instructors were condescending when answering questions. Hailey shared her experience in her electricity and magnetism course:

*Hailey: I just did not feel comfortable with him [the instructor] at all. I didn't feel like I could ask him questions without fear of like being mocked or something like that. So, if I had any questions about that, I would talk to my friends… this professor [also] didn't mind mocking like humanities majors and classes and stuff like that, and saying like, "Oh, like they're not as hard as physics and stuff" …he could be pretty condescending even in class. I just didn't really want to be around it.*

Hailey noted that she did not think that talking to this professor would have helped her because, *"he wouldn't have listened…there just seemed to be no connection between him and the students… it didn't really seem like he had any concern about making a positive learning environment".* She mentioned that it is also true for many other physics instructors because *"They seem like set in their ways, and not really open to other opinions".* Hailey's experiences indicate that it is necessary to provide professional development and train physics instructors to help them realize their responsibility in building an inclusive and equitable learning environment and they need to make intentional efforts to build such an environment in which all students feel safe and can thrive.

**(b) Cold calling**

This code refers to situations in which women described instructors calling students to answer questions at random or asked them to share answers in front of their peers. Even though the instructors may think this is a good strategy to engage students, many interviewed female students noted that they did not like it and this approach made for a very stressful environment for them. For example, Evelyn shared how students like her felt when a physics instructor randomly called on students.

*Evelyn: " He [the instructor] had this Socratic method, which is like he kind of puts people on the spot. It is kind of stressful, because it feels like going in class… you're just like a sitting duck, waiting to be called out…I know a lot of people don't like that and think it's like a stressful environment.*

Maya shared a similar experience in which she and her female peers were concerned about being called randomly in a physics course:

*Maya: I've spoken with all the women who I sit with at the beginning of the semester. When we heard that he [the instructor] likes to call on people at random, there was one spot in like an initial thing where you wrote your name, year, background and stuff. And then [there was a question] "Do you have any concerns about this course?" All of us were like, can we write that we don't want to be called on at random? We literally talked about writing that in the concerns for the course.*

Maya mentioned that despite writing that, they still ended up being called randomly to answer questions in the course. She further noted that the reason they do not like cold calling is because they worry about getting the answers wrong in front of the whole class adding, "*but I guarantee our professor isn't even really thinking about it… They're just doing it*".

In addition to being randomly called in class, similar situations can arise in office hours. For example, Mary shared her experience in a physics instructor's office hours:

*Mary: His office hours were just largely very overwhelming to attend. He would ask you to write stuff on the board in front of the whole group. It was very intimidating, especially because I was there to ask for help. And so like, I didn't know how to do the stuff he was trying to ask us to do.*

Mary further noted that male peers were already showing off and she did not want her male peers to think she is not smart and *"I felt like I had to prove it to them that I was smart enough"*. As we can see from the interviewed women's experiences, cold calling may increase students' anxiety, particularly if they are from marginalized groups, which is consistent with prior studies [139,140]. Moreover, the anxiety caused by cold calling may increase stereotype threats for students from underrepresented groups such as women.

#### 4.5.1.4 Feeling Negatively Recognized about Their Abilities and Potential

In our analysis, 16% of the interviewed female students reported experiences that we coded under the subtheme *negative recognition for students' ability and potential*, which includes two codes: *underestimating students' ability* and *fixed mindset about students' potential* [114].

**(a) Underestimating students' abilities**

This code refers to situations in which instructors doubt students' ability to do well in a task or have low expectations of students. For example, Paola shared an experience of consulting with her physics department academic course advisor about the courses she wanted to take.

*Paola: I said [to the advisor], "I'm gonna take that computational physics class… it's like coding and I have taken the entire CS [computer science] core [courses], so that's fine. And I've programmed in Python for work". And he's like, "Well, it's really project heavy." … It just felt like he's just trying to dissuade me. He just kind of kept saying things like, well, it's really project heavy. And I was like, I'm taking CS 1501, which is like an algorithm's implementation course, which is like a weeder class for CS. I got a 90 for my midterm. It's really project heavy. I can do coding projects like I've been doing that for a while. I've had to code like three years ago, it's fine… He's like, "well, it's gonna be hard" and I was like, what? I know how to code. You can look at my transcript, I've taken more programming classes than probably most of the guys. So why are you saying this to me?*

As we can see from Paola's experience, the academic course advisor implied that the course she wanted to take was too difficult for her. Even though after Paola showed that she had enough coding experience to take this course, the advisor still doubted her ability, which made her frustrated. Paola further added that this actually was not the first time she had such an experience pertaining to physics related course work. Her previous academic course advisor in the first year before she declared the physis major (students usually declare physics major in their second year) also didn't trust that she could take an advanced calculus course, "*He was like, are you sure you should take calc III?*". Even though Paola took this course anyway, she said that she was *"annoyed"* by people who keep doubting her ability.

During the interviews, other female students also related similar experiences and they also reported examples of other female students going through similar experiences. For example, Evelyn shared an experience of one of her female peers who did undergraduate research that required extensive coding being dissuaded from taking the computational physics course with another core class in physics.

*Evelyn: My friend went to her advising appointment. She wanted to take the computational methods [in physics] that [is] like coding class, and her advisor was like telling her she's really going to struggle with it, like it's really difficult and like he doesn't think that she should take it with another core class… And she was really offended because…she was like, "why does he think that I'm not gonna be able to handle this coding class". And it's particularly strange because her [undergraduate] research is coding. She codes in Python every day for hours, and she's really, really good at it. So it was weird that he [the advisor] would kind of assume that she's gonna really struggle with this class and she couldn't handle the workload with other things. Particularly because other people like men were like, "he's [the advisor] never said anything about that to me" …that was really frustrating."*

Evelyn mentioned that witnessing this makes her think that the professor may have lower expectations of female students and added that "*Actually, it happened two semesters in a row. He did it again this semester. He was like, I think you should wait on that class, because coding is very difficult. And she's in it now and it's a breeze.*" Many interviewed female students mentioned that this type of experience makes them really want to prove to others that they deserve to be in physics, and they are as capable as men are, but this feeling sometimes caused extra pressure on them as female students in physics courses. These findings are consistent with prior studies showing that negative stereotypes about women in physics may cause women to assume that they have to make extra efforts to succeed in physics relative to male students and their achievements are not a reflection of how good they are in physics unlike the achievements of "successful" men who excel in these fields without making effort [25]. Likewise, women in physics courses might undergo additional stress and struggle to demonstrate their skills to be valued equally as men in a classroom in which they are underrepresented. These pressures may partially explain the finding in a prior study that female students are more likely to drop out of STEM majors such as physics than their male peers who earn the same grades [141].

In addition to lack of recognition and encouragement, the interviewed female students reported that sometimes micromanagement can also make them feel that physics professors do not trust their ability. For example, Mckinley shared her experience working as an undergraduate researcher with a physics professor in his lab.

*Mckinley: If I was asking him [the professor] a question on how to input a command into a Linux computer… He was like leaning over my shoulder and like telling me one key at a time what to type instead of just trusting me to be able to spell the word … It was weird being micromanaged in the places where I've felt pretty confident in myself. So it was just like a frustrating experience.*

Mckinley noted that on one hand this professor micromanaged her on the things that she was confident with, on the other hand, there was no guidance when she really needed help with the research. Finally, Mckinley left that lab because she did not think that she was given a project that played to her strengths. In addition, she added that even though that professor has once said that everybody can learn physics, she felt that *"he didn't actually put that into practice in his own lab, which was frustrating."*

**(b) Fixed mindset about students' potential**

This code refers to situations in which female students described instructors emphasizing brilliance rather than effort. For example, McKinley reported that some of her physics instructors think that only some people can do well in physics.

*Mckinley: And then there really are professors here who think that some people will just never be able to learn certain concepts.*

Prior studies have shown that instructors' mindset about whether all or only some of their students can excel in their courses can influence students' motivation and achievement, and underrepresented students are more likely to be demotivated and have negative experiences in classes taught by fixed mindset instructors [114].

### 4.5.2 Impact of Negative Perceived Recognition or Lack of Positive Recognition from Instructors/TAs on Students' Self-Efficacy and Interest

In the last section, we discussed different types of negative perceived recognition or lack of positive recognition from instructors and/or TAs. In our interviews, we found that these experiences of not being appropriately recognized often had a negative impact on female students' learning and physics motivational beliefs such as their self-efficacy and interest. In addition, the lack of appropriate recognition can influence female students' persistence and retention in physics and other STEM majors.

For example, Raina shared the feeling students might have when physics instructors assume that something is easy for students:

*Raina: The professors just expect you to completely understand it and so that can be very discouraging sometimes because, they'll just like go on and assume you know things…That can cause a lot of students to just be like I'm doing so poorly, like I am not confident and stuff like that.*

In addition to discouraging students, negative recognition can also cause students to doubt their ability to do well in physics. For example, Hailey shared her feelings when she received negative recognition from her physics instructor.

*Hailey: [I am] like having this professor who didn't seem [to think] my questions were valid…It just made me feel like stupid… and like other people in the class were smarter than me and like I wasn't capable of doing well.*

Evelyn had similar feelings when she was told by physics instructors that something she had struggled with was easy:

*Evelyn: They [physics instructors] say like, this is trivial, this is easy, this should be obvious. So then, if you don't think that, you immediately, you're like, oh, there's something wrong with me. I'm like, I'm missing this super obvious thing.*

As we can see from these examples, when physics instructors belittle students' difficulties, questions or struggles, students' self-efficacy can be negatively impacted, and they may even think "*there's something wrong with me*". The situation can be worse if the same instructors show more interest in and patience for men's questions because it can convey to women that they are not as capable as men. Amy shared her feelings after she was consistently ignored by her male TA, who *"brushed off"* her questions and answered the questions of male students in her group.

*Amy: …anytime that I made a small mistake, I felt like it was completely larger than it was, and I felt like I had made a grave mistake and I was not as smart as my classmates. Even if it was just typing something incorrectly into my calculator. Every little thing that I did incorrectly or* *every time that I wasn't listened to, I thought it was because of me, and because of my academic success and my personal knowledge…*

Amy added that these types of experiences made her feel that she did not belong in physics and also made her start to lose her interest in physics and her desire to major in engineering because physics courses are required for engineering students:

*Amy: They made me feel as if I didn't belong, which made me question, why am I here in the first place? Am I really interested in this? Do I really want to be here, meaning like in the engineering school…? And for me that [the experiences in the physics course] was a big reason why I was thinking of switching out of engineering…*

As we can see, not being appropriately recognized and validated by physics instructors or TAs can negatively impact student's self-efficacy and interest and also increase anxiety and stereotype threats of these students from marginalized groups. This can further influence their persistence and retention in physics and other related STEM majors. In addition, female students' self-efficacy and interest can also be impacted if instructors let the class be dominated by men and let women feel marginalized. For example, Katie shared her experience in a physics course.

*Katie: I guess in that class there were a lot of people who would just ask questions to show that they're smart like that they know what they're talking about. I was like, I don't even understand your question, I don't know what you're asking … They were all male, I don't even know how many girls were in that class, but it seemed like you only heard from the boys ever. It just kind of seems like they're [instructor and male students] like off in their own little world talking about something that the rest of us don't understand.*

Katie further added that the physics instructor talked very fast in class, and she had to re-watch the lecture videos and put it on half speed (these were online classes during the pandemic).

At the same time, the way some of the male students who dominated the class talked made her feel that the male students in general seemed to know what was going on in the class. Moreover, the professor gave very low scores on exams. Katie said that she put in a lot of effort in this course but still got 30% on a test and 60% after curve. These experiences strongly impact her interest and self-efficacy in physics.

*Katie: I hated physics. I just hated it so much. Anytime somebody would bring it up, I would like shudder… I cannot put into words how much I hated that class. I was like I can't do engineering, because I don't like this class … I really was like about to switch and do something else…I was like maybe this isn't for me, some people really seem to know what they're talking* about, *like maybe it's just a me problem.*

As we can see, both Amy and Katie considered switching out from an engineering major because of their negative experiences in physics courses. In addition, our interviews suggest that physics majors may also switch out of physics due to negative recognition. For example, Elaine shared many experiences pertaining to how she felt negatively recognized in physics courses, for example, how her questions were belittled by physics instructors in office hours and how the physics courses were dominated by men. She mentioned that the only reason she is still a physics major is that she has dreamed of becoming a physicist from age 12 and she wrote her dream down, keeps it with her and reads it whenever she feels discouraged and wants to quit.

*Elaine: I think what I'm trying to say is, I would not have been here still if I hadn't been so focused on my reasons for being here. Because there have been so many times where I've been like, I don't even know why I'm doing this…*

Elaine also mentioned that she had female peers switching out of the physics major because of lack of recognition and support.

*Elaine: I know four separate girls who were planning on majoring in physics and astronomy, but after the intro courses, they were too discouraged and switched out… which is really sad because I think they could have done it…it's hard, since there's not a lot of support.*

As we can see, negative recognition from instructors/TAs can impact students' self-efficacy and interest in physics, which have been shown to be very important for student engagement, performance, and retention in physics. Our interviews show that female students may switch out from their majors because of the discouraging experiences they had in physics courses. Therefore, instructors and TAs should realize that it is their responsibility to appropriately recognize and support students and make intentional efforts to build an inclusive and equitable learning environment in which all students can thrive.

### 4.5.3 Positive Recognition from Instructors/TAs and Its Influence on Students' Self-Efficacy

In previous sections, we discussed different types of negative perceived recognition/lack of positive recognition and their impact on female students' self-efficacy and interest. In our interviews, four female students shared their experiences of being positively recognized by physics instructors, which helped their learning and boosted their self-efficacy. In addition, some female students also shared their experiences in other courses (not physics) that made them feel positively recognized. In this section, we will discuss examples of how some instructors positively recognized students.

In our interviews, we found that explicitly recognizing and validating students' ability can help them build their physics self-efficacy. For example, Hailey shared a positive experience communicating with her quantum mechanics instructor:

*Hailey: My quantum professor, he's doing an excellent job at seeing what the students want…I feel like really supported in that environment, especially like, when I go to office hours… this professor literally said to me, like, "I think that you're just struggling with some of complex math stuff, but you're good on the concepts, you're doing well for where you should be". That was really encouraging because quantum is difficult for me at this time and it would definitely be easy to just feel overwhelmed…But the fact that this professor literally was like, "No, you are doing well, and you're capable of doing this.", it just makes me want to try so much more. And it makes me feel like I really am capable of doing it.*

Hailey further added that this kind of validation is not typical of what she is used to getting from physics instructors so it feels strange to her:

*Hailey: … I feel like supported in the class, [it] is just like so strange, but in the best way…I enjoy the class so much more and I feel really motivated to study especially compared to my other core physics classes…*

As we can see, positive recognition from instructors has the power to encourage and motivate students to work hard and boost their self-efficacy. Hailey used the word "strange" to describe her feeling of being supported by only one of her physics course instructors, i.e., such affirmation was not typical which is sad because this should be the situation in all of her physics courses.

In addition to explicitly recognizing students' ability, encouraging students to pursue their goal is also positively recognizing them. For example, Katie shared her experience of talking to a physics instructor when she felt overwhelmed:

*Kaylah: I really liked him as a professor…I was like, have a hard time getting concepts and stuff and I went to his office hours a couple of times at the end of the semester. I was like* *telling him how much I was kind of struggling or like worried about this class. And I told him, I wanted to be a physics and astronomy major. And he was like, "Oh, I don't think you should be discouraged by this one class. And if you really want to do this, you should keep trying". And so I [kept trying and] did pass that class…*

As we discussed in earlier sections, interviewed female students often reported that asking questions to some physics instructors could be quite intimidating. On the other hand, Evelyn's account below suggests that a physics instructor who acknowledges that struggles are normal may reduce students' fear of asking questions or making mistakes:

Evelyn: *When I was in his class, he told me he went to Harvard I think and he was like, "Yeah, I got like a 20% on my first test"… that just shocked me cuz I, I was like, wow… this person that I thought was just like this, you know, perfect academic like genius also struggled in the same way that I struggled and is now like a professor, you know, doing research… that just really altered my perspective of it. Because now it's easy to think of your professors as just like a flawless genius and they're just like constantly judging and grading you, but it's helpful when they open up and tell you what their experiences are and that your experience is normal, and you don't need to freak out about whatever you're thinking…*

In addition to physics instructors, the interviewed female students also shared experiences of feeling positively recognized by instructors of other courses. For example, Suzie shared her experience in her chemistry class:

*Suzie: I like to hear when my professor knows what students tend to struggle with and why they struggle with something. For example, if she sees something that in general students have trouble with, she'll say, students tend to have trouble with this part specifically, so focus more on this when you're studying. And I enjoyed that because it tells me specifically where I might struggle* *with and I don't feel as bad when I do struggle with that topic, because I know that other students in the past tend to struggle with it...*

As we can see, Evelyn's and Suzie's instructors help students to realize struggles are normal by sharing their own experience or sharing prior students' struggles. Sometimes, instructors can even help students by normalizing struggle by showing in a visual manner that many students have similar questions that they do not know the answers to:

*Raina: My chemistry teacher would always encourage us to ask questions…Like one time, when someone had a question and asked it, he stopped the class and was like, "How many of you are also thinking this question?" and like half the class raise their hands. So it was like half of the class had the same question, but everyone thought that they were like the only person that had the question, so it was like very helpful to like visually see how so many other people were also confused on the same topic like. I feel like that definitely made it very easy for me to just like realize, oh like I'm not the only one that's confused all the time.*

These examples show that after students realize that struggles are normal, they are not the only ones who are confused, and the difficulties can finally be overcome by working hard, they tend to be less nervous about not knowing or making mistakes. This can be helpful for developing their self-efficacy and interest. In addition, when students ask questions, instructors who recognize students' effort and ideas come across as more approachable for students. For example, Mila shared how her math instructor explicitly recognized students' questions and/or thoughts:

*Mila: …my math professor, he always says, "you know, that's a really good question let's look into it and kind of go through it" rather than just, "this is the answer"…And then he kind of makes you feel like, "Okay, thank you for thinking about this in a different way, we can answer* *this so you can understand it better", not just I'm answering your question to shut you up and get you out of the way.*

Helen shared a similar positive experience in her math methods in physics class, which influenced her self-efficacy in this class:

*Helen: The [math methods] professor is very open to questions, and he's really nice … he will not make you feel like stupid for asking the question or anything like that. And like, anytime you want to, like go up to his office and ask a question, he is happy to help. And that just like makes a difference in how well I feel like I can do in the class.*

The interviewed female students also mentioned that interpersonal communication with instructors and advisors is very helpful for their learning because their personal life and academic life are so deeply intertwined. For example, Charlize shared her experience with her academic advisor through her sorority.

*Charlize: She's really nice about like, sitting down, just because I had some classes that didn't go so well. And so she sat down with me… just to like plan out the rest of my steps to graduation and things like that… Every time we go, we print out another copy of the physics major requirements and figure out what I've done and what I haven't done and how to make my next semester, like, minimize the stress from classes and things like that.*

Charlize added that personal attention on specific issues she is having has been extremely beneficial to her:

*Charlize: That's been really helpful just because I feel like it's like, for me, it's really having her because she asks questions about my life and like I can tell her all like the issues that I'm having and things like that. And she'll listen and like talk to me and like, more like interpersonal interaction added to that has been a really helpful thing for me…*

As we can see from Charlize's experience, in addition to clearly communicating that difficulties are a normal part of learning and being open to questions, it is also important for advisors and instructors to provide students with appropriate guidance and support and let them know that they believe that they can do well, and they are always there to support them and brainstorm with them about the challenges they may be facing.

## 4.6 Summary and Discussion

There have been many efforts to understand the underrepresentation of women in STEM fields such as physics. Studies have shown that students' motivational beliefs such self-efficacy and interest can influence their persistence and retention in a field [20,37,41,44,45,48,63,73]. Prior quantitative studies have shown that female students usually reported lower physics self-efficacy and interest than their male peers [38,135,136]. In this study, we provide narratives from female students about their experiences in physics courses, which can help unravel some mechanisms for why they consistently report a lower level of motivational beliefs than men. In particular, we conducted interviews with female students enrolled in physics courses to investigate their experiences and perceived recognition from instructors and TAs and the relationship between perceived recognition and students' motivational beliefs such as self-efficacy and interest. Our results indicate pervasive existence of negative perceived recognition or lack of positive recognition from instructors/TAs for female students in the current learning environment. In particular, the negative perceived recognition or lack of positive recognition was categorized into four categories: (1) feeling belittled for questions or efforts, (2) feeling marginalized due to differential gender dynamics, (3) feeling that the physics learning environment is unsafe, and (4) feeling negatively recognized about their abilities and potential. We found that female students' perceived recognition from instructors/TAs often impacted their self-efficacy and interest and can influence their retention and persistence in physics and other STEM majors. Our findings suggest that it may be beneficial to do professional development for instructors and TAs to help them realize their responsibility to appropriately recognize and validate students and make intentional efforts to develop an equitable and inclusive learning environment in which all students can thrive.

We find that the most common negative perceived recognition from instructors and TAs is feeling belittled for questions or efforts. When students are confused about a new concept or have questions, the words "easy", "obvious" or "trivial" from instructors and TAs can convey to students that if they cannot do such easy problems on their own, they may not be smart enough to do physics. We find that this message from instructors and TAs can be conveyed not only via words, but also via tone and facial expression. This situation is exacerbated when the same instructors/TAs show interest in other students' questions (especially those from male students) because due to societal stereotypes and biases, the words "genius" and "brilliant" are associated with men, and thus the dichotomy in the instructor/TA responses to female and male students' questions can reinforce the negative recognition received by women [83]. Moreover, our interviews show that instructors' use of the words "easy", "obvious" or "trivial" for physics problems can be learned quickly and used abundantly by students (particularly male students) when communicating with other students and contributes to the toxic culture of physics. According to our interviews, belittling students' questions usually caused female students to feel nervous about not knowing physics concepts and made them feel less comfortable about asking more questions to the instructors/TAs to the extent that some of them completely stopped going to the office hours.

Our interviews also show that when female students made efforts or improved, lack of positive recognition and affirmation by instructors/TAs can also be discouraging. Students may sometimes worry about whether lack of positive recognition is because their work is not good enough even though they might actually have done well. Thus, a lack of positive recognition can impact students' self-assessment [25]. Moreover, we find that if instructors and TAs convey a doubt about female students' ability to do well in physics or convey a low expectation of them, students can also start doubting their ability even more and their persistence and retention in physics related majors may also be impacted. These findings are consistent with prior quantitative studies showing that women with A grades have the same self-efficacy level as men with C grades [38], and women quit the STEM majors with significantly higher grade point average than men and sometimes the women who quit on average have the same grade point average as their male peers who stay [141]. Eileen Pollack, the first woman to get a BS degree in physics at Yale University, has provided an excellent illustration of the effects of a lack of positive recognition in her memoir [142]. As a child, Pollack wanted to be a theoretical physicist, but after graduating, she eschewed her childhood dreams and decided instead to pursue graduate work in English. In her book, she recounts how she felt when she was dismissed by her instructors and even her undergraduate thesis adviser after solving a theoretical problem for her thesis: "When at last I found the answer, I knocked triumphantly at my adviser's door. Yet I don't remember him praising me in any way. I was dying to ask if my ability to solve the problem meant that I was good enough to make it as a theoretical physicist. But I knew that if I needed to ask, I wasn't." There are many Eileen Pollocks out there who will not narrate the impact of the negative recognition or lack of recognition on their career trajectories in memoirs.

In our interviews, we also find that perception of an unsafe learning environment can be perceived by female students in physics courses as negative recognition by instructors/TAs who are in powerful positions. For example, female students narrated that some instructors or TAs were “condescending” or “intimidating”, which usually caused them to be afraid of making mistakes and asking questions, which are antithetical to learning physics. Moreover, we found that cold calling can also make female students feel unsafe. When instructors or TAs put female students on the spot, they often feel very nervous and do not want to make mistakes in front of their male peers, which can cause anxiety and rob students of their cognitive resources (since working memory where information is processed while problem solving is limited and part of that working memory can be taken up by the anxiety) and influence students’ performance [5,75]. If the female students do make mistakes while being cold called or cannot answer the questions asked, this experience can make them feel frustrated and increase the stereotype threat which can further impact physics self-efficacy and interest.

During the interviews, female students noted that the physics courses and office hours were usually dominated by men, and some even reported being talked down by their male peers. Also, in most cases, instructors and TAs did not intervene to correct students’ behaviors and support female students, and sometimes instructors and TAs belittled the problems that these students communicated with them. As a result, these female students explicitly mentioned feeling negatively recognized by instructors/TAs in physics courses. A recent study [143] shows that in an equitable and inclusive physics department, when students fail to interact with each other equitably, faculty members intervene in the student-student interactions and insist on students’ behaving appropriately and learning the norms. This type of intervention by instructors and TAs in which they take responsibility to protect students from, e.g., sexist and racist microaggressions has the potential to create a more welcoming environment and ensure that all students, regardless of the demographic groups they belong to, can excel. Our interviews suggest that instructors and TAs usually did not intervene in the negative student-student interactions even when they were made aware of them and sometimes they themselves treated men and women differently, e.g., by displaying higher expectations of men, showing more interest in men's questions or paying more attention to men in general. These issues in the unsupportive learning environment dominated by men usually impacted female students' sense of belonging, self-efficacy and interest further.

We note that even though instructors and TAs sometimes negatively recognize or fail to recognize or validate students unconsciously, they still impact students' self-efficacy and interest. Therefore, it is important for instructors or TAs to realize that it is the impact on the students that matters and not their intention. Our study discussed four types of common negative perceived recognition or lack of positive recognition by female students from instructors and TAs, which can help physics educators to be intentional about not negatively recognizing students and focus on positively recognizing and validating students, which is particularly critical for students from marginalized groups such as women. Our findings can also help researchers to further understand the role played by perceived recognition in shaping students' self-efficacy and interest.

We also discussed examples of positive perceived recognition shared in the interviews that boosted female students' motivational beliefs, for example, situations in which they felt explicitly recognized for their effort and ability, encouraged to pursue challenging goals, and situations in which instructors/TAs were open to questions and interpersonal communication with students, acknowledged that difficulties are normal and surmountable, and were always there to support students in their learning. Prior studies have also shown that a social belonging intervention, which focuses on establishing a classroom climate in which adversity is framed as normal and likely to be overcome by working hard and working smart and taking advantage of all of the resources, can reduce gender differences in students' performance [144]. Based upon our interview findings and how much female students valued the fact that a few instructors attempted to normalize adversity, this type of intervention might also be helpful in improving female students' self-efficacy and interest.

Our findings suggest that it is important to make intentional efforts to appropriately recognize and validate students. 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 [145]. For example, instructors can explicitly recognize students by directly acknowledging their efforts and questions and expressing faith in their ability to excel. They can also implicitly recognize students by valuing their opinions and assigning a leadership position or a challenging task to students in small groups that makes them feel excited [145]. 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 can 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 [112]. We emphasize again that in any professional development workshops for instructors/TAs focusing on these issues, it is important for instructors/TAs to reflect upon and internalize that it is not their intentions that matter but the impact they are having on their students.

## 5.0 Do Female and Male Students' Physics Motivational Beliefs Change in a Two-Semester Introductory Physics Course Sequence?

### 5.1 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 [1,2,4,6,13,20,22,36,37,63,71]. 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 [20,37,41,44,45,48,63]. 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 [24,28,74-77]. 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 [21,22,58,63]. Students' identity in an academic domain, such as physics, is students' views about whether they see themselves as a "physics person" [21,22]. 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") [22]. 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 [78,79]. Moreover, they found that beliefs about performance and competence are actually not distinct and predict students' physics identity as a single construct [21,57]. 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. [81]. Based on the studies discussed above, other studies have been conducted to examine the impact of physics identity on students' career intentions [64,125] as well as other possible factors that could affect physics identity such as out-of-class science and engineering activities [125] and students' sense of belonging [90,146]. In addition, the identity framework has also been adapted to studies of math and engineering identities [124,147-149].

Competency belief is defined as the extent to which a person feels they have the necessary attributes to succeed [150]. 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 [151-153]. 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 [39]. 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 [94,154], 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 [20,36,37]. In particular, self-efficacy has been shown to influence students' engagement and performance in a given domain [41,44]. 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 [45]. In addition, studies show that students' self-efficacy predicts their career choices and persistence toward their short-term and long-term career goals [41].

Another component of identity is interest, which is defined by positive emotions accompanied by curiosity and engagement in a particular topic [47]. Interest has also been shown to influence students' learning [41,47,48]. 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 [52]. In addition, studies have shown that students' interest is not completely independent of self-efficacy [41,54]. According to Eccles's Expectancy-Value Theory (EVT) [53,54], 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 [155]) 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 [79]. 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 [58,81]. 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 [58].

However, many studies have shown that female students did not feel that they were recognized appropriately in science even before they entered college [66,83]. For example, a report of the National Science Foundation [84] 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 [84]. Studies show that these stereotypes and biases also exist in the university context [81,85]. 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 [85]. 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 [11,66,156]. 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 [38]. 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 [66].

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 US. Each semester in the US 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 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 [39,153,157]. 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 [21,81], in which the relationship between gender and identity is mediated through self-efficacy, interest, and perceived recognition. As shown in Figure 9 (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 (Figure 9 (b)), which is similar to the model in Kalender et al's prior work [81]. Due to 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 [158-160]. Therefore, in Model 2, self-efficacy is the predictor of interest, which is different from the model used in the prior work [81], 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.

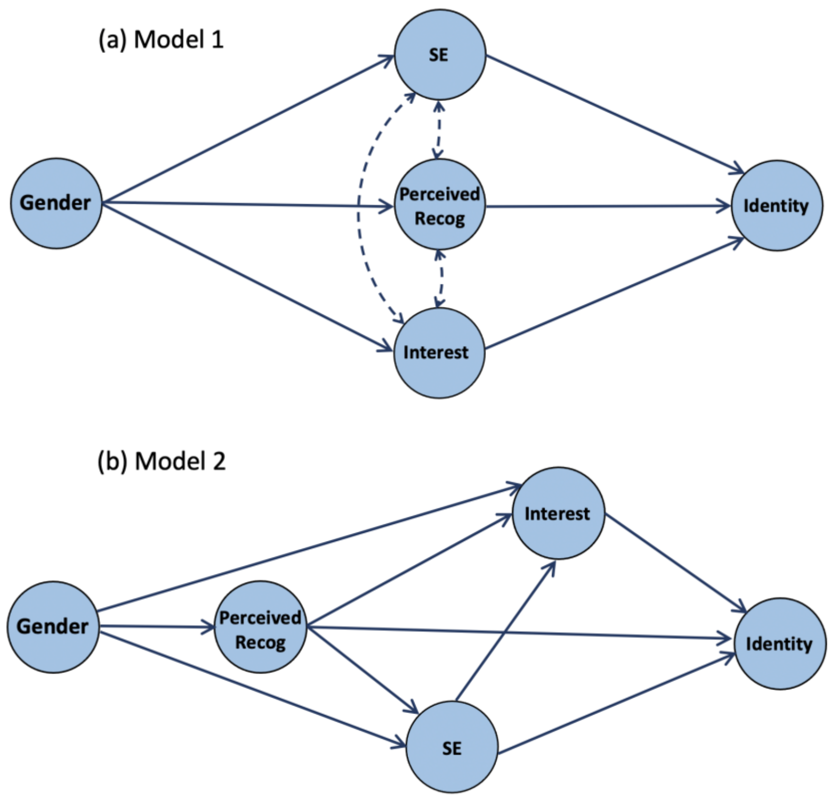



**Figure 9 Schematic representation of the path analysis part of the SEM models that shows how the relationship between gender and physics identity is mediated by perceived recognition (Recog), self-efficacy (SE) and interest. (a) In Model 1, all the three predictors are correlated with each other. (b) 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.**

## 5.2 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 US 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?

## 5.3 Methodology

### 5.3.1 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 US. 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 hours per week) with recitations (1 hour 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 [161]. Since the ICC values are significantly smaller than 0.1, the instructor level effects can be ignored [161]. 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 (IRB) 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.

### 5.3.2 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 10. The survey items were adapted from the existing motivational research [92-94,162-164] and have been revalidated in our prior work [37,95,96,127,165]. 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) [97], Pearson correlation between different constructs and Cronbach's alpha (which is a measure of the internal consistency of each construct with several items) [98-100]. As shown in Table 10, factor loadings (lambda) of each construct in physics 1 and physics 2 are very similar.

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 [82,92]. 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 [21]. Perceived recognition corresponds to whether a student thinks other people see them as a physics person [21,63], and it includes three items which correspond to family, friends and TA/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 [102].

**Table 10 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. †The response options for this question are "Never, Once a month, Once a week, Every day". ‡The response options for this question are "very boring, boring, interesting, very interesting".**

| Construct and Item | Lambda | |
|---|---|---|
| | 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 † | 0.676 | 0.687 |
| In general, I find physics ‡ | 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 |

### 5.3.3 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 [81], we checked the response option distances for our survey constructs by using item response theory (IRT) to support the use of means across ratings [105,166]. 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 [106,107]. 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 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 [105,107]. 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 [105].

Next, we calculated the Pearson correlation coefficients pairwise between the constructs for physics 1 and physics 2 separately. As shown in Table 11, 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 [109]. 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 Godwin et al.'s [58] and Kalender et al.'s [81] prior findings that perceived recognition (external identity) is the largest predictor of physics identity (internal identity).

**Table 11 Zeroth order correlation coefficients of the constructs in the mediation model.**

| Observed Variable | Physics 1 | | | | Physics 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | 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 | -- |

Finally, we used structural equation modeling (SEM) to analyze the predictive relationships among the constructs [87]. 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 [87]. 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 [98]. 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 $\beta$ for paths between each pair of constructs and the value of each $\beta$ 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 [91,167,168]. 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 [21,58,81,90]. According to Kline, a typical sample size in studies where SEM is used is about 200 [91], 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 [169]. As noted, our data fit the measurement model very well. Moreover, Table 10 shows that all factor loadings are higher than 0.5, which is considered as acceptable [169], 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 [108]. 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 [170]. In addition, we also performed the same

SEM models using diagonally weighted least square (DWLS) 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 Figure 9 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.

## 5.4 Results

### 5.4.1 Gender Difference in Motivational Beliefs

Table 12 shows the descriptive statistics for students' motivational beliefs at the end of physics 1 and physics 2. As shown in Table 12, 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 [104]. In particular, female students' average scores pertaining to perceived recognition and physics identity 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 [171] 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 B for the descriptive statistics. Appendix B 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 C, which show consistent results with the descriptive statistics shown in Table 12.

**Table 12 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* ~ 0.2 be considered a small effect size, *d* ~ 0.5 a medium effect size and *d* ~ 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 | | |

### 5.4.2 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 multi-group 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 Figure 10. A summary of all direct and indirect effects can be found in Table 13 (for physics 1) and Table 14 (for physics 2). The model fit indices suggest a good 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 Figure 10 or Table 13 and Table 14, 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 12 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 12. 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, Figure 10 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 Hazari et al.'s [21] and Kalender et al.'s prior work [81].

By comparing the SEM path analysis results of physics 1 (Figure 10 (a) or Table 13) and physics 2 (Figure 10 (b) or Table 14), 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 12. In addition, we 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 ($\beta = 0.29$).

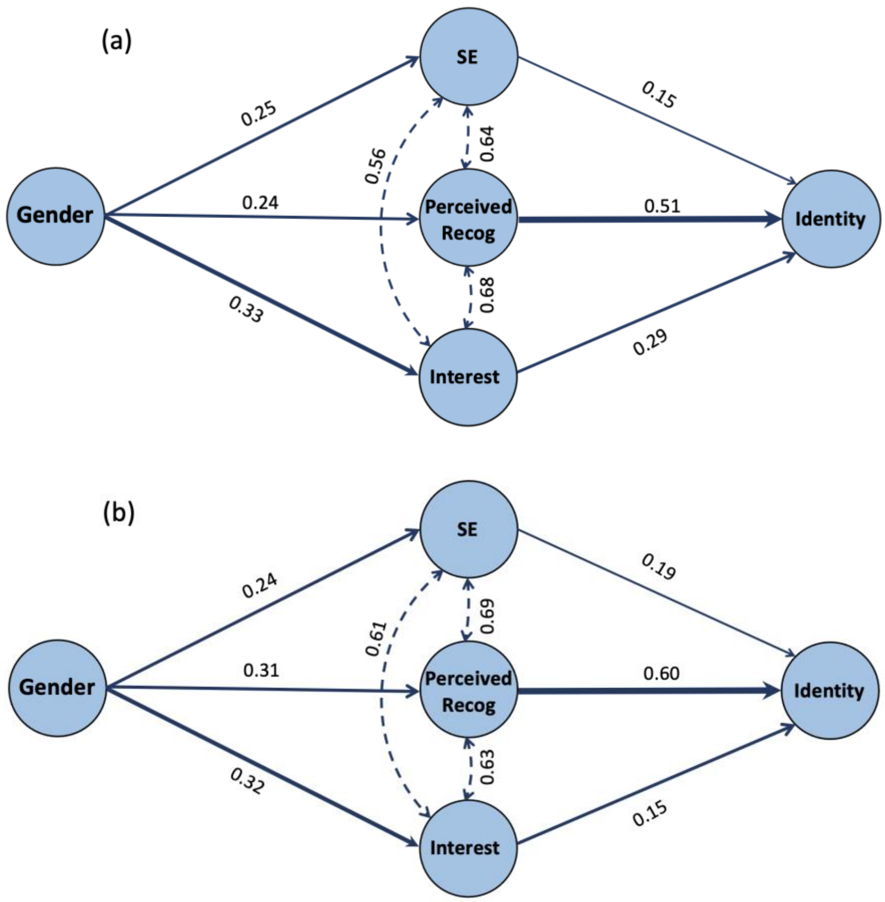



**Figure 10 Results of the path analysis part of SEM Model 1, in which there are only covariances between each pair of constructs: perceived recognition (Recog), self-efficacy (SE) and interest. (a) Shows the results of using physics 1 data, and (b) shows the results for physics 2 data. The dashed lines represent residual covariances between constructs. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients ($\beta$ values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the $\beta$ value. All $\beta$ values shown are significant with $p <$ 0.001.**

**Table 13 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 |

**Table 14 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 |

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 Figure 3. A summary of all direct and indirect effects can be found in Table 15 (for physics 1) and Table 16 (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 regression coefficients from gender to self-efficacy and interest.

By comparing the path analysis results of Model 2 in physics 1 (Figure 11 (a) or Table 15) and physics 2 (Figure 11 (b) or Table 16), 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 (Figure 10) 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, Figure 11 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.

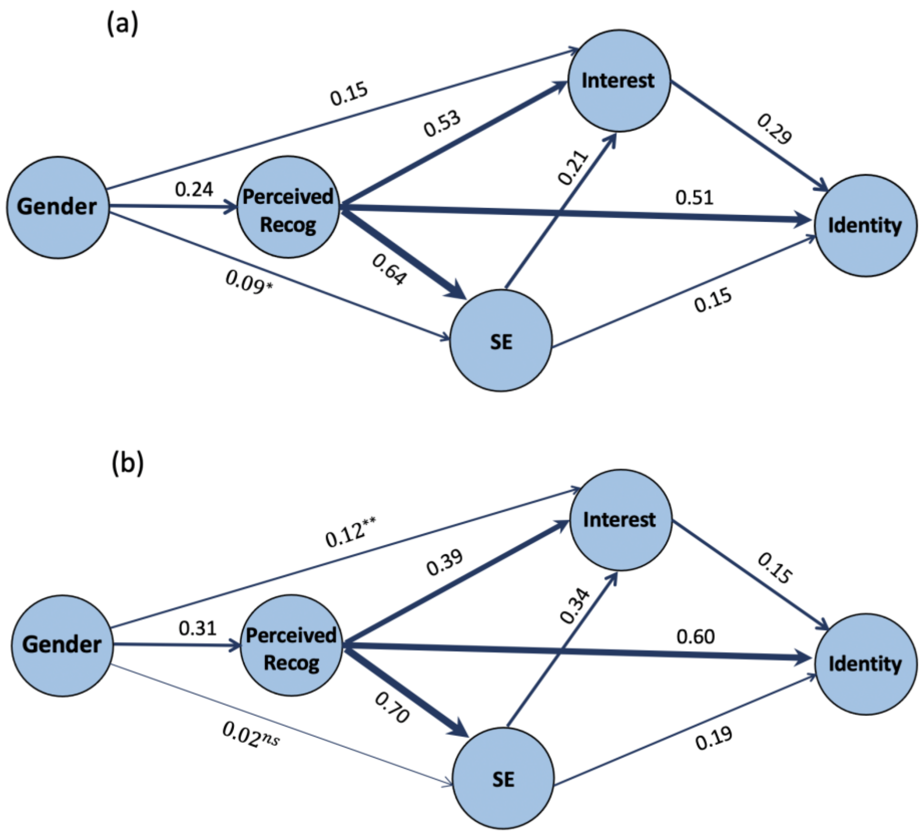



**Figure 11 Results of the path analysis part of SEM Model 2, in which perceived recognition (Recog) predicts self-efficacy (SE) and interest, and self-efficacy predicts interest. (a) Shows the results of using physics 1 data, and (b) shows the results for physics 2 data. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients ($\beta$ values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the $\beta$ with $0.001 \le p < 0.01$ indicated by **, $0.01 \le p < 0.05$ indicated by *, and $p \ge 0.05$ indicated by ns (not significant). All the other regression lines show relations with $p < 0.001$.**

**Table 15 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 16 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 |

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 17). 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.

**Table 17 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 (SE), perceived recognition (Recog), and interest. In Model 2, the arrows indicate the direction of the predictive relationships.**

| Models | Courses | Constructs | $R^2$ |
|---|---|---|---|
| Model 1<br>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<br>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 |

### 5.5 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 [21,81] and performed structural equation modeling (SEM) 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 12). 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/TAs sometimes showed more interest in male students' questions and answered male students' questions with more attention than when they answered their questions [71,76]. The interviewed female students also reported that men in their physics courses were generally praised more by the instructors/TAs than women, and sometimes instructors/TAs called men who answered the questions "brilliant", which made them feel as though they were not brilliant [71,76]. Due to 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 counsellors, 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 B), 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 [145]. 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 [145]. 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 [112]. 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 [172,173]. Research in physics education has shown that traditional lectures, even when perceived as good lectures, have limited success in helping students learn physics [174-176]. 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 [75,138], 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/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 [177-180]. Instructors can also tailor these short interventions in their classes to help all students develop positive motivational beliefs and learn physics equitably.

## 5.6 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 B 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 post- motivational beliefs in the second year studied in Appendix B 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 US, 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 US. More studies should be conducted in different types of institutions such as small colleges and universities in the US 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.

# 6.0 How Perception of Being Recognized or Not Recognized By Instructors as a "Physics Person" Impacts Male and Female Students' Self-Efficacy and Performance

## 6.1 Introduction

Physics has historically been portrayed as a field for brilliant men. Many prior studies have focused on the reasons for women's underrepresentation in physics and related disciplines from different perspectives and strategies to improve the learning environments so that all students can excel in physics courses [3,4,20-22,36-38,63,72,78,181,182]. Due to societal stereotypes, women often have significantly lower physics self-efficacy than men even when they perform similarly and many shy away from physics related majors and careers [20,36-38]. Moreover, being recognized by the instructor as a student who can excel in physics can be valuable for all students in physics courses. However, it is particularly important for underrepresented students including women and ethnic and racial minorities partly due to the societal stereotypes associated with who can excel in physics and lack of role models [3-5,20-22,36-38,63,72,75,78,81,181-185].

Perceptions regarding lack of positive recognition pertaining to whether a student can excel in physics and unintended belittling of students by instructors has a greater potential to negatively impact underrepresented students including women. For example, Eileen Pollock, the first woman to get a BS degree in physics at Yale University, decided to pursue graduate work in English despite finishing her physics undergraduate degree summa cum laude. In her memoir [142], she recounted the negative impact of not being positively recognized by her instructors, "By the start of my senior year, I was at the top of my class, with the most experience conducting research. But not a single professor asked me if I was going on to graduate school. When I mentioned shyly to

Professor Zeller that my dream was to apply to Princeton and become a theoretician, he shook his head and said that if you went to Princeton, you had better put your ego in your back pocket, because those guys were so brilliant and competitive that you would get that ego crushed, which made me feel as if I weren't brilliant or competitive enough to apply." Being a woman, it is not surprising that Pollock would interpret such statements to imply that she was essentially being told that she was not capable of excelling like the brilliant men at Princeton!

Pollock [142] also noted lack of positive recognition from her thesis advisor, "Not even the math professor who supervised my senior thesis urged me to go on for a Ph.D. I had spent nine months missing parties, skipping dinners and losing sleep, trying to figure out why waves — of sound, of light, of anything — travel in a spherical shell, like the skin of a balloon, in any odd-dimensional space, but like a solid bowling ball in any space of even dimension. When at last I found the answer, I knocked triumphantly at my adviser's door. Yet I don't remember him praising me in any way. I was dying to ask if my ability to solve the problem meant that I was good enough to make it as a theoretical physicist. But I knew that if I needed to ask, I wasn't." She added that she was "certain this meant I wasn't talented enough to succeed in physics, I left the rough draft of my senior thesis outside my adviser's door and slunk away in shame." This example illustrates another missed opportunity in which a thesis advisor failed to positively recognize her accomplishments and she went from feeling "triumphant" about having solved her thesis problem to feeling she wasn't talented enough to succeed in physics. What is also worth noting is that while writing her book, when Pollock asked her former advisor what he thought of her thesis, he stated that it was exceptional and when pressed further confessed that he had never encouraged *anyone* to pursue further studies.

Here we discuss a study on how the perception of being positively recognized or not recognized appropriately by the instructor or teaching assistant (TA) as a “physics person” or a person who can excel in physics impacts male and female students’ self-efficacy and performance at the end of a two-term college calculus-based introductory physics sequence. At the large public university where this study was conducted, Physics 1 involves mechanics and waves and physics 2 involves electricity and magnetism. A majority of students were first year engineering, physics, chemistry and math majors.

In addition to individual semi-structured think-aloud interviews with 30 student volunteers (20 women and 10 men), we used Structural Equation Modeling (SEM) with gender mediation. SEM is an approach involving simultaneous multiple regressions and can be used to predict relationships among different variables [87]. We focused on matched students who took both physics 1 and physics 2, i.e., there were 233 female and 464 male students followed from physics 1 to the end of physics 2. In four sections of the courses from which we present data, four instructors were involved who primarily used traditional lecture-based method. As shown in Fig. 1, we used this method [87] to investigate how male and female students’ self-efficacy at the end of physics 2 (post self-efficacy 2 or Post SE 2) and their physics 2 grade (Grade 2) are mediated by the perceived recognition as a “physics person” by the course instructor/TA. Fig. 1 shows that our model controls for student high school math SAT scores (SAT-Math), high school GPA (HS-GPA), physics 1 grade (Grade 1) and self-efficacy at the end of physics 1 (post self-efficacy 1 or Post SE 1). We collected data from a validated motivational survey [21,38,81,182,183] to measure students’ self-efficacy at the end of both physics 1 and 2 and students’ perceived recognition by the instructor and TA at the end of physics 2. However, details of the survey and measurement part of SEM will be described elsewhere.

**Table 18 Mean values of high school GPA and SAT math scores, college grades and post self-efficacy in physics 1 and 2, and perceived recognition in physics 2 by gender, along with p-values showing statistical significance of *t*-tests and effect sizes (given by Cohen's d with a positive value favoring male students) showing the strength of gender contrast [22]. Score ranges for all variables are also shown.**

| Predictors and Outcomes (Score Range) | Mean | | *p* value | Cohen's *d* |
|---|---|---|---|---|
| | Male | Female | | |
| High School GPA (0-5) | 4.20 | 4.34 | < 0.001 | -0.34 |
| SAT Math (400-800) | 713 | 706 | 0.130 | 0.13 |
| Post Self-Efficacy in Physics 1 (1-4) | 3.06 | 2.83 | < 0.001 | 0.47 |
| Physics 1 Grade (0-4) | 2.93 | 2.74 | 0.001 | 0.28 |
| Perceived Recognition (1-4) | 2.55 | 2.14 | < 0.001 | 0.55 |
| Post Self-Efficacy in Physics 2 (1-4) | 2.91 | 2.65 | < 0.001 | 0.47 |
| Physics 2 Grade (0-4) | 2.73 | 2.48 | < 0.001 | 0.31 |

## 6.2 Results

Table 18 shows the averages and effect sizes [87] for the differences between male and female students for their high school GPA and SAT math scores, their college physics 1 and physics 2 grades, their post self-efficacy at the end of physics 1 and physics 2 and their perceived recognition by TA/instructor in physics 2. Cohen suggested that typical values of d=0.2, 0.5 and

0.8 represent small, medium and large effect sizes, respectively [87]. Table 18 shows that female students had a higher average high school GPA than male students but lower average grades in both Physics 1 and 2 with small effect size. Table 18 also shows that there is a statistically significant gender gap in students' post self-efficacy scores both in physics 1 and physics 2 favoring male students with moderate effect sizes. In addition, Table 18 shows that female students also had lower perceived recognition by instructor/TA compared with male students with a moderate effect size.

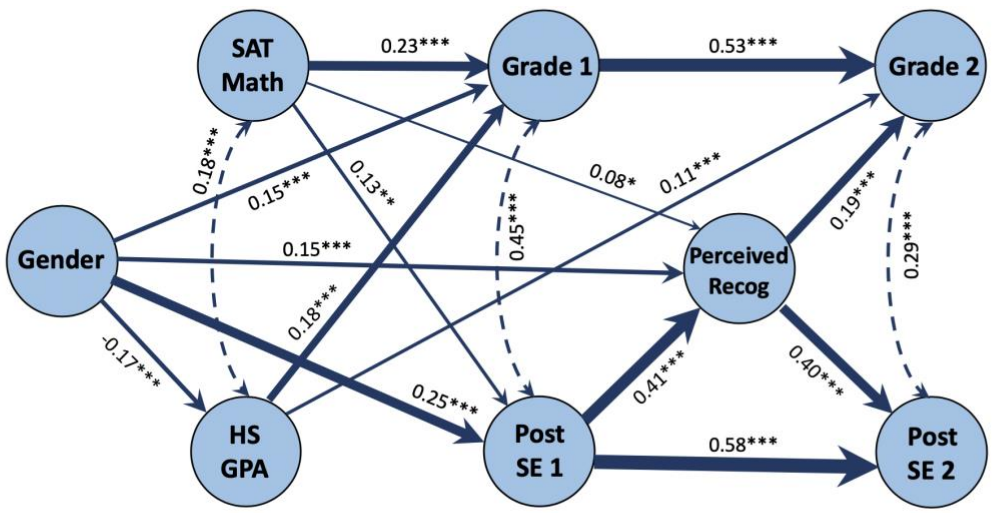



**Figure 12 Results of the structural equation modeling with gender mediation showing all of the paths connecting various predictor and outcome variables for which the regression coefficients β are statistically significant (*p*-values are indicated by *** for *p*<0.001, ** for *p*<0.01 and * for *p*<0.05 values) [22]. Each arrow with the line connecting two variables in the diagram indicates the direction of regression. Numbers shown with regressions connecting two variables are standardized values of β that can be compared with each other and thicker lines qualitatively signify stronger β. Each dashed line with double arrow connecting two variables indicates the correlation between them, and the number on the line is standardized covariance between them.**

Figure 12 depicts the path analysis part of the SEM with gender mediation [87] for students' self-efficacy and grade at the end of physics 2 and how these are mediated by the perceived recognition by the course instructor/TA, controlling for students' high school math SAT scores, high school GPA as well as their grade and self-efficacy at the end of physics 1. While details of SEM will be presented elsewhere, our SEM model fits are CFI = 0.945, TLI = 0.925, RMSEA = 0.061 and SRMR = 0.035, which represent good fits [87].

In Figure 12, the standardized regression coefficient β between each dependent (outcome) and independent (predictor) variable can be interpreted as the amount by which the dependent variable changes if the independent variable changes by one standard deviation [87]. We note that the β between perceived recognition by instructor/TA and student grade in physics 2 (Grade 2) is larger than the β between the high school GPA and Grade 2 suggesting that perceived recognition by their instructor/TA plays a more important role in predicting the course grade in physics 2 (see Figure 12). Similarly, the impact of students' perceived recognition by instructor/TA on their post self-efficacy in physics 2 (β for Perceived Recognition to Grade 2) is large ( $\beta = 0.40$), and is roughly as large as the $\beta = 0.58$ between post self-efficacy in physics 1 and physics 2 (see Figure 12). Thus, by recognizing students for their effort and progress when appropriate and being careful not to belittle them in unintended ways, instructors/TAs can impact their students' post self-efficacy and grade in physics 2 at the end of a two-term physics sequence in a positive way. In fact, instructor/TA's positive recognition for effort and progress at appropriate times can go a long way in supporting women's self-efficacy and identity and creating a learning environment in which all students can thrive [21,81].

We note that prior investigation suggests that women with an A grade have comparable physics self-efficacy to men with a C grade in introductory physics [38] and this difference in self-efficacy has the potential to disproportionately drive out women from physics related majors and careers. In particular, the perception of not feeling positively recognized by the instructors/TAs as a person who can excel in physics has the potential to not only deteriorate students' self-efficacy and performance in a physics course but can also have a long term negative impact. While the details will be presented elsewhere, we note that even in individual interviews, we find that women are less likely than men to feel positively recognized by instructors/TAs (similar to Table 18), and it has the potential to disrupt their paths away from physics related majors and/or careers. In the individual interviews we conducted with students, women sometimes reported that they had contemplated switching out of their majors (either engineering or physics) because of negative experiences in their physics courses while men did not express similar concerns. The interviewed women sometimes noted that men in their physics courses were generally praised more by the instructor/TA than women and men often dominated asking questions and answering most of the questions instructors asked. Moreover, some women noted that instructors often did not make an effort to improve their sense of belonging, e.g., by ensuring that everyone was given an opportunity to contemplate the answers to the questions asked in small groups and then each time a different group was asked to answer the questions. Some interviewed women reported that sometimes instructors/TAs called men who answered the questions "brilliant", which made them feel as though they were not brilliant. They also noted that when they went to the course instructor or the TA to ask for help on physics problems, they were told that the problems were "easy", "obvious" or "trivial", which they perceived as disparaging or belittling in that they felt they were being told that they are not smart enough to do physics if they could not do such easy problems on their own. Some of them reported never going back to their instructor or TA after that for help.

### 6.3 Discussion and Summary

Women are severely underrepresented in physics and related disciplines in which there are pervasive gender stereotypes about who can excel and systemic disadvantages to women from these stereotypes accrue from a young age. We presented both quantitative and qualitative evidence focusing on the impact of male and female students' perceived recognition by the instructor and TA on their physics self-efficacy and performance at the end of a two-term calculus-based introductory physics sequence. We find gender differences in students' perceptions of being recognized.

Positively recognizing students who are underrepresented in physics is particularly important because lack of positive recognition has negatively impacted their entry and retention in physics related disciplines for decades. We strongly recommend that physics instructors and advisors have high expectations of all students. However, they should not be parsimonious in providing micro-affirmations to students whenever an opportunity arises and praise them for their effort instead of brilliance or intelligence [186]. Building an inclusive and welcoming environment, being empathetic, learning students' names as well as offering mentoring, guidance and support for next steps that focus on individual student's strengths, interests and growth can help in improving students' perception of feeling recognized [186]. Thus, instructors should set high standards but make it clear to students that they know they can achieve those high goals by working hard and working smart [187]. This is important for students to feel that the instructor was applying high standards to them and not just giving empty praise [187].

Also, calling problems that students are struggling with "trivial" or "obvious" is likely to have a student feel disparaged even if no harm was intended. What instructors and advisors must realize is that what is important is the impact on students of what they say and not whether harm was intended! Instructors should realize that labeling a question that a student is asking “trivial” has the potential to create stereotype threat [5,185] for students who are underrepresented in physics, increase their anxiety and decrease their physics self-efficacy.

We also encourage established members of the field to tackle head-on any issues of bias or microaggression that they observe in their classes rather than staying silent. Otherwise, due to societal stereotypes about physics, lack of perceived recognition as someone who can excel in physics has the potential to most hurt the underrepresented students, e.g., women who have few role models. They often enter physics courses with lower self-efficacy and doubts about whether they have what it takes to excel in physics courses. Our findings suggest that without explicit thought and action by instructors and TAs to appropriately recognize students as people who can excel in physics, not only is the gender gap in a physics course grade likely to be maintained but the self-efficacy gender gap is also likely to persist. This self-efficacy gender gap has the potential to not only have short term negative impact on women in physics courses but can also have long term negative impact, e.g., on their career choices.

# 7.0 How Female and Male Students' Sense of Belonging in Introductory Physics Class Predicts Their Academic Performance

## 7.1 Introduction and Theoretical Framework

Prior studies have shown that women are often underrepresented in many science, technology, engineering, and mathematics (STEM) courses and disciplines [2-4,22,23,36,38,63,185]. For example, even though women earn approximately 60% of all bachelor's degrees in the US, only 20% of the physics undergraduate degrees are earned by women [13]. In addition, some studies have reported gender disparity in students' performance in STEM disciplines [14,15,188]. Prior research also suggests that individuals' course enrollment and performance in STEM can be influenced by their motivational beliefs such as their sense of belonging in that domain [21,22,37,43,189-192]. For students from underrepresented groups, sense of belonging might be undermined due to negative stereotypes about who can excel in certain STEM fields and lack of encouragement and role models, leading to withdrawal from STEM fields [24,25,28,29,32,74,193]. Hence, investigating students' sense of belonging is critical to understanding and addressing diversity, equity, and inclusion issues in STEM disciplines.

By equity in learning, we mean that not only should all students have equitable opportunities and access to learning resources, but they should also have an equitable and inclusive learning environment with appropriate support and mentoring so that they can engage in learning in a meaningful and enjoyable manner and the learning outcomes should also be equitable. By equitable learning outcomes, we mean that students from all demographic groups (e.g., regardless of their gender identity or race/ethnicity) who have the pre-requisites to enroll in the course have comparable learning outcomes, which is consistent with Rodrigues et al.'s equity of parity model [33]. The STEM learning outcomes not only include students' academic performance but also include their motivational beliefs in the domain such as student sense of belonging. If students do not believe that they belong in a science or engineering course, they are unlikely to fully engage in learning and are less likely to be enthusiastic about future studies and careers related to these fields [194]. Thus, investigation of students' sense of belonging is important to understanding their engagement, performance, and persistence, and can provide guidelines for developing an inclusive and equitable learning environment in which students from all demographic groups can excel.

### 7.1.1 Prior Studies on Students' Sense of Belonging

Students' sense of belonging refers to whether students feel that they belong within a particular community [195]. The feeling of belongingness is related to being valued, respected, and accepted in academic spaces [195]. Prior studies have shown that students' sense of belonging can influence their persistence and retention in a domain. For example, one study showed that even after controlling for other known correlates of persistence (e.g., financial difficulties, standardized test performance), students who reported higher sense of belonging also reported stronger intentions to persist at the university [196]. Furthermore, research has suggested that, in STEM settings such as physics, students from underrepresented groups, e.g., women in some of the disciplines, often experience threats to their belonging, and they are also more likely to be impacted negatively by lack of a sense of belonging [197,198]. For example, a study shows that women who reported lower sense of belonging specifically in their STEM major (e.g., physics) were more likely to expect to switch out of their major [199]. Seymour and Hewitt conducted a study to investigate why undergraduate students leave science, and they found that a common reason for women's attrition was feeling less welcome and accepted in the traditional masculine culture in STEM [200]. Their work also suggests that this feeling of being an outsider is not only related to the fact that female students often have few same-sex peers in STEM fields, but it is also affected by the traditional "weed-out system" of many STEM fields (e.g., physics) that encourages behaviors such as self-promotion and competitiveness, which are often in conflict with traditional female gender roles [200].

In addition, students' sense of belonging has also been shown to predict other motivational beliefs. For instance, Freeman and colleagues [201] found that students' sense of belonging in a specific college class was positively associated with their self-efficacy, task utility, and intrinsic motivation. In physics, prior studies have shown that students' sense of belonging predicts their physics identity and perceived utility value of academic tasks [90,202]. Prior studies also show possible association between sense of belonging and students' academic outcomes [203-206]. For example, Pittman and Richmond [207] found that students who reported a higher sense of belonging were not only more motivated than peers who were low in belonging, but they also reported higher grades. Regression analysis can help us understand how strong this association is in different disciplines in different contexts (e.g., in traditionally taught courses vs. when explicit effort is made in the courses to increase the sense of belonging of students, particularly for those who are from marginalized backgrounds such as women in physics). It is important for instructors to know the factors (e.g., sense of belonging which has been found to be important in other contexts and can be measured using validated surveys in different types of courses) that can influence students' learning outcomes and to what extent they influence the outcomes because this can help instructors make intentional effort to improve their courses for all students. Several studies have used multiple regression analysis or structural equation modeling to study the direct effect of sense of belonging on students' overall grade [208,209] or grade in a specific course such as math [146]. However, very few studies have been done to study the effect of students' sense of belonging on learning outcomes (such as conceptual understanding and quantitative problem solving skills) at the end of a physics course. Even though this association may be mediated by other variables (e.g., students' motivation to learn, self-regulation, engagement in class, initiative, interaction with peers and instructors, the level of anxiety, etc.), it is useful to first study the total effect of sense of belonging on students' learning outcomes.

### 7.1.2 Gender Difference on Concept Inventories in Physics

Even though teaching students to solve quantitative physics problems is often a central objective in traditional physics instruction, students' conceptual understanding of the physics content knowledge is also a very important learning outcome [210]. However, prior studies have shown that female students often have lower average score than male students on physics concept inventories [211-215]. Moreover, a prior study investigating students' performance on physics conceptual and quantitative tests showed that women had significantly lower conceptual scores than men, while men and women performed equally on the quantitative test questions [216]. Several factors have been proposed to explain the gender difference in students' performance on these physics concept inventories. For example, studies have shown that differential background and preparation of men and women can account for part of the gender gaps [217,218]. In addition, prior studies suggested that more interactive teaching methods are beneficial for all students [219,220] and may also help reduce the gender gaps in students' conceptual understanding [2,217,221]; however, this effect has not been consistently reproduced in other studies [14,181,222]. In particular, a study shows that in an inequitable and non-inclusive learning environment, female students may benefit less than male students from interactive learning because they may not feel safe to express themselves in collaborative group discussions, and the gender gap was even larger at the end of the course than that in a traditional lecture-based course even though all students benefited compared to a traditionally taught course [138]. Prior studies show that factors such as sense of belonging can also contribute to the gender difference in students' performance in physics [180,223]. However, there are very few studies focusing on the effect of students' sense of belonging on their conceptual understanding and quantitative problem solving in a physics course.

### 7.1.3 The Present Study and Theoretical Model

Inspired by the above studies, we conducted a study focusing on how students' sense of belonging predicts students' conceptual understanding and course grades at the end of a college introductory physics course after controlling for gender, high school grade point average (GPA), Scholastic Aptitude Test (SAT) math scores, and their sense of belonging and conceptual understanding at the beginning of the course. In this study, students' conceptual understanding was measured by the Force Concept Inventory (FCI), which is one of the most commonly used concept inventories in physics for introductory mechanics [224]. This study was conducted in a traditionally taught course, in which students' course grades are largely based on their performance on midterm and final exams, which mainly focus on quantitative problem solving. Moreover, the quantitative exams are spread out throughout the course and average over student performance on all the topics covered in this course. Therefore, course grade is also a good measure of students' quantitative understanding. We also investigated how students' sense of belonging and FCI scores changed from the beginning to the end of the course and whether there were gender differences in the constructs studied. As shown in Figure 13, high school GPA and SAT math scores are the two high school constructs, and there is a covariance between them. Students' pre-sense of belonging and pre-FCI were measured at the beginning of the course, and their post-sense of belonging and post-FCI were measured at the end of the course. There is a regression path from sense of belonging to FCI for both pre and post, which helps us study how students' sense of belonging predicts their performance on the FCI test. Grade is a learning outcome predicted by the other constructs. Gender is not predicted by any construct, so there is no path pointing to it. In Figure 13, each construct can be predicted by all of the constructs on its left. From left to right, all possible paths were considered in the structural equation modeling (SEM) model, but only some of the paths are shown for clarity in Figure 13.

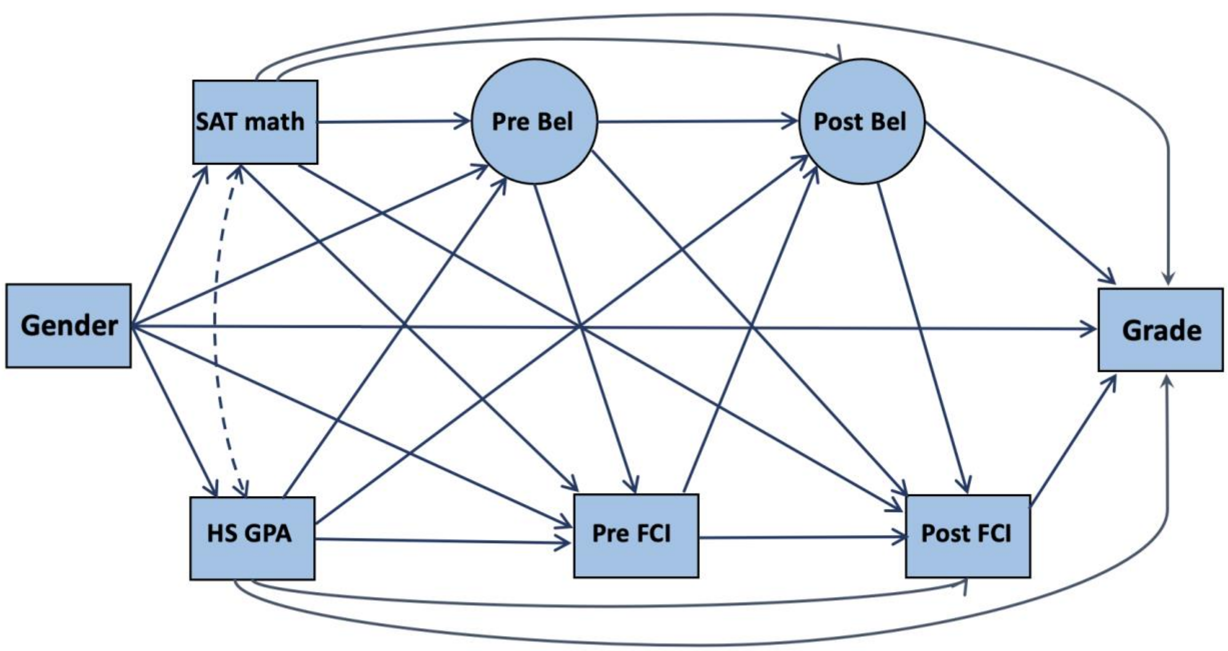



**Figure 13 Schematic representation of the theoretical framework. From left to right, all possible regression paths were considered, but only some of the paths are shown here for clarity. HS GPA represents high school grade point average. Bel represents sense of belonging.**

## 7.2 Research Questions

Our research questions regarding students' sense of belonging and academic performance in a calculus-based introductory physics course at a large state-related university in the US are as follows. This course is generally mandatory and taken by engineering, physical science and mathematics majors in the first semester of their first year of undergraduate studies.

**RQ1.** Are there gender differences in students' sense of belonging and academic outcomes in an introductory physics course, and how does students' sense of belonging predict their academic outcomes after controlling for students' gender, high school GPA, SAT math scores, and their sense of belonging and FCI scores at the beginning of the course?

**RQ2.** Does gender moderate the relationship between any two constructs in the model? In other words, does the strength of relationship given by the standardized regression coefficients between any two constructs in the model differ for women and men?

## 7.3 Methods

### 7.3.1 Participants

The data used in this study were collected from students enrolled in a first-semester college calculus-based introductory physics course at a large research university in the US. This course is taken primarily by students majoring in engineering, physical sciences, and mathematics for whom it is mandatory. This course is a traditional lecture-based course (4 hours per week) with recitations (1 hour per week), in which students typically work on physics problems with the help of a teaching assistant (TA). This course mainly includes mechanics topics such as kinematics, forces, energy and work, rotational motion, gravitation and oscillations and waves. There were 533 students (211 female students and 322 male students) who participated in the study at both the beginning and end of the course (matched students from pre to post). In our study, most students were 18 or 19 years old, and the average age was 18.7 years old. Most students were first year students, who had just graduated from high school. Students were predominantly White (74%), with the remaining students coming from a number of other ethnic or racial backgrounds: Asian (12%), Hispanic (5%), African American (4%), Multiracial (4%) and Others (1%). Students' demographic data were provided by the university. Students' names and IDs were de-identified by an honest broker who provided each student with a unique new ID. Thus, researchers could analyze students' data without having access to students' identifying information.

### 7.3.2 Measures

In this study, we investigate how students' sense of belonging predicts their academic outcomes (measured by FCI scores and course grades) after controlling for students' gender, high school GPA, SAT math scores, and their sense of belonging and FCI scores at the beginning of the course. Students' pre-college scores including high school grade point average (GPA) and Scholastic Assessment Test (SAT) math scores were provided by the university, which are widely used for college admissions in the United States. Students' course grades are largely based on their performance on midterm and final exams, which mainly focus on quantitative problem solving. The course grades were obtained from the university records. The conversion between the letter grade and corresponding grade point is given in Table 19. Students' conceptual understanding and their sense of belonging were measured using Force Concept Inventory (FCI) and a validated survey, respectively. Both the conceptual test and survey were administered to students in the first and last recitation class of the semester.

**Table 19 Letter grades and corresponding grade points.**

| | F | D- | D | D+ | C- | C | C+ | B- | B | B+ | A- | A/A+ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade Point | 0 | 0.75 | 1.00 | 1.25 | 1.75 | 2.00 | 2.25 | 2.75 | 3.00 | 3.25 | 3.75 | 4 |
| Definitions | Failure | | | | | Minimum level to pass | | | | | | Superior attainment |

### 7.3.2.1 Conceptual Test

The Force Concept Inventory (FCI) which consists of 30 multiple choice questions was administered to measure students' conceptual understanding of Newtonian mechanics [224], in contrast to their ability to solve quantitative problems that are typically used in regular course exams (and which can sometimes be solved algorithmically without conceptual understanding of the underlying concepts). The Force Concept Inventory (FCI) is one of the most commonly used multiple choice surveys for assessing students' conceptual understanding of Newtonian mechanics [225]. FCI has been validated with extensive interviews and by comparison with the former mechanics diagnostic test [224,226]. Face and content validity of FCI was established through the support of the numerous physics instructors who have used the test and who have generally agreed that the test measures students' understanding of the force concept [224,226]. In addition, FCI correlates well with other kinematics and forces related assessments such as force and motion conceptual evaluation (FMCE) [227,228]. Reliability of FCI has been well established through extensive use of the tests. Hake's large survey data give convincing support for reliability and similar pre- and post-test scores have been found in the USA in many institutions where the style of instruction has been similar [229]. Moreover, FCI shows global test-retest stability [230] and low global context dependence [231].

Moreover, it is well established that student knowledge structures are often loosely organized, incoherent, and context-dependent [232-234]. Thus, FCI is not going to show that students' knowledge structure is coherent if factor analysis of FCI concepts is conducted based upon cognitive task analysis from an expert perspective. In particular, based upon the fact that student knowledge is context dependent, the FCI results will show the structure and organization of students' context-dependent knowledge [235]. In particular, FCI provides a standard against which students' context-dependent conceptual understanding can be compared in detail [235]. Students' performance on FCI can show what aspects of Newtonian force students approach correctly and where and how the students use their common-sense beliefs. In other words, the FCI score is a measure of disparity between student concepts and the Newtonian force concepts [235]. Moreover, prior studies show that students with low FCI scores usually have fragment and incoherent concepts about force and motion, while students scoring 85% or higher on FCI exhibit a fair degree of coherence in their conception of force and motion [236,237], which are consistent with Vosniadou's study showing that pre-Newtonian thinkers have conceptions that are not fully coherent or consistent [238]. Therefore, FCI is useful to measure the degree to which the student has assimilated the Newtonian mechanics concepts.

#### 7.3.2.2 Survey Instruments with belonging items

In this study, we used a validated survey to measure students' sense of belonging at the beginning (pre) and end (post) of the course. The survey questions for sense of belonging are listed in Table 20. These survey questions were adapted from existing motivational research [164] and were re-validated in our prior work [37,95,96,127,165]. As shown in Table 20, for both pre and post, all of the CFA item loadings are above 0.6 and most of them are above 0.7, which means that our construct extracts sufficient variance from the items [239].

Students' sense of belonging pertains to their feelings of whether they belonged in the physics class [195]. We measured it using four items that were scored on a 5-point Likert scale: "not at all true, a little true, somewhat true, mostly true and completely true" (Cronbach's alpha = 0.81 for pre and Cronbach's alpha = 0.86 for post). Two items pertaining to sense of belonging ("I feel like an outsider in this class" and "Sometimes I worry that I do not belong in this physics class") were reverse coded, which means that a higher score in these two items represents a lower sense of belonging. Students' sense of belonging score is the average score of all items related to this construct.

**Table 20 Survey questions for sense of belonging, along with CFA factor loadings for students' pre and post responses to the survey. Lambda (factor loading) represents the correlation between each item and the 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.**

| No. | Survey Items for Physics Sense of Belonging | Lambda | |
|---|---|---|---|
| | | Pre | Post |
| 1 | I feel like I belong in this class. | 0.704 | 0.832 |
| 2 | I feel like an outsider in this class. | 0.704 | 0.733 |
| 3 | I feel comfortable in this class. | 0.770 | 0.773 |
| 4 | Sometimes I worry that I do not belong in this physics class. | 0.689 | 0.782 |

### 7.3.3 Data Analysis

In this study, we first used a *t*-test [103,240] to compare students' sense of belonging and FCI scores at the beginning and end of the course as well as conducted an analysis of gender differences using descriptive statistics. Then, we used structural equation modeling (SEM) [87] to study how students' sense of belonging predicts their FCI scores (conceptual understanding) and course grades (quantitative understanding) at the end of the course after controlling for students' gender, high school GPA and SAT math scores as well as their sense of belonging and FCI scores at the beginning of the course.

Before performing the SEM, we calculated the Pearson correlation coefficients pairwise between the constructs [100]. The results show that even though these constructs have correlations with each other, the correlations are not so high that the constructs could not be examined as separate constructs [109] (see Appendix D for detailed results). To analyze the regression relationships among the constructs, we performed integrated SEM analysis. First, we evaluated the measurement model for pre- and post-sense of belonging using confirmatory factor analysis (CFA) and obtained factor scores for pre- and post-sense of belonging [241,242]. Then, we performed a path analysis using these factor scores. The two-steps SEM is implemented because, while not strictly a SEM, which would require simultaneous estimation of factor loadings and regression coefficients, the factor scores approach enhances stability of the solution avoiding the effects of misspecifications that always happen in complex models with multiple predictors, mediators and dependent variables as in our case [241,243]. Even though FCI also includes 30 items, we use FCI total score rather than factor score to conduct the path analysis. This is because FCI includes many conceptual dimensions such as kinematics, Newton's first law, second law, third law, superposition principle for vector addition, and different kinds of force [224]. The authors of FCI emphasized that the total FCI score is the most reliable single index of student understanding because it measures coherence across all dimensions of Newtonian force concept [235]. Lasry et al.'s study also showed that FCI total score is highly reliable [230]. Therefore, in this study we use total score for FCI.

The path analysis in SEM gives regression coefficients $\beta$ for paths between each pair of constructs and the value of each $\beta$ is a measure of the strength of that relationship. Compared with a multiple regression model, the advantage of SEM is that we can estimate all regression links for multiple outcomes simultaneously, which improves the statistical power. The level of SEM model fit can be represented by the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residuals (SRMR), and CFI > 0.9, TLI > 0.9, RMSEA < 0.08, and SRMR < 0.08 are considered as acceptable [98]. We first analyzed the saturated SEM model as shown in Figure 13 (i.e., included all possible regression pathways), and then dropped the pathways that were not statistically significant to obtain a model that produced an acceptable fit to the data and contained only statistically significant regression paths. Before performing mediation analysis involving gender as shown schematically in Figure 13 (in which gender predicts each construct in the model through direct and indirect paths), we first tested the gender moderation relations between each pair of the constructs using multi-group SEM (to investigate any interaction effects with gender), which includes testing of factor loadings, indicator intercepts, residual variances and regression coefficients.

## 7.4 Results

### 7.4.1 Descriptive Statistics for Students' Sense of Belonging and Academic Performance

Table 21 shows the descriptive statistics of students' sense of belonging and FCI scores at the beginning and end of the course. As shown in Table 21, female students had statistically significantly lower scores in both sense of belonging and FCI than male students at the beginning of the course, and the gender differences were maintained at the end of the course. In addition, we found that from pre to post, both female and male students' FCI scores increased, while their sense of belonging decreased. Even though female students showed a larger improvement in FCI scores than male students according to the Cohen's *d*s, the FCI normalized gains were the same for men and women. In Appendix E, we report the percentages of female and male students who selected each answer choice from a 5-point Likert scale for each survey item for sense of belonging, which show consistent results with the descriptive statistics shown in Table 21.

Table 22 shows the descriptive statistics of students' high school GPA, SAT math scores, and grades in the physics course. As shown in Table 22, there was no statistically significant gender difference in students' SAT math scores, and female students had a higher average high school GPA than male students. In addition, even though the average course grade of women is lower than that of men, this gender difference is not statistically significant.

**Table 21 Descriptive statistics of pre- and post-sense of belonging (Bel) and FCI scores for female and male students. Cohen suggested that typically values of d ~ 0.2, 0.5 and 0.8 represent small, medium and large effect sizes, respectively [240]. Hake suggested that value of g < 0.3, 0.3 < g < 0.7, and g > 0.7 represent small, medium and large normalized gains, respectively [229]. A minus sign indicates that students' average score decreased from pre to post.**

| Gender | Pre-Bel | Post-Bel | Statistics | | |
|---|---|---|---|---|---|
| | Mean | Mean | Cohen's *d* | *p* value | |
| Male | 4.01 | 3.71 | -0.37 | <0.001 | |
| Female | 3.74 | 3.42 | -0.35 | <0.001 | |
| *p* value | <0.001 | 0.001 | | | |
| Cohen's *d* | 0.36 | 0.30 | | | |
| Gender | Pre-FCI | Post-FCI | Statistics | | |
| | Mean | Mean | Normalized gain (*g*) | Cohen's *d* | *p* value |
| Male | 63% | 72% | 0.23 | 0.42 | <0.001 |
| Female | 50% | 61% | 0.23 | 0.60 | <0.001 |
| *p* value | <0.001 | <0.001 | | | |
| Cohen's *d* | 0.63 | 0.50 | | | |

**Table 22 Descriptive statistics of female and male students' high school GPA, SAT math scores and course grades. A minus sign indicates that female students have a higher average score than male students.**

| Grades (Score Range) | Mean | | $p$ value | Cohen's $d$ |
|---|---|---|---|---|
| | Male | Female | | |
| High School GPA (0-5) | 4.12 | 4.31 | < 0.001 | -0.48 |
| SAT Math (400-800) | 706 | 703 | 0.593 | 0.05 |
| Grade (0-4) | 2.57 | 2.44 | 0.104 | 0.14 |

### 7.4.2 Moderation Analysis of Gender

We first conducted a moderation analysis to test whether gender moderates the relationship between any two constructs in the model (i.e., do the strength of relationships given by the standardized regression coefficients between any two constructs in the model differ for women and men?). We used the R [110] software package "lavaan" to conduct multi-group SEM. We initially tested for measurement invariance. In other words, we analyzed whether the factor loadings, intercepts, and residual variances of the observed variables are equal in the model where we measured the latent constructs so we can confidently perform multi-group analysis. The analysis involved introducing certain constraints in steps and testing the model differences from the previous step. In each step, we compared the model to both the previous step and the freely estimated model, that is, the model in which all parameters are freely estimated for each gender group. First, to test for "weak" or "metric" measurement invariance, we ran the model where only factor loadings were fixed to equality across both gender groups, but intercepts and errors were allowed to differ. The model was not statistically significantly different from the freely estimated model according to a likelihood ratio test, so weak measurement invariance holds (Chi-square difference ($\Delta\chi^2$) = 8.817, degree of freedom difference ($\Delta dof$) = 6, and non-significant $p$ = 0.184). Next, we tested for "strong" or "scalar" measurement invariance by fixing both factor loadings and intercepts to equality across gender groups. This model was not statistically significantly different from either the metric invariance model ($\Delta\chi^2$ = 6.977, $\Delta dof$ = 6, $p$ = 0.323) or the freely estimated model ($\Delta\chi^2$ = 15.794, $\Delta dof$ = 12, $p$ = 0.201), so strong measurement invariance holds. Finally, to test for "strict" measurement invariance, we fixed factor loadings, intercepts, and residual variances to equality. This model was not statistically significantly different from either the scalar invariance model ($\Delta\chi^2$ = 7.892, $\Delta dof$ = 8, $p$ = 0.444) or the freely estimated model ($\Delta\chi^2$ = 23.685, $\Delta dof$ = 20, $p$ = 0.256), so strict measurement invariance holds. Therefore, since all levels of measurement invariance hold for this model, we continued on to perform multi-group comparisons.

We ran a multi-group SEM in which all regression estimates were fixed to equality for female and male students in addition to the factor loadings and intercepts, and we compared this model with the freely estimated model. There was no statistically significant difference between the two models, so we reported the model in which regression pathways are equal for men and women. The model fit parameters for this case were acceptable (RMSEA = 0.053, SRMR = 0.065, CFI = 0.962, TLI = 0.960). The multi-group SEM results suggest that regression pathways among the constructs do not have differences across gender when compared to the freely estimated model ($\Delta\chi^2$ = 44.479, $\Delta dof$ = 34, $p$ = 0.108) or to the strict model ($\Delta\chi^2$ = 20.794, $\Delta dof$ = 14, $p$ = 0.107). Therefore, the results shown above indicate that in our model, strong measurement invariance holds and there is no difference in any regression coefficients by gender, which allowed us to perform the mediation analysis involving gender using SEM (as shown schematically in Figure 13).

### 7.4.3 SEM Model Controlling for Both SAT Math and High School GPA

In this section, we discuss our use of structural equation modeling (SEM) to investigate how students' sense of belonging predicts their course grades (quantitative understanding) and FCI scores (conceptual understanding) at the end of the course after controlling for students' gender, high school GPA and SAT math score as well as their sense of belonging and FCI scores at the beginning of the course. Fig. 2 shows the path analysis results of the integrated SEM model. The model fit indices suggest a good fit to the data: CFI = 1.000 (>0.90), TLI = 1.000 (>0.90), RMSEA = 0.000 (<0.08) and SRMR = 0.015 (<0.08). The solid lines represent regression paths, and the numbers on the lines are regression coefficients ($\beta$ values), which represent the strength of the regression relations.

As shown in Fig. 2, students' sense of belonging at the end of the course directly predicts their grades and post-FCI even after controlling for their pre-FCI, SAT math and high school GPA. In particular, the direct effect of post-sense of belonging on grades is $\beta = 0.23$, which is comparable to the effects of SAT math, high school GPA and post-FCI on students' grades. In addition, students' post-sense of belonging also directly predicts their post-FCI with $\beta = 0.12$, which provides an indirect path from post-sense of belonging to grades (post-Bel → post-FCI → Grade). We note that gender predicts high school GPA with a negative regression coefficient ($\beta = -0.23$), which means that female students on average had a somewhat higher high school GPA than male students. This result is consistent with the descriptive statistics shown in Table 22. We note that SAT math directly predicts post FCI ($\beta = 0.08$) and grade ($\beta = 0.21$), and high school GPA also directly predicts grade ($\beta = 0.23$). In addition, the Pearson correlation coefficient between SAT math and grade is 0.50 and between high school GPA and grade is 0.38. Therefore

[244,245], SAT math accounts for around 10% variance in students' course grades, and high school GPA accounts for around 9% variance in students' course grades. In Appendix F, we reported the results of full SEM (which estimates factor loadings and regression coefficients simultaneously), which are very similar to the results of the integrated SEM model shown in Figure 14.

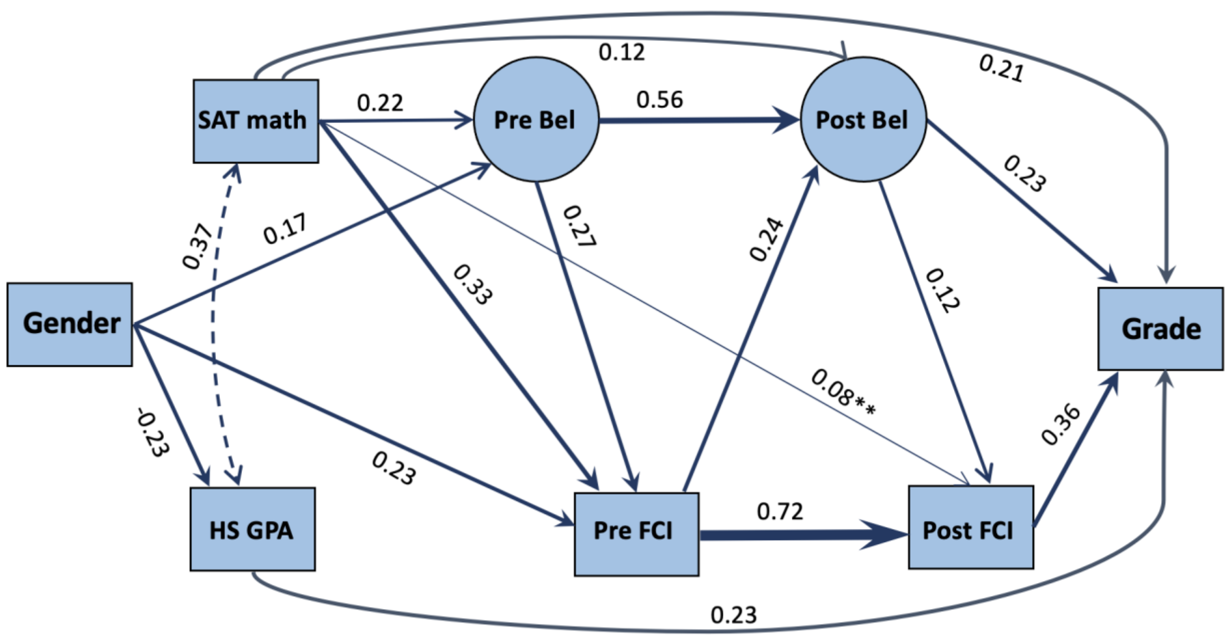



**Figure 14 Results of the path analysis part of the integrated SEM model. Bel represents sense of belonging. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. All the other regression lines show relations with $p < 0.001$**

To further understand how the model explains the variance in students' learning outcomes, we calculated the coefficients of determination $R^2$ (fraction of variance explained) for each construct in the SEM model ( Table 23). We found that pre-sense of belonging and pre-FCI have relatively small $R^2$ values compared to those for post-sense of belonging, post-FCI and grades. This is because they are input constructs that we are controlling for, which are predicted by fewer constructs than the outcome constructs. As shown in Table 23, $R^2$ values of all outcome constructs are reasonably high, which means that our model explains much variance in the outcome constructs.

**Table 23 Coefficient of determination ($R^2$) for various constructs in the SEM model. All $R^2$ values are significant with p values < 0.001. Bel represents sense of belonging.**

| Constructs | | | | |
|---|---|---|---|---|
| Pre Bel | Pre FCI | Post Bel | Post FCI | Grade |
| 0.07 | 0.30 | 0.53 | 0.67 | 0.51 |

## 7.5 Discussion

In this study, we focused on female and male students' sense of belonging and how it predicts students' learning outcomes including both conceptual understanding (measured by Force Concept Inventory) and quantitative problem solving (measured by course grades) in a college calculus-based introductory physics course. In particular, we studied how students' sense of belonging predicts their FCI scores and grades at the end of the course (post) after controlling for their gender, high school GPA, SAT math scores, and their sense of belonging and FCI scores at the beginning of the course (pre). We also investigated the gender differences in different constructs and whether gender moderates the relationship between any two constructs in our model.

Our results show that students' post-sense of belonging statistically significantly predicts their post-FCI scores and grades, and there is no statistically significant gender difference in the relationship between any two constructs in the model. The results also show that there are statistically significant gender differences disadvantaging women in both sense of belonging and FCI scores at the end of the course. Since sense of belonging is a predictor of FCI scores, the gender difference in students' sense of belonging may at least partially explain the gender gap in FCI scores. Due to reasons such as societal stereotypes and biases about who can do well in physics and lack of female role models, female students may have lower sense of belonging even before they enter college. Moreover, studies have shown that students from marginalized demographic groups such as women are more likely to be impacted negatively by negative cues relevant to belonging [197,198,246,247]. Thus, this type of gender gap in students' sense of belonging in physics is likely to exist unless intentional efforts are made by instructors to create an inclusive and equitable learning environment, in which all students can develop a high sense of belonging. If we could eliminate the gender difference in sense of belonging by creating a learning environment in which all students feel safe to engage in collaboration and discussions with peers and instructor, the gender difference in FCI scores may also decrease. In addition, our results show that both men's and women's sense of belonging decreased from pre to post, which means that the current physics learning environment is not helping students build a higher sense of belonging. Considering that FCI scores are predicted by sense of belonging, it is possible that students' post-FCI scores will be higher if we could improve students' sense of belonging in the class.

In addition, students' post-sense of belonging also statistically significantly predicts their course grades. We note that there is no gender difference in students' SAT math scores, and women actually have a higher average high school GPA than men; however, at the end of the physics course, women have somewhat lower course grades even though this gender difference is not statistically significant. Since there is a gender difference disadvantaging women with regard to sense of belonging, and sense of belonging statistically significantly predicts students' course grades, it is possible that female students would perform better in a more inclusive and equitable learning environment in which they have a higher sense of belonging.

We note that increased sense of belonging can be related to a variety of factors e.g., higher motivation to learn, self-regulation, engagement, initiative, more productive interactions with peers and instructors, and reduced anxiety, etc. For example, a possible hypothesis for why students' sense of belonging predicts their learning outcomes is that when students do not feel that they belong in the class, they may not fully take advantage of the interactions with their instructors and peers. For example, in group problem solving or class discussions, students with low sense of belonging may not fully participate in learning and discussions because they may not feel safe to share their ideas. For students from underrepresented groups, e.g., women in physics, it might be even more difficult to participate in group learning when the learning environments are not equitable and inclusive and discussions are dominated by students from the dominant groups. In addition, if appropriate attention is not paid to equity and inclusion, students with low sense of belonging may also feel less comfortable asking questions to instructors or TAs in class or office hours because they may worry that their questions are not good enough and feel nervous about showing that they do not know something. Moreover, the lack of sense of belonging may also result in anxiety [248], which can rob students of their cognitive resources and reduce their level of cognitive engagement while learning.

Our study hints at the important role played by instructors in improving students' sense of belonging by creating an inclusive and equitable learning environment. In particular, instructors need to keep in mind how the societal stereotypes and biases about who belongs in physics and who can excel in physics impact the stereotyped groups. They should also be mindful of the factors and cues that can influence students' sense of belonging and make explicit efforts to improve their classes. For example, instructors should provide every student with equal opportunity to ask and answer questions during the class and avoid letting a small group of students dominate the class discussion. In addition, studies have shown that framing adversity as something that all students experience and emphasizing that embracing struggle and using it as a stepping stone will help them become better over time at solving challenging physics problems can help students interpret adversity as normal and temporary instead of as a sign that the students do not belong and do not have what it takes to excel [144,194].

Furthermore, physics is one of the disciplines with the worst stereotypes about requiring a natural ability to be successful [11,83], and studies have shown that the idea of ability being fixed and unchangeable can increase students' concerns about belonging, especially for students from underrepresented groups who have few role models [249,250]. Thus, it is critical to build a learning environment which emphasizes that abilities are malleable and can be changed through deliberate practice and effort [146]. In addition, instructors can also show students non-stereotypical role models from diverse demographic groups, personalities and interest, since this has been shown to increase students' sense of belonging [251,252]. Instructors can apply these strategies in their classes and intentionally make efforts to develop an inclusive and equitable learning environment, in which all students recognize that failing and struggling are normal and they can overcome them by working hard, working smart using deliberate strategies, going to instructors' office hours, working with peers and taking advantage of all of the resources provided. In an equitable and inclusive learning environment, instructors convey to students that they have high expectations for all of them, know that each one of them will meet those expectations by embracing challenges, and are always there to support all students as needed. In an equitable and inclusive learning environment, students can be encouraged to view struggles as stepping stones to learning and be proud of struggling since it is a sign that they are on their way to learning.

In this study, we focus on the relationship between students' sense of belonging and learning outcomes in a college level calculus-based introductory physics course. Our results show that sense of belonging is a strong predictor of students' learning outcomes. Since this effect might be mediated by other factors such as students' motivation to learn, self-regulation, engagement in class, initiative, interactions with peers and instructors, the level of anxiety, etc., it would be helpful to investigate what factors mediate this effect and what role is played by each mediator in future studies. In addition, this study is based on students' self-reported responses to the motivational survey. It would be helpful to interview more students to get a deeper understanding of the mechanism of how the learning environment shapes students' sense of belonging and how sense of belonging influences students' learning outcomes in the physics course. It would also be helpful to use belonging interventions with control groups to further study the effect of sense of belonging on students conceptual understanding and course grades. In future studies, it would also be valuable to investigate students' sense of belonging and its impact on learning outcomes in other courses, such as algebra-based physics courses, where women are the majority group, often making up 60% or more of the classroom. In addition, it would be also useful to investigate students' sense of belonging and its impact on learning outcomes in advanced physics courses, which are usually taken by physics major students beyond the first year. Similar studies in other disciplines and in other countries would also be helpful for developing a deeper understanding of the role played by sense of belonging in student learning.

## 7.6 Conclusions

In this study discussed in the context of an introductory physics course at a large research university in the US, we find that students' sense of belonging plays an important role in predicting their physics conceptual understanding and grades (quantitative understanding) in the college physics course, and this relationship is comparable for men and women. Our results show that in the current learning environment, female students felt a significantly lower sense of belonging than male students, which can contribute to the gender difference in students' learning outcomes at the end of the course. These findings suggest that women's lower sense of belonging can explain their physics learning and hint at the importance of instructors being intentional about creating an inclusive and equitable learning environment in which students from all demographic groups have a high sense of belonging and can excel.

# 8.0 How Perception of Learning Environment Predicts Male and Female Students' Physics Self-Efficacy, Interest and Identity in Calculus-Based Introductory Physics Courses

## 8.1 Introduction and Theoretical Framework

Several prior studies have focused on underrepresented groups such as women in science, technology, engineering, and mathematics (STEM) courses, majors and careers [1-7,13,21,22,36-38,63,135,181,184,185,253,254]. Prior research suggests that individuals' career enrollment and achievement in STEM can be influenced by their domain specific motivational beliefs such as self-efficacy, interest and identity [18-22,34,37,41,43,45,48,63,73,255-261]. For students from underrepresented groups, these motivational characteristics might be undermined due to lack of encouragement, negative stereotypes, and inadequate prior preparation, leading to withdrawal from STEM fields [24-32,74,262]. Hence, investigating students' motivational characteristics is critical to understanding and addressing diversity, equity, and inclusion in STEM disciplines.

Prior research suggests that self-efficacy is an important motivational characteristic of students in order for them to excel in a domain [20,36-38]. In particular, self-efficacy is the belief in one's capability to be successful in a particular task, course, or subject area [39,40], and it has been shown to influence students' engagement and performance in a given domain [41-43,122,258]. Students with high self-efficacy in a domain often enroll in more difficult courses in that domain than those with low self-efficacy because they perceive difficult tasks as challenges rather than threats [45].

Another motivational characteristic is interest, which is defined by positive emotions accompanied by curiosity and engagement in a particular discipline [46,47]. Studies have shown that interest can also influence students' learning [41,47-50]. For example, one study shows that making science courses more relevant to students' lives and transforming curricula to promote interest in learning can improve students' achievement [52]. In addition, studies have shown that students' interest is not independent from self-efficacy [41,54].

According to Eccles' Expectancy-Value Theory (EVT) [53,54], interest is paired well with self-efficacy as connected constructs that predict students' academic outcome expectations and career aspirations. In this theory, students' persistence and engagement in a task or field can be influenced by their expectancy of success and by their estimation of the task's value. The expectancy here refers to learners' belief in their ability to succeed in the given task [54], which is closely related to self-efficacy. Value in this theory includes four components: intrinsic value, attainment value, utility value, and cost [54]. Intrinsic value represents students' personal interest in the task or field. Attainment value refers to how important students themselves feel it is for them to develop mastery and do a good job in the field [54]. Utility value pertains to whether this task can help them succeed in various fields [54]. The last value component is cost, which corresponds to the negative aspect of engagement such as the amount of anxiety or opportunity cost due to the time spent on the task [54]. In the Expectancy-Value Theory, people's learning goal, academic performance, and persistence in the field are impacted by their expectancy and the four components of value [54].

Science identity is another important motivational characteristic that influences students' career decisions and outcome expectations [21,22,55-58,63,101]. Students' identity in a specific field such as physics is related to whether they see themselves as a physics person [21,22,55,58,63,101]. Some studies have found that female students are less likely to see themselves as a physics person than male students [63,64]. In prior research, in general, the challenges women face in developing physics identity are related to societal biases and stereotypes [66-68]. For example, negative societal stereotypes and generalizations about who can succeed in physics and other STEM disciplines can lower women's sense of belonging and undermined their experiences, so that they often have lower STEM identity than men [24,69,70]. Thus, studying STEM identity may help us to understand the gender difference in participation in STEM. The well-known science identity framework developed by Carlone and Johnson [22] includes three dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others"). Hazari et al. adapted this model to physics and added interest to this model [21]. They investigated whether the relation between gender and physics identity was mediated by interest, competency belief, and perceived recognition from other people [78,79]. These two studies reveal that individuals' internal identity in science is not only impacted by their own motivational characteristics but also by their perceived recognition from others.

Several studies have shown that female students did not feel that they were recognized appropriately even before they entered college [66,83,183]. We have conducted interviews with women in physics courses and it is clear from these interviews that they felt that in general men were recognized more by the instructors than women and it impacted their self-efficacy, interest and identity as a physics person [71,76]. One of the stereotypical views of science is that it is for high achievers or naturally gifted students [66]. In general, due to societal stereotypes, being a genius or exceptionally smart is associated with boys [156]. In one investigation, boys and girls were externally exposed to these fixed intelligence views starting from early ages, which influenced the development of their science identity [83]. In addition, one study indicated that elementary and high school boys and girls interested in science were treated differently by parents, teachers and friends. 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 [84]. Studies showed that these stereotypes and biases also exist in university context [81,85]. For example, one study found that science faculty members in biological and physical sciences exhibit biases against female students by rating male students significantly more competent [85]. Our prior study also found that in introductory physics course, there is a significant gender difference in perceived recognition from TAs and instructors [81,112]. There is often a feedback loop, e.g., between recognition and identity. For example, the experiences of not being recognized as a science person may further increase the stereotype threat, and these gender-based biases may accumulate over time and become a detriment to female students' science identity.

Moreover, students' interest and self-efficacy have also been found to be connected to their interaction with other people and recognition by them [40,47]. In Hidi and Renninger's four stages model of interest development [47,49,86], people's interest in a field is triggered and maintained by external factors first and then becomes an individual interest. In addition, according to Bandura's social cognitive theory, individual's self-efficacy can be shaped by verbal encouragement from others [151,153]. Kalender et al.'s physics identity framework [81] showed that students' perceived recognition not only strongly predicts their physics identity, but also predicts their physics interest and self-efficacy, which suggests that the gender difference in students' perceived classroom experience may partially explain the gender differences in students' physics self-efficacy, interest and identity at the end of the course. However, it is not clear how much of self-efficacy and interest changed from the beginning to the end of the course, and what role was played by perception of learning environment in this change.

In addition to perceived recognition, some studies have shown that students' sense of belonging and their interaction with peers are also important constructs of learning environment [2,69,70,195,263-265]. For example, if students feel more secure in their belonging in school, they may approach others in the academic environment more and with more positive attitudes, building better interaction and higher perceived recognition [178]. However, there are very few quantitative studies about the effect of learning environment on students' motivational beliefs and what role is played by each factor in the learning environment. To better understand how the learning environment influences students' motivational outcomes and how to develop an inclusive and equitable learning environment, further study is needed.

Our conceptualization of inclusive excellence and equity in learning includes three pillars: equitable access and opportunity to learn, equitable and inclusive learning environment, and equitable outcomes. Thus, by inclusive excellence and equity in learning, we mean that not only should all students have equitable opportunities and access to resources, they should also have an equitable and inclusive learning environment with appropriate support and mentoring so that they can engage in learning in a meaningful and enjoyable manner and the learning outcomes should be equitable. By equitable learning outcomes, we mean that students from all demographic groups (e.g., regardless of their gender identity or race/ethnicity) who have the pre-requisites to enroll in courses have comparable learning outcomes. This conceptualization of equitable outcome is consistent with Rodriguez et. al.'s equity of parity model [33]. The learning outcomes include student performance in courses as well as evolution in their motivational beliefs such as self-efficacy etc. because regardless of performance, students' motivational beliefs can influence their short and long-term retention in their major and careers. In other words, an equitable and inclusive learning environment should provide guidance, support and mentoring to all students as appropriate and ensure that students from all demographic groups have equal sense of belonging regardless of their prior preparation so long as they have the prerequisite basic knowledge and skills. An equitable and inclusive learning environment would also ensure that students from all demographic groups and prior preparation embrace challenges as learning opportunity instead of being threatened by them and enjoy learning. Equitable learning outcomes also include the ability of the courses to empower students from all demographic groups and make them passionate to pursue further learning and careers in related areas. We note that equitable access and opportunity to learn, equitable and inclusive learning environment and equitable outcomes are strongly entangled with each other. For example, if the learning environment is not equitable and inclusive in a particular course, the learning outcomes are unlikely to be equitable.

To improve equity and inclusion in physics classes, we conducted a study to investigate the effect of the perception of learning environment (including sense of belonging, perceived peer interaction, and perceived recognition) on students' self-efficacy, interest and identity by controlling for students' gender and their self-efficacy and interest at the beginning of a calculus-based introductory physics course. We note that the learning environment mentioned here is not necessarily the classroom environment: it also includes students' experiences outside the class. For example, students may work together on their homework after class, and they could also ask for help during instructors' office hours. As shown in Figure 15, the total nine constructs are divided into three groups: what we control for, perception of learning environment, and outcomes. Students' gender, pre-self-efficacy (Pre SE) and pre-interest measured at the beginning of the course are constructs that we control for, which are related to students' beliefs about physics based on their prior experience. Outcomes include students' physics post-self-efficacy (Post SE), post-interest and identity at the end of the course. Perceived recognition (Perceived Recog), perceived peer interaction (Peer Int) and sense of belonging (Belonging) constitute the perception of learning environment. For convenience, perceived peer interaction is shortened to peer interaction in the rest of the paper. We note that out of the three components of the pre-physics identity, pre-self-efficacy and pre-interest have been controlled for. However, perceived recognition is part of the perception of learning environment and thus is not controlled for at the beginning of the course (i.e., students have no experience interacting with their instructors and TAs).

In our study, peer interaction, perceived recognition and sense of belonging were measured at the end of the course. This is because only after the course can students answer these survey questions based on their real experience in the course such as their interaction with peers, TAs and instructors. Because pre- and post- responses are actually students' responses to the same questions at two different time points, it is not surprising if students' pre-self-efficacy and pre-interest partially predict their post-self-efficacy and post-interest. However, if students' self-efficacy and interest changed from pre to post, we want to study whether the perception of learning environment helps to explain the changes and what role is played by each construct in the perception of learning environment.

In this study, we first calculated female and male students' mean scores for each motivational construct. Then, we studied how much of students' self-efficacy and interest changed from pre- to post- and how much of these changes can be explained by the perception of learning environment. We used Structural Equation Modeling (SEM) to study the effect of learning environment on students' post-self-efficacy, interest, and identity by controlling for gender and their pre-self-efficacy and interest. To better understand the role played by each perception of learning environment construct, we first considered a model with perceived recognition as the only perception of learning environment construct to see how much variance in students' self-efficacy, interest and identity was explained by the model. Then, we added peer interaction and sense of belonging into this model one by one to see whether the adding constructs help to explain extra variance in students' motivational outcomes.

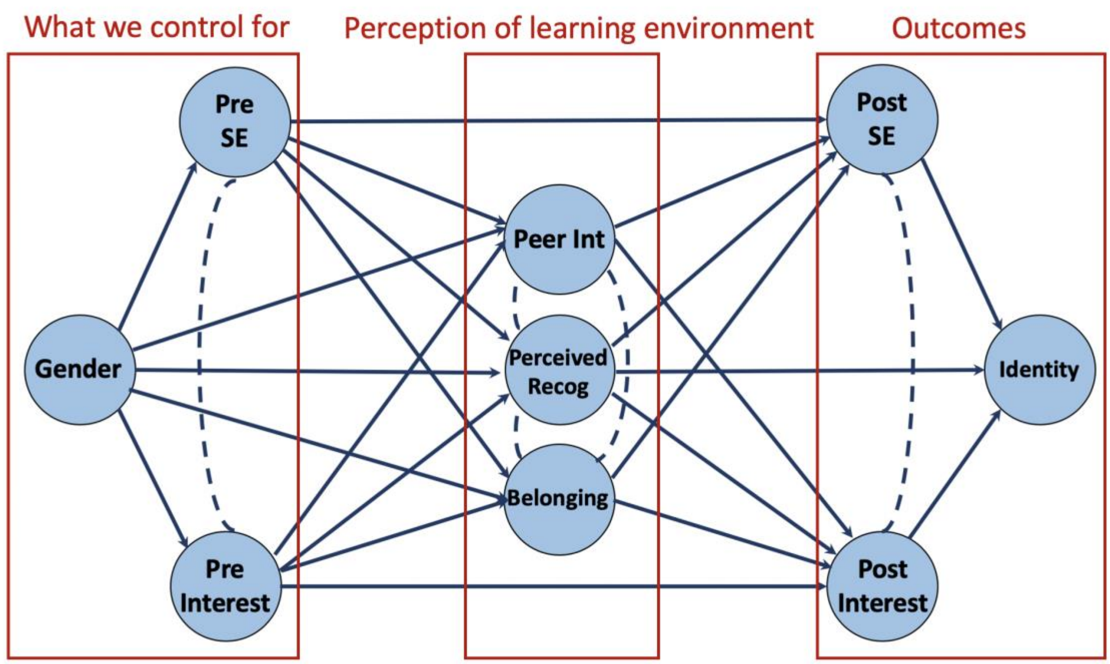



**Figure 15 Schematic representation of the theoretical framework regarding physics identity. From left to right, all possible regression paths were considered, but only some of the paths are shown here.**

## 8.2 Research Questions

Our research questions regarding the calculus-based introductory physics course 1 at a large state-related university are as follows. This course is generally taken by engineering, physical science, and mathematics majors in the first year of undergraduate studies.

**RQ1.** Are there gender differences in students' motivational characteristics and do they change from the beginning to the end of the course (i.e., from pre to post)?

**RQ2.** How do the components of the perception of learning environment (including sense of belonging, peer interaction and perceived recognition) predict students' physics identity as well as post-self-efficacy and post-interest controlling their gender, pre-self-efficacy and pre-interest?

**RQ3.** Does gender moderate the relationship between any two constructs in the model? (i.e., Does the strength of relationship given by the standardized regression coefficient between any two constructs in the model differ for women and men?)

**RQ4.** If gender does not moderate any path in the model, how does gender mediate

a. the factors that were controlled for?
b. the perception of learning environment after controlling for pre-self-efficacy and pre-interest?
c. the motivational outcomes after controlling for everything in the model?

**RQ5.** What unique role is played by each of the three components we have included in the learning environment in predicting students' motivational outcomes?

**RQ6.** Based on the aspects of the perception of learning environment that explain most of the variance in the outcomes, which model is most productive for creating the inclusive environment?

## 8.3 Methodology

### 8.3.1 Participants

In this study, we collected motivational survey data at the beginning and end of the semester from students who took the introductory calculus-based physics 1 course in two consecutive fall semesters. This course is taken mostly by students majoring in engineering, physical science, and mathematics. This course is a traditional lecture-based course (4 hours per week) with recitations (1 hour per week) in which students typically work collaboratively on physics problems. The paper surveys were handed out and collected by TAs in the first and last recitation class of a semester. Course instructors were encouraged to give students course credit or extra credit for completing the survey, and the completion rates are typically 80%-90%. We named the data collected at the beginning of the semester as pre-data and that collected at the end of the semester as post-data. Finally, we combined the two semesters' data and put them into two categories, pre and post. The demographic data of students—such as gender—were provided by the university. Students' names and IDs were de-identified by an honest broker who provided each student with a unique new ID (which connected students' survey responses with their demographic information). Thus, researchers could analyze students' data without having access to students' identifying information.

There were 1364 students participating in the pre-survey and 1219 students participating in the post-survey. There were 1052 students participating both pre- and post- survey. Students may miss the first or the last recitation class for many possible reasons. In this study, we only considered students who are in our post-data set even though some of them might not take the pre-survey. This is because our focus is on students' motivational outcomes at the end of the course, and what role is played by the perception of learning environment in predicting students' motivational outcomes by controlling for their pre-motivational scores. In addition, SEM can provide a better estimation with the more completed outcome data [266]. Thus, it is reasonable to keep students' response to the post-survey rather than the pre-survey. To handle the missing data, we used the full information maximum likelihood method to estimate the model by using all available information [170]. In this method, the population parameters are estimated that would most likely produce the estimates from the sample data that is analyzed [170].

In all, there were 1219 students participating in the post-survey including both semesters. In our final data set, we kept 1203 students (including 427 female students and 776 male students) because the other 16 students did not provide their gender information. We recognize that gender identity is not a binary construct. However, because students' gender information was collected by the university which offered binary options, we did the analysis with the binary gender data in this study. 1.3% of the students who did not provide this information were not included in this analysis.

### 8.3.2 Survey Instruments

In this study, our analysis includes six motivational constructs—physics self-efficacy, interest, peer interaction, perceived recognition, sense of belonging, and identity. The questions for each construct are listed in Table 24. The survey questions were adapted from the existing motivational research [92-94,162-164,267] and were re-validated in our prior work [37,95,96,127,165]. The validation and refinement of the survey involved use of one-on-one student interviews with both introductory and advanced students [37,76,96,268], exploratory and confirmatory factor analyses (EFA and CFA) [97], Pearson correlation between different constructs and Cronbach alpha [99,100]. The peer interaction scale was added later, and therefore additional validation (CFA and Cronbach alpha results) for that scale is reported in Table 24.

Physics self-efficacy represents students' belief about whether they can excel in physics. In our survey, we had four items for self-efficacy (Cronbach alpha = 0.69 for pre-self-efficacy and Cronbach alpha = 0.8 for the post-self-efficacy [99]). These items had response scale of "NO!, no, yes, YES!", which is a 4-point Likert scale (1-4). We also had four items for physics interest (Cronbach alpha = 0.75 for the pre-interest, Cronbach alpha = 0.82 for the post-interest). The question "I wonder about how physics works" had temporal response options: "Never, Once a month, Once a week, Everyday." 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. By choosing the four options, students will get score from 1 to 4 accordingly. For example, if a student finds physics very boring, he/she will get one point for this item. And the more interest a student has in physics, the higher score the student will get for this item. It is noteworthy that Cronbach' alpha is always higher in our post-data than in the pre-data. This may be because students had a clearer judgment about their self-efficacy and interest after one semester of learning so that they could answer those questions in a more consistent way at the end of the semester than at the beginning of the semester.

There is one item for physics identity in this survey (I see myself a physics person). Physics identity corresponds to students' belief about whether they designate themselves as a physics person [21]. This item involved a four-point Likert response on the scale: "strongly disagree, disagree, agree, and strongly agree" and they correspond to 1 to 4 points [102].

In addition, perceived recognition, peer interaction and sense of belonging are the other three motivational constructs in our study. Unlike self-efficacy, interest and identity, these three constructs are directly related to students' experience in the course. Perceived recognition included three items which represent whether a student think other people see them as a physics person [21,63,101] (Cronbach alpha = 0.86). Peer interaction including four items represents whether students have a productive and enjoyable experience when working with peers (Cronbach alpha = 0.91). Sense of belonging is about students' feelings of whether they belonged in the physics class [195], and it included five items that were scored on a 5-point Likert scale: "not at all true, a little true, somewhat true, mostly true and completely true" (Cronbach alpha = 0.86). Two sense of belonging items ("I feel like an outsider in this class" and "Sometimes I worry that I do not belong in this physics class") were reverse coded, which means that a higher score in these two items represents a lower sense of belonging. Students' score of each construct is the average score of all items in this construct.

**Table 24 Survey items for each of the motivational scales. The Cronbach alphas and CFA item loadings (Lambda and p-values of the significance test for each item loading) shown here were calculated with the post-data. †The response options for this question are "Never, Once a month, Once a week, Every day". ‡The response options for this question are "Very boring, boring, interesting, Very interesting".**

| Construct and Item | Lambda | *p* value |
|---|---|---|
| Physics Identity | | |
| I see myself as physics person. | 1.000 | <0.001 |
| Physics Self-Efficacy (Cronbach alpha = 0.80) | | |
| I am able to help my classmates with physics in the laboratory or in recitation. | 0.722 | <0.001 |
| I understand concepts I have studied in physics. | 0.726 | <0.001 |
| If I study, I will do well on a physics test. | 0.727 | <0.001 |
| If I encounter a setback in a physics exam, I can overcome it. | 0.669 | <0.001 |
| Physics Interest (Cronbach alpha = 0.82) | | |
| I wonder about how physics works † | 0.664 | <0.001 |
| In general, I find physics ‡ | 0.795 | <0.001 |
| I want to know everything I can about physics. | 0.796 | <0.001 |
| I am curious about recent physics discoveries. | 0.693 | <0.001 |
| Physics Perceived Recognition (Cronbach alpha = 0.86) | | |
| My family sees me as physics person. | 0.902 | <0.001 |
| My friends see me as physics person. | 0.899 | <0.001 |
| My physics TA and/or instructor see me as physics person. | 0.693 | <0.001 |
| Physics Sense of Belonging (Cronbach alpha = 0.86) | | |
| I feel like I belong in this class. | 0.831 | <0.001 |
| I feel like an outsider in this class. | 0.697 | <0.001 |
| I feel comfortable in this class. | 0.830 | <0.001 |
| I feel like I can be myself in this class. | 0.616 | <0.001 |
| Sometimes I worry that I do not belong in this physics class. | 0.712 | <0.001 |
| Physics Peer Interaction (Cronbach alpha = 0.91)<br>My experience and interaction with other students in this class… | | |
| made me feel more relaxed about learning physics. | 0.717 | <0.001 |
| increased my confidence in my ability to do physics. | 0.910 | <0.001 |
| increased my confidence that I can succeed in physics. | 0.928 | <0.001 |
| increased my confidence in my ability to handle difficult physics problems. | 0.846 | <0.001 |

### 8.3.3 Quantitative Analysis of Survey Data

First, we calculated the mean score for each construct for each student. Then we used a regular two-sample *t*-test [103,104] to compare students' pre and post-score and to compare responses for female and male students. Finally, we used Structural Equation Modeling (SEM) [87] to study the effect of the perception of learning environment on students' motivational outcomes by controlling for gender and their pre-self-efficacy and interest. The SEM includes two parts: confirmatory factor analysis (CFA) and path analysis.

To validate the items on our survey, we performed the CFA for each construct. The model fit is good if the fit parameters are above certain thresholds. 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 good fit [266]. In our study, CFI = 0.943, TLI = 0.934, RMSEA = 0.05 and SRMR = 0.041 which represents a good fit. Thus, there is additional quantitative support for dividing the constructs as proposed. In addition, as show in Table 24, all of the CFA item loadings are above 0.5 and most of them are above 0.7, which means that our constructs extract sufficient variance from the items [239]. The CFA results allowed us to perform the path analysis.

Before performing the path analysis, we calculated the Pearson correlation coefficients pairwise between each two constructs [100]. As shown in **Error! Reference source not found.**, e ven though these constructs have strong correlations with each other, the correlations are not so high that the constructs could not be separately examined in SEM [109]. It is noteworthy that in **Error! Reference source not found.**, there are three very strong correlations. The correlation c oefficient between the pre-interest and post-interest is 0.87, which means that students' interest at the end of the course is highly related to their interest at the beginning of the course. The correlation coefficient between physics identity and perceived recognition is 0.84, which is consistent with Godwin et al. and Kalender et al.'s prior work [58,81] finding that perceived recognition is the largest predictor of physics identity. Another large coefficient is between students' post-self-efficacy and sense of belonging. According to prior work done by Kalender et al., these two constructs are indeed strongly correlated with each other [183].

**Table 25 Zeroth order correlation coefficients of the constructs in the mediation model.**

| Observed Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Physics identity | -- | -- | -- | -- | -- | -- | -- | -- |
| 2. Pre-Self-efficacy | 0.55 | -- | -- | -- | -- | -- | -- | -- |
| 3. Pre-Interest | 0.60 | 0.66 | -- | -- | -- | -- | -- | -- |
| 4. Post-Self-efficacy | 0.68 | 0.61 | 0.45 | -- | -- | -- | -- | -- |
| 5. Post-Interest | 0.70 | 0.49 | 0.87 | 0.63 | -- | -- | -- | -- |
| 6. Perceived Recognition | 0.84 | 0.46 | 0.59 | 0.70 | 0.68 | -- | -- | -- |
| 7. Peer Interaction | 0.51 | 0.37 | 0.31 | 0.67 | 0.47 | 0.49 | -- | -- |
| 8. Sense of Belonging | 0.62 | 0.46 | 0.38 | 0.83 | 0.55 | 0.63 | 0.68 | -- |

To analyze the relations among the constructs, we performed the full SEM. Apart from CFA, the path analysis in SEM gives regression coefficients $\beta$ for paths between each pair of constructs and the value of each $\beta$ is a measure of the strength of that relationship. 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. The level of SEM model fit can also be represented by CFI, TLI, RMSEA and SRMR. We first analyzed the saturated SEM model that includes all of the 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. Before performing gender mediation analysis, we first tested the gender moderation relations between each pair of constructs using multi-group 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 of 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 (see Appendix G for detailed multi-group SEM analysis results).

One advantage of SEM is that it shows not only the direct regression relation between two constructs but also all of the indirect relations mediated though other constructs, which allowed us to calculate the total regression effect by adding the direct and indirect regression coefficients up. In this study, we first considered a model with perceived recognition as the only perception of learning environment construct to see how students' physics self-efficacy, interest and identity were predicted by it. Then, we added peer interaction or sense of belonging as additional constructs in the perception of learning environment. Finally, our model included all of the perception of learning environment constructs. We analyzed the variance in each motivational outcome denoting students' motivational characteristics explained by each model to understand the unique role played by each perception of learning environment component and to determine if all three perception of learning environment components are productive.

## 8.4 Results

### 8.4.1 Gender Differences in Motivational Characteristics

The results of the regular two-sample *t*-test are shown in Table 26 and Table 27. We also used a paired sample *t*-test for matched pairs to compare students' pre- and post-self-efficacy and interest, and the results are similar to those using the regular *t*-test (see Appendix H for results of the paired-sample *t*-test for matched pairs of pre and post responses). As shown in Table 26, female students had significantly lower average interest and self-efficacy scores in both pre- and post-survey than male students, and these gender differences increased by the end of the semester. The effect size given by Cohen's *d* [104] of gender difference in physics interest increased from 0.54 to 0.60, and the effect size of gender difference in self-efficacy increased from 0.32 to 0.53. In addition, even though students' interest and self-efficacy dropped generally from pre to post (see Table 26), female students' interest and self-efficacy dropped ($d$ = 0.52 for self-efficacy and $d$ = 0.30 for interest) even more than male students' ($d$ = 0.27 for self-efficacy and $d$ = 0.19 for interest). Table 27 shows the average scores on the other constructs (perception of peer interaction, perceived recognition, sense of belonging and identity) in the post-survey. As shown in Table 27, female students also had significantly lower average scores in all of the four constructs than male students, and the effect sizes are all in the medium range [104].

**Table 26 Descriptive statistics of pre- and post-self-efficacy and interest for female and male students.**

| Gender | Pre-Interest (1-4) | Post-Interest (1-4) | Statistics | | Pre-SE (1-4) | Post-SE (1-4) | Statistics | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Mean | *p* value | Cohen's *d* | Mean | Mean | *p* value | Cohen's *d* |
| Male | 3.19 | 3.08 | <0.001 | 0.19 | 3.12 | 2.98 | <0.001 | 0.27 |
| Female | 2.89 | 2.70 | <0.001 | 0.30 | 2.96 | 2.68 | <0.001 | 0.52 |
| *p* value | <0.001 | <0.001 | | | <0.001 | <0.001 | | |
| Cohen's *d* | 0.54 | 0.60 | | | 0.32 | 0.53 | | |

**Table 27 Descriptive statistics of peer interaction, perceived recognition, sense of belonging and identity for female and male students.**

| Gender | Post-peer Int (1-4) | Post-perceived recog (1-4) | Post-sense of belonging (1-5) | Post-identity (1-4) |
|---|---|---|---|---|
| Male | 2.97 | 2.60 | 3.73 | 2.63 |
| Female | 2.68 | 2.24 | 3.33 | 2.17 |
| p value | <0.001 | <0.001 | <0.001 | <0.001 |
| Cohen's d | 0.44 | 0.49 | 0.46 | 0.56 |

### 8.4.2 Perception of Learning Environment Mediation Models

In this section, we will show the predictive relationships among the constructs using SEM models. Because many studies have shown that perceived recognition is a strong predictor of students' motivational beliefs and identity [81,269-271], all of the models shown in the main text of this paper include perceived recognition as one of the perception of learning environment constructs (see Appendix I for results of other SEM models). Firstly, perceived recognition was the only perception of learning environment construct in the model. Then we added peer interaction or sense of belonging to the perception of learning environment one by one to analyze how each helped to predict students' identity, post-self-efficacy and interest. Finally, we included all of the three constructs in our model and studied how these constructs mediated the outcomes together and what role was played by each of them.

Figure 16 shows the path analysis results of the SEM model in which perceived recognition is the only perception of learning environment construct. The model fit indices suggest a good fit to the data: CFI = 0.939 (>0.90), TLI = 0.927 (>0.90), RMSEA = 0.055 (<0.08) and SRME = 0.043 (<0.08). The solid lines represent regression paths, and the numbers on the lines are regression coefficients ($\beta$ values), which represent the strength of the regression relations. As we can see in Figure 16, there are two paths going from pre-self-efficacy to post-self-efficacy. One path goes from pre- to post- directly, and the other path goes through the perceived recognition. The regression coefficient of the direct path from pre-self-efficacy to post-self-efficacy is 0.35. The regression coefficient of the indirect path can be calculated by multiplying the regression coefficients from pre-self-efficacy to perceived recognition ($\beta$ = 0.28) and the regression coefficient from perceived recognition to post-self-efficacy ($\beta$ = 0.50), which gives us

$0.28 \times 0.50 = 0.14$. Similarly, the direct effect from pre-interest to post-interest is $\beta$ = 0.72, and the indirect effect is $0.42 \times 0.25 = 0.11$.

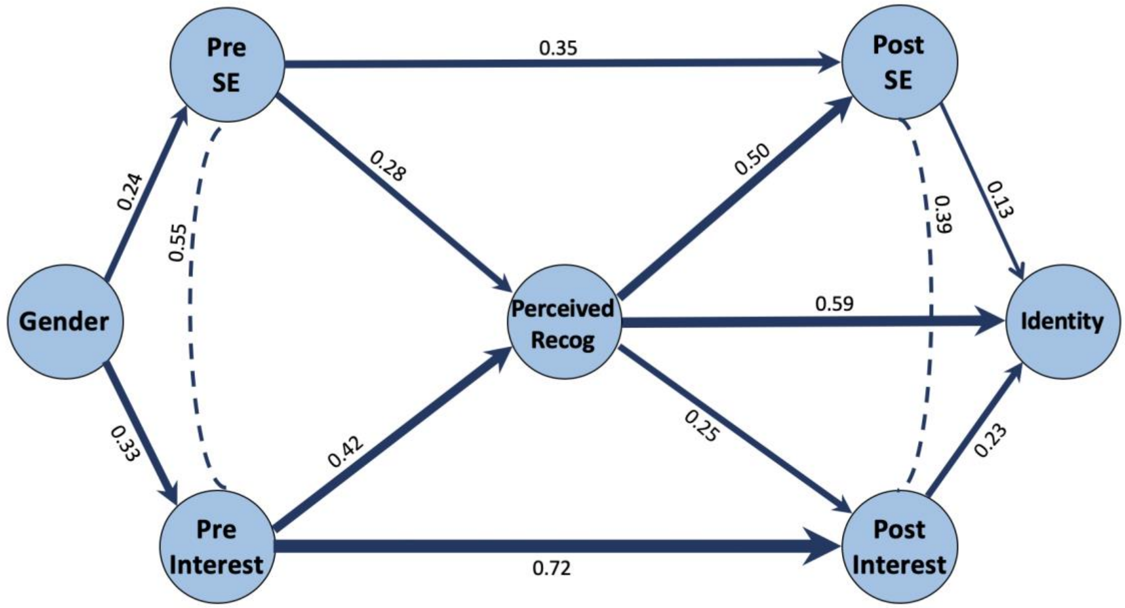



**Figure 16 Schematic diagram of the path analysis part of the structural equation modeling between gender and being a physics person through self-efficacy, interest and perceived recognition. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of β value (standardized regression coefficient). All β values are significant with p < 0.001.**

We then added peer interaction to the model. The results of the SEM model are presented visually in Figure 17 This model also fit the data very well. CFI = 0.951 (>0.90), TLI= 0.943 (>0.90), RMSEA = 0.048 (<0.08) and SRMR = 0.040 (<0.08). The results show that peer interaction directly predicts post-self-efficacy and interest. By comparing this model with the one only including perceived recognition, we can see that after adding peer interaction to the model, the direct effect of pre-self-efficacy on post-self-efficacy became weaker (the direct $\beta$ value dropped from 0.35 to 0.29). In addition, the direct effect of perceived recognition on post-self-efficacy also decreased (the $\beta$ value from perceived recognition to post-self-efficacy dropped from

0.50 to 0.35). This is due to the shared variance between peer interaction and the other direct predictors (pre-self-efficacy and perceived recognition) of post-self-efficacy. The regression coefficient from a predictor to an outcome represents the expected changes in the outcome as a result of changes in the predictor in standardized deviation units while controlling for the correlated effects of other predictors [244]. When peer interaction was added to the model, the direct effect of pre-self-efficacy and perceived recognition on post-self-efficacy decreased because more correlated effect had been controlled for. Similarly, the direct effect of pre-interest and perceived recognition on post-interest also decreased because the effect that correlated with peer interaction had been controlled for. It is noteworthy that the direct effect of perceived recognition, post-interest and post-self-efficacy on identity did not change after adding peer interaction. This is because peer interaction does not predict identity directly.

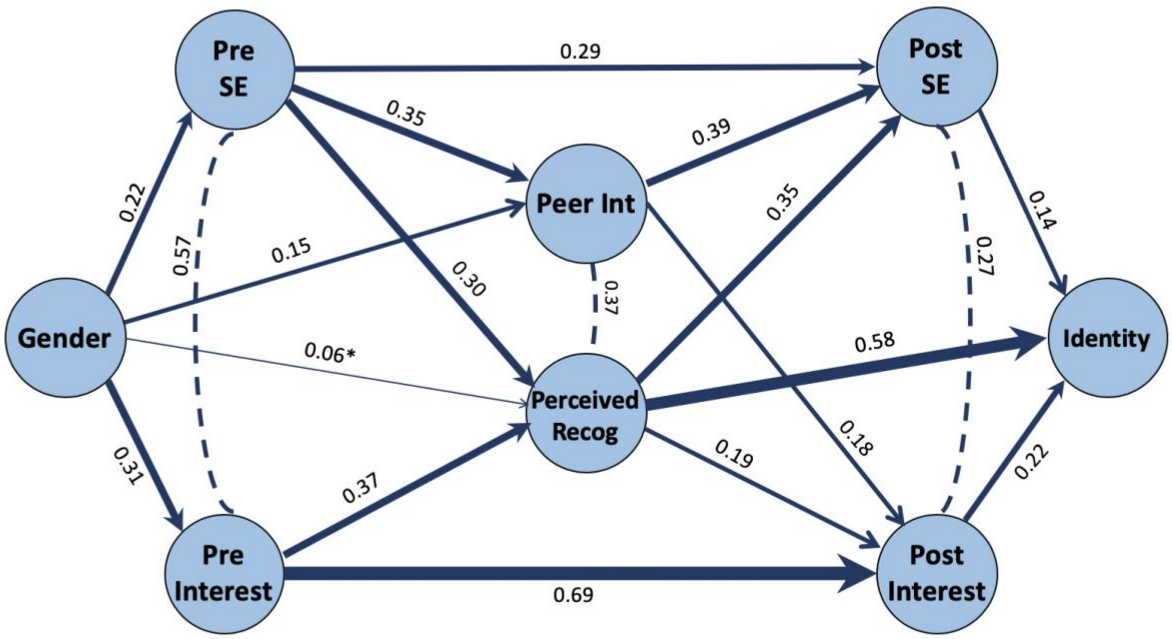



**Figure 17 Schematic diagram of the path analysis part of the structural equation modeling between gender and being a physics person through self-efficacy, interest, perceived recognition and peer interaction. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of β value (standardized regression coefficient) with 0.01< p < 0.05 indicated by *. Other regression lines show relations with p < 0.001.**

Then, we analyzed a SEM model which only includes perceived recognition and sense of belonging as shown in Figure 18 The model also fits the data well (CFI = 0.930 (>0.90), TLI = 0.919 (>0.90), RMSEA = 0.054 (<0.08) and SRMR = 0.043 (<0.08)). Figure 18 shows that sense of belonging is predicted by pre-self-efficacy and interest, and it directly predicts post-self-efficacy and interest. In addition, there is a strong correlation between sense of belonging and perceived recognition. Thus, there is more correlated effect being controlled for when estimating the regression coefficients from the predictors to the outcomes. For example, compared to the model only including perceived recognition (Figure 16), adding sense of belonging decreased the direct regression coefficients from pre-self-efficacy, pre-interest, and perceived recognition to post-self-efficacy and post-interest. It is noteworthy that even though pre-self-efficacy directly predicts identity in this model as shown in Figure 18, because the regression coefficient of this path is very small ($\beta$ = 0.06), the regression coefficients from post-self-efficacy, post-interest and perceived recognition to identity are still almost the same as in the first two models discussed earlier.

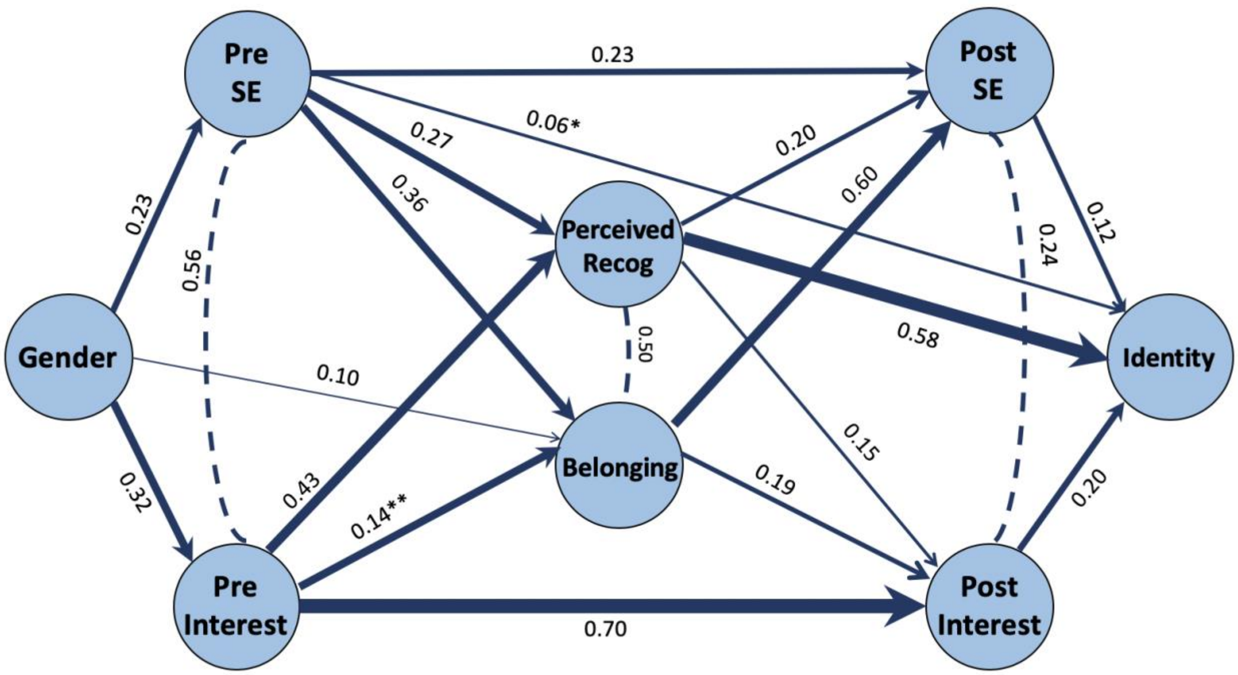



**Figure 18 Schematic diagram of the path analysis part of the structural equation modeling between gender and being a physics person through self-efficacy, interest, perceived recognition and peer interaction. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of β value (standardized regression coefficient) with 0.01< p < 0.05 indicated by * and 0.001 < p < 0.01 indicated by **. Other regression lines show relations with p < 0.001.**

Finally, we analyzed a SEM model which includes all of the three perception of learning environment constructs. Figure 19 shows the results visually. This model also fits the data very well (CFI = 0.941 (>0.90), TLI = 0.932 (>0.90), RMSEA = 0.049 (<0.08) and SRMR = 0.042 (<0.08)). As shown in Figure 19, all three perception of learning environment constructs directly predict post-self-efficacy and post-interest, and the direct effect from pre-self-efficacy to post-self-efficacy and from pre-interest to post-interest are weaker than any models discussed earlier. In other words, as we added more constructs to the perception of learning environment, the strength of the direct paths from pre to post decreased. This result indicates that the perception of learning environment is mediating the effect of students' pre-self-efficacy and pre-interest on their motivational outcomes. Because there is no other construct predicting identity directly apart from perceived recognition, post-self-efficacy and post-interest, the regression coefficients from the three predictors to identity are almost the same in all of the above models. This result is consistent with Godwin et al. and Kalender et al.' prior work on physics identity framework [58,81]. In addition, as shown in Fig. 2-5, as more perception of learning environment constructs are added, the residual covariance between post-self-efficacy and post-interest decreases while the residual covariance between pre-self-efficacy and pre-interest almost stays the same. This is because the residual covariance represents covariance between the constructs that is not explained by their predictors. This means that as we added more perception of learning environment constructs, more variance in post-self-efficacy and post-interest was explained.

Figure 19 shows that gender not only directly predicts the three perception of learning environment constructs but also indirectly predicts them through pre-self-efficacy and pre-interest. For example, the direct effect of gender on peer interaction is 0.15. The indirect effect can be calculated by multiplying the coefficient 0.22 (gender → pre-self-efficacy) and coefficient 0.35 (pre-self-efficacy → peer interaction). Thus, the total effect of gender on peer interaction is $0.15 + 0.22 \times 0.35 = 0.23$. Similarly, the total effect of gender on perceived recognition is $0.06 + 0.22 \times 0.32 + 0.31 \times 0.35 = 0.24$, and the total effect of gender on sense of belonging is $0.14 + 0.22 \times 0.45 = 0.24$. These results are consistent with the descriptive statistics shown in Table 27, which shows that there are statistically significant gender differences in all three perception of learning environment constructs and the effect size are all in the medium range [104].

Although there were large gender differences in students' pre- and post- self-efficacy and interest, Figure 19 shows gender mediation and clarifies that gender only directly predicts pre-self-efficacy, pre-interest, and the three learning environment constructs. Thus, Figure 19 reveals that the gender differences in students' post-self-efficacy, post-interest, and identity shown in Table 26 were mediated by the different components of the perception of the learning environment. We note that due to reasons such as societal stereotypes and biases about who belongs in physics and can excel in it, women already had lower self-efficacy and interest at the beginning of the course. Moreover, Table 26 shows that the gender differences in students' self-efficacy and interest actually increased from the beginning to the end of the course. Thus, our results indicate that the current learning environment disadvantages female students more than male students and is not equitable and inclusive.

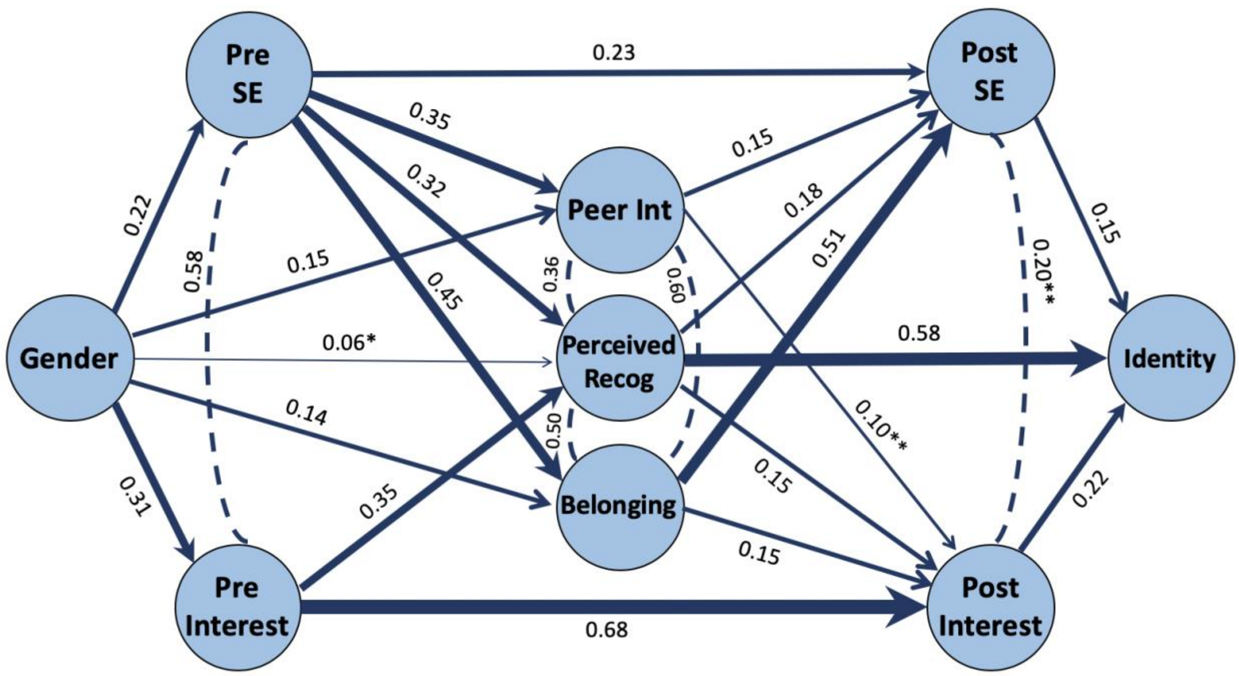



**Figure 19 Schematic diagram of the path analysis part of the structural equation modeling between gender and being a physics person through self-efficacy, interest, perceived recognition, peer interaction and sense of belonging. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of β value (standardized regression coefficient) with 0.01< p < 0.05 indicated by * and 0.001 < p < 0.01 indicated by **. Other regression lines show relations with p < 0.001.**

In this study, we investigated how students' perception of learning environment predicts their motivational outcomes by controlling for gender, self-efficacy and interest at the beginning of the course. Even though two out of the three components of pre-identity (pre-self-efficacy and pre-interest) have already been controlled for, it would still be helpful to check the model in which pre-identity is also controlled for. However, because the identity construct was added to our survey at the end of the first fall semester we studied, we do not have the pre-identity data for that semester. Thus, we checked the model including pre-identity with only the second fall semester's data (see Appendix J for results of the path analysis part of this SEM model). We find that the direct effect from pre-identity to post-identity is very small and there are no qualitative changes with regard to how the perception of learning environment predicts outcomes compared with the model without pre-identity.

To summarize how the outcome constructs were predicted through both direct and indirect paths, we calculated the regression coefficient for each path in the model that includes all three learning environment constructs (Table 28). The indirect effect of pre-self-efficacy on students' post-self-efficacy is 0.34 which is even larger than the direct effect 0.23. This means that most effect of pre-self-efficacy on post-self-efficacy was mediated by the perception of learning environment. It is noteworthy that the effect of students' sense of belonging on post-self-efficacy ($\beta = 0.51$) is almost the same as the effect of pre-self-efficacy on post-self-efficacy ($\beta = 0.57$). Unlike post-self-efficacy, students' post-interest is mainly predicted by their pre-interest directly, and the effect of perception of learning environment constructs is small. In addition, we found that self-efficacy, interest and perceived recognition are the three main predictors of identity, and perceived recognition is the major predictor, which is consistent with prior studies about physics identity [21,58,81].

**Table 28 Regression coefficients ($\beta$) of direct and indirect paths for three outcomes predicted by various predictors in the model which includes perception of learning environment (sense of belonging, peer interaction, and perceived recognition).**

| Predictor | Outcome | Direct | Indirect | Total (Direct+Indirect) |
|---|---|---|---|---|
| Pre-SE | Post-SE | 0.23 | 0.34 | 0.57 |
| Pre-SE | Post-Interest | 0.00 | 0.15 | 0.15 |
| Pre-SE | Identity | 0.00 | 0.29 | 0.29 |
| Pre-Interest | Post-SE | 0.00 | 0.06 | 0.06 |
| Pre-Interest | Post-Interest | 0.68 | 0.05 | 0.73 |
| Pre-Interest | Identity | 0.00 | 0.17 | 0.17 |
| Peer Int | Post-SE | 0.15 | 0.00 | 0.15 |
| Peer Int | Post-Interest | 0.10 | 0.00 | 0.10 |
| Peer Int | Identity | 0.00 | 0.04 | 0.04 |
| Perceived Recog | Post-SE | 0.18 | 0.00 | 0.18 |
| Perceived Recog | Post-Interest | 0.15 | 0.00 | 0.15 |
| Perceived Recog | Identity | 0.58 | 0.06 | 0.64 |
| Sense of Belonging | Post-SE | 0.51 | 0.00 | 0.51 |
| Sense of Belonging | Post-Interest | 0.15 | 0.00 | 0.15 |
| Sense of Belonging | Identity | 0.00 | 0.11 | 0.11 |

Table 28 shows that the perception of learning environment is mediating the effect of students' pre-self-efficacy and pre-interest on their motivational outcomes. To further understand the role played by each learning environment construct, we calculated the coefficients of determination $R^2$ (fraction of variance explained) for each construct in seven different SEM models with different combinations of components of the perception of learning environment (Table 29). We found in all models, input constructs have very small $R^2$ ($R^2$ for pre-self-efficacy is 0.05 and $R^2$ for pre-interest is 0.10). This is because they are input constructs that we are controlling for, which are only explained by gender. The three perception of learning environment constructs are not only explained by gender but also by pre-self-efficacy and pre-interest. In all models, the variance in peer interaction is explained 17%, the variance in perceived recognition is explained 39%, and the variance in sense of belonging is explained around 24%.

$R^2$ for all outcome constructs are pretty high, which means that our models have explained most variance in them. It is noteworthy that $R^2$ values for post-interest are around 0.79 in all of the models, which means that the models which include any of the three learning environment constructs can explain 79% of the variance of post-interest. However, there is 75% of the variance of post-self-efficacy explained by the model which only includes sense of belonging, which is larger than that explained by the other two single construct models and is very close to that explained by the model including all of the three perception of learning environment constructs. In addition, we found that the models including sense of belonging always explain more variance in post-self-efficacy than the models without sense of belonging. For example, the model including perceived recognition and sense of belonging explains 77% of variance of post-self-efficacy, and the model including peer interaction and sense of belonging also explains 77% of variance of post-self-efficacy; however, the combination of peer interaction and perceived recognition only explains 68% of variance of post-self-efficacy. Similarly, the model which only includes perceived recognition explains 74% of the variance of identity, and adding peer interaction or sense of belonging does not help explain the variance in identity further. Table 29 shows that both sense of belonging and perceived recognition play unique roles in the perception of learning environment in explaining students' motivational outcomes. However, peer interaction co-varies with sense of belonging and perceived recognition and uniquely explains very small percentages of the variance in the outcomes. However, this does not mean that peer interaction is not important. In contrast, the co-variation actually suggests a possibility that students' sense of belonging and perceived recognition can potentially be shaped by helping students interact meaningfully with peers (which in turn can improve their motivational outcomes). Thus, we believe the model including all of the three perception of learning environment constructs is productive.

**Table 29 Coefficient of determination ($R^2$) for various constructs in different models for impact of the perception of learning environment. All $R^2$ values are significant and *p* values < 0.001.**

| Construct | Models | | | | | | |
|---|---|---|---|---|---|---|---|
| | Recog | Peer | Bel | Peer+Recog | Peer+Bel | Recog+Bel | Peer+Recog+Bel |
| Pre-SE | 0.05 | 0.04 | 0.04 | 0.05 | 0.05 | 0.05 | 0.05 |
| Pre-Interest | 0.11 | 0.10 | 0.10 | 0.10 | 0.09 | 0.10 | 0.10 |
| Peer Int | / | 0.17 | / | 0.17 | 0.17 | / | 0.17 |
| Recog | 0.39 | / | / | 0.39 | / | 0.39 | 0.38 |
| Belonging | / | / | 0.24 | / | 0.25 | 0.24 | 0.25 |
| Post-SE | 0.56 | 0.61 | 0.75 | 0.68 | 0.77 | 0.77 | 0.79 |
| Post-Interest | 0.79 | 0.77 | 0.78 | 0.80 | 0.78 | 0.80 | 0.80 |
| Identity | 0.74 | 0.61 | 0.61 | 0.74 | 0.61 | 0.75 | 0.75 |

## 8.5 Summary and Discussion

We studied students' physics motivational beliefs in a calculus-based introductory physics course. In particular, we studied how the perception of learning environment—including peer interaction, perceived recognition, and sense of belonging—predicts students' physics self-efficacy, interest and identity at the end of the course by controlling for their gender as well as their self-efficacy, interest at the beginning of the course.

We found that both male and female students' self-efficacy and interest dropped from pre to post but female students' dropped even more than male students' (Table 26). In addition, we found significant gender differences disadvantaging female students in all motivational constructs in our models with the effect sizes for all of them in the medium range (Table 26) [104]. The gender differences in students' perception of learning environment may partially explain the finding that the gender differences in students' self-efficacy and interest increased by the end of the course. These results show that the current learning environment influenced students' motivational belief in a negative way and was even more detrimental to female students' feelings.

To further understand how the perception of learning environment predicts students' motivational outcomes, we performed structural equation modeling (SEM) to analyze predictive relationships among the constructs. Our results show that the perception of learning environment constructs directly predict students' motivational outcomes even after controlling for students' gender, pre-self-efficacy and pre-interest. In particular, we found that in the learning environment, students' sense of belonging is the major predictor of their post-self-efficacy, and students' perceived recognition is the major predictor of their physics identity (Figure 19).

Finally, we studied the role played by each of the three perception of learning environment components in explaining students' motivational outcomes. By comparing the fraction of variance in motivational outcomes explained by each model, we found that perceived recognition uniquely contributed most in explaining identity, and sense of belonging uniquely contributed most in explaining post-self-efficacy. Our results are consistent with Kalender et al.'s prior work [81,183] which showed that perceived recognition is the major predictor of identity and there is a strong correlation between students' sense of belonging and self-efficacy. However, they did not take into account students' motivational characteristics at the beginning of the course. In our study, we found that even after controlling for students' gender, pre-interest and pre-self-efficacy, the perception of learning environment still plays a very important role in predicting students' motivational outcomes. We note that even though peer interaction co-varies with sense of belonging and perceived recognition and uniquely explains very small percentages of the variance in the outcomes, this does not mean that peer interaction is not important. Many instructors may not know how to implement strategies to improve students' sense of belonging. The co-variation suggests a possibility that students' sense of belonging and perceived recognition may possibly be shaped by helping students interact meaningfully with peers (which in turn can improve their learning outcomes). Thus, we believe the model including all of the three learning environment constructs is productive.

In this study, we found that students' post-interest was mainly predicted by their pre-interest and the perception of learning environment predicted only a small amount. This means that our current learning environment barely helped students build their physics interest. Actually, both male and female students' interest in physics dropped by the end of the course. Even though our results show that the perception of learning environment had a large effect on students' post-self-efficacy, the drop in students' average self-efficacy scores after instruction shows that this overall effect of the current learning environment in introductory physics is actually negative.

Another important finding is that even though there are significant gender differences in students' post-self-efficacy, interest and identity, gender does not directly predict them. This means that all of the gender differences in students' motivational outcomes can be explained by the gender differences in the predictors. This finding indicates that we may be able to bridge the gender gap in students' self-efficacy, interest and identity by developing an inclusive and equitable learning environment. Unfortunately, our results show that the current learning environment did not bridge the gap and actually enlarged it. Due to reasons such as societal stereotypes and biases about who belongs in physics and can excel in it, women already had lower self-efficacy and interest at the beginning of the course. The current learning environment disadvantaging female students may further impact their motivational beliefs.

Our findings suggest that an inclusive and equitable learning environment is very important for helping students improve their motivational belief in physics. An ideal situation would be that all students get full scores in self-efficacy, interest and identity after the course; however, this also means that there would be no variance in these motivational outcomes at all and the effect of the perception of learning environment would also be zero. A more realistic situation is that we improve the learning environment so that everyone develops a significantly higher self-efficacy, interest and identity at the end of the course. As we found, the perception of learning environment directly predicts students' motivational outcomes at the end of the course (as shown in Figure 19), so it is reasonable to expect that we can enlarge this effect (which means the regression coefficients from the perception of learning environment to students' post-self-efficacy, interest and identity could be even larger). In addition, we should try to reduce the effect of gender, pre-self-efficacy and interest on the perception of learning environment. Thus, students can equally benefit from the inclusive and equitable learning environment regardless of the gender and what their motivational beliefs were at the beginning of the course. We note that even though we find that the perception of learning environment did not explain students' interest much, this does not mean that interest cannot be changed. People's interest can be trigged and maintained by external factors [86]. For example, effective evidence-based instructional conditions or learning environments that include group work, puzzles, computers, and so on have been found to trigger situational interest [115,272-276]. Thus, it is important for instructors and researchers to develop an engaging evidence-based learning environment in which students can develop their interest in physics.

This investigation was conducted in a traditionally taught lecture-based course. It would definitely help to incorporate more research-validated active engagement pedagogies, but that is not enough. For example, according to prior studies [75,138], 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/anxious if the environment is not equitable and inclusive. Instructors will need to create an equitable and inclusive learning environment keeping in mind how the societal stereotypes and biases about who belong in physics and who can excel in physics impacts the stereotyped groups. Some social-psychological interventions such as value-affirmation intervention and ecological-belonging intervention have been shown to reduce gender gaps in students' performance [144,180]. However, further studies are needed to understand whether they help to improve students' perceptions of learning environment and motivational outcomes in a course.

In our study, responses to the survey were all self-reported by students. It would be helpful to interview more students to get a deeper qualitative understanding of what they had experienced during the learning process and how the perception of learning environment influenced their learning behavior and motivational characteristics. In addition, it would also be interesting to look at the relation between the perception of learning environment and students' motivational outcomes in algebra-based physics courses. Is the self-efficacy and interest of female students higher in algebra-based physics courses than that in calculus-based physic courses? Are the gender differences in students' motivational beliefs smaller in algebra-based physics courses because female students are overrepresented in these courses? Even though the general societal stereotype threat still exists even for female students in algebra-based physics courses, if their motivational beliefs can be protected by a classroom where they have many female peers, this may itself be a useful finding.

# 9.0 How Inclusiveness of Learning Environment Mediates Gender Differences in Students' Physics Motivational Beliefs and Grades

## 9.1 Introduction

Prior studies have shown that women are often underrepresented in many science, technology, engineering, and mathematics (STEM) courses and disciplines [1-12]. For example, even though women earn approximately 60% of all bachelor's degrees in the US, only 20% of the physics undergraduate degrees are earned by women [13]. In addition, several studies have also reported gender disparity in students' performance in some STEM disciplines [14,15,188]. Moreover, prior studies also showed that female students leave STEM fields at higher rates than male students [9,277]. These studies suggest that we are largely missing out on the talents of half of the population, which not only hinders the development of STEM fields because of the loss of talent and diversity, but also hinders women from pursuing many great career opportunities. Therefore, efforts to promote participation, achievement, and continuation of women in STEM fields are important for the development of both individuals and the society as a whole. Some prior studies suggest that individuals' performance in STEM can be influenced by their motivational beliefs such as self-efficacy, interest and identity in that domain [18-23,34,35,41,43,45,256,278-281]. Students from underrepresented groups in STEM such as women may not have enough encouragement and role models to help them develop strong motivational beliefs in STEM. In addition, the societal stereotypes and biases in STEM may further undermine their motivational beliefs and lead to withdrawal from STEM courses, majors or careers [9-11,24-32,74,262]. Thus, investigation of students' motivational beliefs is important for better understanding the underrepresentation of women and minority students in STEM, and can be useful for formulating guidelines for developing an inclusive learning environment and promoting diversity and equity in STEM fields.

By inclusive learning environment, we refer to an environment in which all students feel welcome, valued, and supported. By equity in learning, we mean that not only should all students have adequate opportunities and access to resources, and have an inclusive learning environment with appropriate support and mentoring so that they can engage in learning in a meaningful and enjoyable manner, but the course outcomes should be equitable. Therefore, inclusiveness is necessary but not sufficient for equity since inclusiveness does not guarantee equitable course outcomes. By equitable course outcomes, we mean that students from all demographic groups (e.g., regardless of their gender identity or race/ethnicity) who have the prerequisites to enroll in the course, on average, have comparable outcomes, which is consistent with Rodriguez et al.'s equity of parity model [33]. The STEM course outcomes include student performance and their STEM motivational beliefs at the end of the courses because regardless of the performance, the motivational beliefs can influence students' short and long-term retention in STEM disciplines [34,35]. We note that adequate opportunity and access to resources, inclusive learning environment and equitable outcomes are strongly entangled with each other. For example, if the learning environment is not inclusive, the outcomes are unlikely to be equitable. In this study, we aim to understand how students' perception of the inclusiveness of the learning environment predicts their course outcomes including both academic performance and motivational beliefs.

### 9.1.1 Students' Motivational Beliefs Related to STEM Learning

The Expectancy-Value Theory (EVT) [53,54] is one of the most prominent approaches to the study of students' motivational beliefs. In the EVT, expectancy refers to students' belief in their ability to succeed in a given task [54]. Value refers to the subjective task value for students, which can be differentiated into four components: intrinsic value, attainment value, utility value, and cost [54]. Intrinsic value refers to students' interest in the task and the enjoyment they experience from performing the task. Attainment value reflects how important students themselves feel it is for them to develop mastery and do a good job in the field [54]. Utility value pertains to students' perception of whether the task can help them achieve some other goals [54]. The last value component is cost, which refers to the assessments of how much effort and time will be taken to engage in the task as well as the amount of opportunity cost and stress caused by the task [54]. In the EVT, students' learning goals, academic engagement and performance, and persistence in a field are impacted by their expectancy of success and the four components of value [54].

The expectancy component of EVT is closely related to the concept of self-efficacy in Bandura's social cognitive theory, which is defined as one's belief in one's ability to succeed in a specific area or accomplish a task [39,40]. Prior research suggests that self-efficacy is an important motivational belief of students for them to excel in a domain [20,36-38]. Studies have shown that students' engagement and performance can be influenced by their self-efficacy [41-44,282]. For example, students who have high self-efficacy tend to see difficulties as challenges and believe that productive struggles can help them improve, so they often choose to take harder courses and ask to do more challenging problems than students with low self-efficacy, who usually see difficulties as threats and obstacles to success [45].

Another important motivational belief is interest, which refers to students' curiosity, enjoyment and engagement in a specific area [46,47]. Interest is closely related to intrinsic value in EVT. Studies have shown that interest can also influence students' learning [41,47-51,283]. For example, one study showed that students' performance can be improved by connecting physics courses to students' daily lives or using evidence-based curricula to make the courses more engaging and interesting [52]. Prior studies (both experimental and correlational) have also shown that interest can be affected by self-efficacy [160,284]. Some other studies show that interest may also lead to the development of self-efficacy [285,286].

In addition, students' identity in a specific field such as physics is another important motivational belief that influences their career decisions and outcome expectations [21,22,55-62]. Students' physics identity is related to whether they see themselves as a physics person [21,22,55,58,63]. Some studies have found that female students often report lower physics identity than male students [63-65]. This gender difference in physics identity has been shown to be related to societal biases and stereotypes about who belongs in and can succeed in physics [66-68]. In physics and other STEM fields, these stereotypes can negatively influence women's experiences, which may lower their sense of belonging and identity and lead to withdrawal from STEM fields [24,69,70]. Therefore, investigating students' STEM identity may help us understand the gender difference in participation in STEM. We now turn to our theoretical framework for investigating the factors that can affect students' identity, self-efficacy, interest and their academic performance.

### 9.1.2 Theoretical Framework

In Carlone and Johnson's (2007) science identity framework, students' science identity includes three interrelated constructs: competence (belief in one's competence), performance (belief in ability to perform), and recognition (recognition of self and by others as a "science person"). Hazari et al. (2010) adapted this model to physics and added interest to this model. In addition, Hazari et al. (2010) developed quantitative measures for these constructs and found that competence and performance factored into a single construct. Moreover, they separated recognition of self and by others and used a single item ("I see myself as a physics person") to measure students' overall physics identity [21]. In their later studies using structural equation modeling, they found that students' overall physics identity was predicted by the physics identity constructs (interest, competence/performance, and perceived recognition from other people) [58,90,124,125]. The advantage of keeping all identity constructs in structural equation modeling is that one can investigate the predictive relationships between each identity construct and overall identity and other factors (e.g., grade) as well as the predictive relationships among the identity constructs. This physics identity framework has been used to study physics identity of students in high school physics classes [287,288] as well as college students with a variety of majors [64,125,147,289,290], and studies have shown that students' physics identity is an important predictor of their engineering identity [58,124]. These studies reveal that individuals' overall identity in STEM fields is not only impacted by their own motivational characteristics but also by their perceived recognition from others.

Several studies have shown that female students did not feel that they were recognized appropriately even before they entered college [66,83,183]. One stereotypical view of science is that it is for students who are very smart or have a natural gift in science [66]. In general, due to societal stereotypes, being brilliant or exceptionally smart is usually associated with boys [156]. One investigation showed that the gendered notions of brilliance are endorsed by children as young as 6 years old and have an immediate effect on their interest and identity in science [83]. These stereotypes and biases also exist in the university context [81,85,112]. For example, one study showed that science faculty participants rated men as more competent and would like to offer higher starting salary and more mentoring to male applicants than the (identical) female applicants even though only the names were different in the hypothetical information they were provided [85]. For female students, the experiences of not being recognized as a science person and the gender-based biases may accumulate over time and negatively influence their science identity.

Moreover, students' interest and self-efficacy have also been found to be connected to their interaction with other people and recognition by them [40,47]. In the four-phase model of interest development, Hidi and Renninger pointed out that external factors such as group work and tutoring can trigger and maintain people's interest [47,49,86]. In addition, according to Bandura's social cognitive theory, an individual's self-efficacy can be shaped by verbal encouragement from others [151,153]. Authors' physics identity framework [65,81] showed that students' perceived recognition not only strongly predicts their overall physics identity, but also predicts their physics interest and self-efficacy.

In addition to perceived recognition, students' sense of belonging and their interaction with peers have also been shown to be important aspects of the inclusiveness of the learning environment [2,69,70,195,263-265]. For example, if students have a high sense of belonging in class, they may interact with others more and with more positive attitudes, and they may also develop a higher perceived recognition [178]. In our previous study (Authors, 2021), we found that students' sense of belonging, perceived effectiveness of peer interaction, and perceived recognition in an introductory physics course statistically significantly predicts their motivational beliefs such as self-efficacy, interest at the end of a physics course after controlling for their motivational beliefs at the beginning of the course. As noted earlier, in addition to motivational beliefs, students’ academic performance at the end of a physics course is also a very important course outcome. Although prior studies have shown that perceived recognition, sense of belonging, and peer interaction are important for student learning, it is not well understood how these three constructs together predict students’ course outcomes (including both academic performance and motivational beliefs) after controlling for students’ academic preparation, what role is played by each one of them in predicting students’ course outcomes after controlling for the other two, and whether there are any gender differences in these predictive relationships. Moreover, very few studies have investigated how students’ motivational beliefs evolve in a two-semester course sequence and the role played by perceived recognition, sense of belonging, and peer interaction in this evolution. Answering these questions is important for developing a deeper understanding of how to improve students’ course outcomes by cultivating an inclusive learning environment in which all students can thrive.

### 9.1.3 The Present Study

In this study, we include students’ self-reported perceived recognition by instructors and teaching assistants, sense of belonging, and perceived effectiveness of peer interaction as three aspects of students’ perception of the inclusiveness of the learning environment, and we used quantitative methods to investigate how the perceived inclusiveness of the learning environment predicts students’ course outcomes in a calculus-based introductory physics sequence (including physics 1 and physics 2) at a large state-related university in the US. In this study, we include students’ academic performance (measured by course grades) and motivational beliefs (including physics self-efficacy, physics interest, and overall physics identity) at the end of the physics sequence as course outcomes. Specifically, we try to address the following research questions:

**RQ1.** Are there gender differences in students' academic performance and motivational beliefs and do they change from physics 1 to physics 2?

**RQ2.** How do the different components of the perception of the inclusiveness of the learning environment in physics 2 (including perceived recognition, sense of belonging, and perceived effectiveness of peer interaction) predict students' academic performance and motivational beliefs in physics 2 after controlling for students' gender, high school preparation, and their performance and motivational beliefs in physics 1?

**RQ3.** If gender does not moderate any predictive relationship in RQ2 (the regression coefficients among the constructs are not different for women and men), how does gender directly or indirectly predict

a. students' high school preparation and their academic performance and motivational beliefs in physics 1?
b. the perception of the inclusiveness of the learning environment after controlling for students' high school preparation and their academic performance and motivational beliefs in physics 1?
c. students' academic performance and motivational beliefs in physics 2 after controlling for everything else in RQ2?

This study was conducted in a two-term college calculus-based introductory physics sequence (including physics 1 and physics 2) at a large public university. These courses are generally mandatory and taken by students majoring in engineering, physical science, and mathematics in their first year of university. Physics 1 mainly includes mechanics, while the main content of physics 2 is electricity and magnetism. In our prior work, we found that students' physics motivational beliefs decreased from the beginning to the end of physics 1, and these changes were mediated by the inclusiveness of the learning environment in physics 1 [136,261]. Students' physics motivational beliefs may further change for better or worse from physics 1 to physics 2 based upon the inclusiveness of the learning environment, e.g., depending on whether students had a high sense of belonging in the course, whether they felt recognized, e.g., by instructors and teaching assistants and whether their interaction with peers was positive. We note that for most students in the calculus-based introductory physics sequence, physics 2 might be their last formal physics course in college, so their motivational beliefs at the end of physics 2 are very important not only for their engagement in the following courses in physical science and engineering, but also for their short and long-term academic goals in STEM disciplines.

Therefore, in this study, we investigated the effect of the perception of the inclusiveness of the learning environment (including students' sense of belonging, perceived effectiveness of peer interaction, and perceived recognition) on students' grades and motivational beliefs (including physics self-efficacy, physics interest, and overall physics identity) in physics 2 after controlling for students' gender and pre-college test scores (including high school Grade Point Average (GPA) and Scholastic Assessment Test (SAT) math scores) as well as their self-efficacy, interest, and grades in physics 1. For convenience, perceived effectiveness of peer interaction is shortened to peer interaction in the rest of the paper. We note that the learning environment here is not only the classroom environment but also includes students' experiences outside the class. For example, students may work together on their homework after class, and they could also ask for help during TAs'/instructors' office hours.

As shown in Figure 20, the thirteen constructs are divided into three groups: what we control for, perception of the inclusiveness of the learning environment, and outcomes. Students' gender, SAT math scores, high school GPA (HS GPA), and their self-efficacy, interest and grades in physics 1 (Self-efficacy (SE) 1, Interest 1, and Grade1) are constructs that we control for. Outcomes include students' self-efficacy, interest, overall physics identity and grades in physics 2 (Self-efficacy (SE) 2, Interest 2, Identity, and Grade 2). Perceived recognition (Perceived Recog), peer interaction (Peer Int) and sense of belonging (Belonging) constitute the perception of the inclusiveness of the learning environment. It is expected that students' responses to the motivational survey in physics 1 and physics 2 are correlated because they are students' responses to the same questions pertaining to the same motivational construct at two different time points. However, if students' motivational beliefs changed from physics 1 to physics 2, we want to study whether the perception of the inclusiveness of the learning environment helps to explain the changes and what role is played by each construct in the inclusiveness of the learning environment. In addition, since previous studies suggest that self-efficacy and interest can influence student learning [41,44,47,49], we also model a direct path from self-efficacy and interest to grade in both physics 1 and physics 2.

In this study, we first investigated how students' motivational beliefs changed from physics 1 to physics 2 and whether there were gender differences in the constructs studied. Then, we used Structural Equation Modeling (SEM) to study the effect of the perception of the inclusiveness of the learning environment on students' motivational beliefs and grades in physics 2 after controlling for students' gender, high school GPA, SAT math scores, and their motivational beliefs and grades in physics 1.

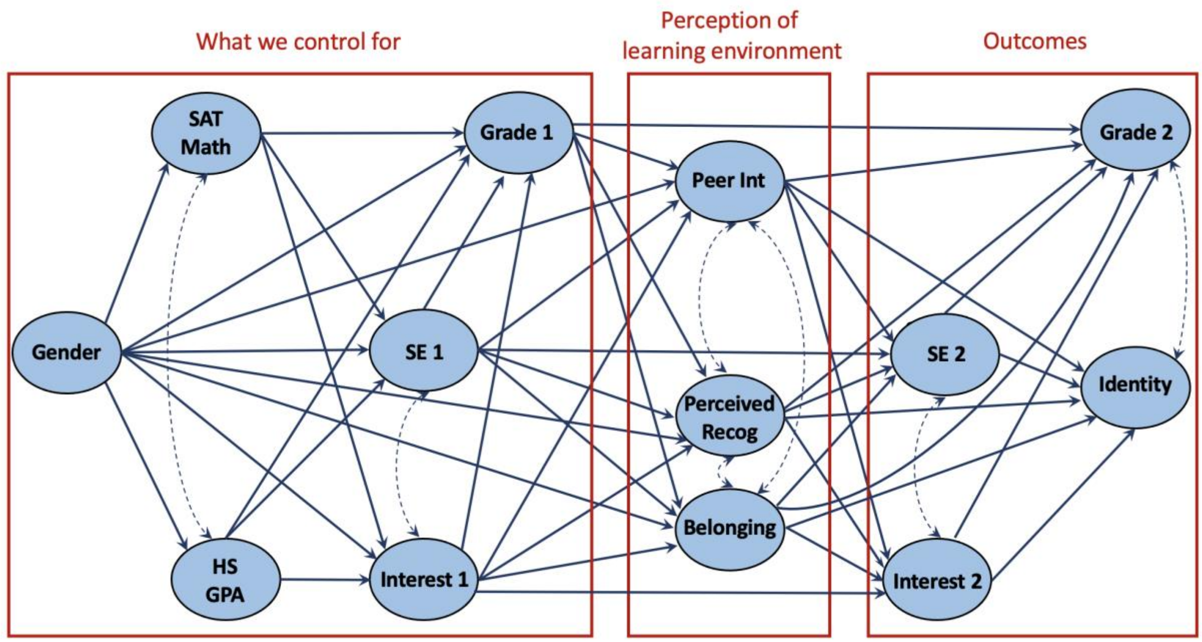



**Figure 20 Schematic representation of the theoretical framework. The solid lines represent regression paths, and the dashed lines represent covariances. From left to right, all possible regression paths were considered, but only some of the paths are showed here for clarity.**

## 9.2 Methods

### 9.2.1 Participants and Data Sources

The motivational survey data used in this study were collected at the end of each course of a two-term college calculus-based introductory physics sequence (including physics 1 and physics 2) in two consecutive school years at a large research university in the US. These courses are taken mostly by students in engineering school and school of arts and sciences (students majoring in physical sciences and mathematics) for whom they are mandatory. Physics 1 mainly includes mechanics, while the main content of physics 2 is electricity and magnetism. The paper surveys were handed out and collected by TAs during the last recitation class of a semester. In particular, students' self-efficacy and interest in physics 1 and physics 2 were measured at the end of each course, and their perceived recognition, peer interaction, sense of belonging and overall physics identity were measured at the end of physics 2. This is because only after the course can students answer the survey questions pertaining to inclusiveness of the learning environment based on their real experience in the course such as their interaction with peers, TAs and instructors. Table 30 shows when each construct was assessed throughout the course sequence. The demographic data of students—such as gender—were provided by the university. Students' SAT math scores, high school GPA, and course grades in physics 1 and physics 2 were also obtained from the university records. Students' names and IDs were de-identified by an honest broker who generated a unique new ID for each student. Thus, researchers could analyze students' data without having access to students' identifying information.

**Table 30 Time points when different constructs were assessed.**

| Constructs | When the constructs were assessed |
|---|---|
| Gender, SAT Math, High School GPA | Pre-college |
| Grade 1, Self-efficacy 1, Interest 1 | At the end of Physics 1 (December) |
| Peer interaction, Perceived recognition, Sense of Belonging | At the end of Physics 2 (April) |
| Grade 2, Self-efficacy 2, Interest 2 | At the end of Physics 2 (April) |

There were 1203 students in physics 1 and 921 students in physics 2 participating in the survey. However, in this study, we only focused on 697 students (233 female students and 464 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 and academic performance 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 directly enrolling 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. As noted, the participants of this study were undergraduate students in engineering school and school of arts and sciences and most of them were in their first year of university when the study was conducted. Students in engineering school usually declare their majors in the first year, while most students in the school of arts and sciences usually do not declare their majors until the second year. In this study, 524 of the students are engineering majors. For the 173 students who are in school of arts and sciences, only 56 of them had declared their majors when the study was conducted, and the rest of them were undeclared majors. The average age of the participants of this study was 18 years old. Only 11 students were above 20 years old and only 1 student was above 30 years old.

### 9.2.2 Survey Instruments

In this study, our analysis includes three motivational constructs (physics self-efficacy, physics interest, and overall physics identity) and three perception of the inclusiveness of the learning environment constructs (peer interaction, perceived recognition, and sense of belonging). The survey questions for each construct are shown in Table 31. We adapted these questions from existing motivational research [21,92-94,164] and revalidated them in our prior work [37,65,95,96,127,128,165]. The validation and refinement of the survey involved use of individual interviews with students [37,76,95,96], exploratory and confirmatory factor analysis (EFA and CFA) [97], Pearson correlation between different constructs and Cronbach alpha [99,100].

Physics self-efficacy represents students' belief about whether they can excel in physics. In our survey, we had four items for self-efficacy [92-94] (Cronbach's $\alpha$ = 0.79 for self-efficacy in physics 1 and $\alpha$ = 0.81 for self-efficacy in physics 2 [99]). These items each had four options "NO!, no, yes, YES!", which is a 4-point Likert scale (1-4). We also had four items for physics interest [92,94] (Cronbach's $\alpha$ = 0.82 for interest in physics 1 and $\alpha$ = 0.84 for interest in physics 2). For the item "I wonder about how physics works", students can choose from "Never, Once a month, Once a week, Every day". For the item "In general, I find physics", students can choose from "very boring, boring, interesting, very interesting". The remaining two items under interest had a response scale of "NO!, no, yes, YES!". By choosing the four options, students will get a score from 1 to 4 respectively. For example, if a student finds physics very boring, they will get one point for this item. The more interest a student has in physics, the higher score the student will have for this item. We had one item for overall physics identity [21], which corresponds to students' belief about whether they designate themselves as a physics person. This item had response options "strongly disagree, disagree, agree, and strongly agree", which correspond to 1 to 4 points [102].

In addition, perceived recognition, peer interaction and sense of belonging are the perception of the inclusiveness of the learning environment constructs in our study. Unlike self-efficacy, interest and overall physics identity, these three constructs are directly related to students' experience in the course. Perceived recognition (included in perception of the inclusiveness of the learning environment) included one item which represents whether a student thinks their instructors or TAs see them as a physics person [21,63,101]. Peer interaction (which includes four items) [93] represents whether students have a productive and enjoyable experience when working with peers (Cronbach's $\alpha$ = 0.92). Sense of belonging is about students' feelings of whether they belonged in the physics class [195], and it included five items [164] that each had a 5-point Likert scale: "not at all true, a little true, somewhat true, mostly true and completely true" (Cronbach's $\alpha$ = 0.87). Two sense of belonging items ("I feel like an outsider in this class" and "Sometimes I worry that I do not belong in this physics class") were reverse coded, which means that a higher score in these two items represents a lower sense of belonging. Students' score on each construct is the average score of all items in that construct.

**Table 31 Survey items for each of the motivational scales. The Cronbach alphas and CFA item loadings (Lambda and *p*-values of the significance test for each item loading) shown here were calculated with physics 2 data. †The response options for this question are "Never, Once a month, Once a week, Every day". ‡The response options for this question are "very boring, boring, interesting, very interesting".**

| **Construct and Item** | Lambda | *p* value |
|---|---|---|
| **Overall Physics Identity** | | |
| I see myself as a physics person. | 1.000 | <0.001 |
| **Physics Self-Efficacy** (Cronbach's $\alpha$ = 0.81) | | |
| I am able to help my classmates with physics in the laboratory or in recitation. | 0.707 | <0.001 |
| I understand concepts I have studied in physics. | 0.744 | <0.001 |
| If I study, I will do well on a physics test. | 0.729 | <0.001 |
| If I encounter a setback in a physics exam, I can overcome it. | 0.723 | <0.001 |
| **Physics Interest** (Cronbach's $\alpha$ = 0.84) | | |
| I wonder about how physics works † | 0.706 | <0.001 |
| In general, I find physics ‡ | 0.804 | <0.001 |
| I want to know everything I can about physics. | 0.826 | <0.001 |
| I am curious about recent physics discoveries. | 0.708 | <0.001 |
| **Physics Perceived Recognition** | | |
| My physics TA and/or instructor see me as a physics person. | 1.000 | <0.001 |
| **Sense of Belonging** (Cronbach's $\alpha$ = 0.87) | | |
| I feel like I belong in this class. | 0.812 | <0.001 |
| I feel like an outsider in this class. | 0.710 | <0.001 |
| I feel comfortable in this class. | 0.835 | <0.001 |
| I feel like I can be myself in this class. | 0.689 | <0.001 |
| Sometimes I worry that I do not belong in this physics class. | 0.705 | <0.001 |
| **Physics Peer Interaction** (Cronbach's $\alpha$ = 0.92) | | |
| **My experience and interaction with other students in this class…** | | |
| made me feel more relaxed about learning physics. | 0.734 | <0.001 |
| increased my confidence in my ability to do physics. | 0.928 | <0.001 |
| increased my confidence that I can succeed in physics. | 0.930 | <0.001 |
| increased my confidence in my ability to handle difficult physics problems. | 0.873 | <0.001 |

### 9.2.3 Analysis

First, we calculated the mean score for each construct for each student. Then we used a *t*-test [103,240] to compare students' responses in physics 1 and physics 2 and to compare responses for female and male students. Then, we conducted Structural Equation Modeling (SEM) [87] using the "lavaan" package in software R [110] to study how the perception of the inclusiveness of the learning environment predicted students' motivational and academic outcomes in physics 2 after controlling for students' gender, high school GPA and SAT math as well as their self-efficacy, interest and grades in physics 1.

The SEM includes two parts: confirmatory factor analysis (CFA) and path analysis. First, we performed the CFA for each construct. The CFA model fit is considered adequate if the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are >0.9 and Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) are <0.08 [98]. In our study, CFI = 0.933, TLI = 0.918, RMSEA = 0.054 and SRMR = 0.038, which represents a good fit. This result provides quantitative support for us to divide the motivational constructs and the inclusiveness of the learning environment constructs as proposed. In addition, as shown in Table 31, all of the CFA item loadings are above 0.5 and most of them are above 0.7, which means that our constructs extract sufficient variance from the items [239].

Before performing the path analysis, we calculated the Pearson correlation coefficients pairwise between each pair of constructs [100]. As shown in Table 32, there are relatively strong correlation among non-academic constructs, while the correlation between non-academic constructs and SAT math or high school GPA are relatively small. Even though these non-academic constructs have strong correlations with each other, the correlations are not so high that SEM cannot examine the constructs separately [109]. We note that in Table 32, there are two very strong correlations. The correlation coefficient between Interest 1 and Interest 2 is 0.88, which means that students' interest in physics 2 is highly related to their interest in physics 1. Another large coefficient is for correlation between students' Self-efficacy 2 and sense of belonging. According to the prior work done by Authors, these two constructs are indeed strongly correlated with each other even though they are separate constructs [65,183].

**Table 32 Zeroth order correlation coefficients of the constructs studied. *p* values are indicated by $^{**}$ for $0.001 \leq p < 0.01$, $^{*}$ for $0.01 \leq p < 0.05$, and $^{ns}$ for $p > 0.05$. All the other correlation coefficients have $p < 0.001$.**

| Observed Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. SAT math | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 2. High school GPA | 0.17 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 3. Overall Physics Identity | $0.08^{*}$ | $-0.09^{*}$ | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 4. Grade 1 | 0.29 | 0.21 | 0.28 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 5. Self-efficacy 1 | $0.09^{ns}$ | $-0.01^{ns}$ | 0.56 | 0.47 | -- | -- | -- | -- | -- | -- | -- | -- |
| 6. Interest 1 | $0.02^{ns}$ | $-0.13^{**}$ | 0.57 | 0.21 | 0.59 | -- | -- | -- | -- | -- | -- | -- |
| 7. Grade 2 | 0.25 | 0.21 | 0.31 | 0.59 | 0.30 | 0.18 | -- | -- | -- | -- | -- | -- |
| 8. Self-efficacy 2 | $0.07^{ns}$ | $-0.02^{ns}$ | 0.71 | 0.33 | 0.75 | 0.48 | 0.41 | -- | -- | -- | -- | -- |
| 9. Interest 2 | $0.04^{ns}$ | $-0.07^{ns}$ | 0.67 | 0.21 | 0.46 | 0.88 | 0.25 | 0.64 | -- | -- | -- | -- |
| 10. Perceived Recognition | $0.11^{**}$ | $-0.03^{ns}$ | 0.65 | 0.20 | 0.45 | 0.42 | 0.30 | 0.66 | 0.52 | -- | -- | -- |
| 11. Peer Interaction | $0.06^{ns}$ | $-0.04^{ns}$ | 0.53 | 0.18 | 0.47 | 0.37 | 0.28 | 0.73 | 0.52 | 0.56 | -- | -- |
| 12. Sense of Belonging | $0.09^{*}$ | $0.02^{ns}$ | 0.60 | 0.29 | 0.58 | 0.34 | 0.38 | 0.83 | 0.52 | 0.61 | 0.64 | -- |

To analyze the relationships among the constructs, we performed the full Structural Equation Modeling (SEM). Apart from CFA, the path analysis part of SEM estimates the predictive relationships between different constructs. The strength of each relationship is represented by a regression coefficient $\beta$. One advantage of SEM is that it simultaneously estimates factor loadings for items and all of the regression links for multiple outcomes, which improves the statistical power compared with other statistical methods such as multiple regression. Another advantage of SEM is that it shows not only the direct regression relation between two constructs but also all of the indirect relations mediated through other constructs, which allowed us to calculate the total regression effect by summing the direct and indirect regression coefficients. The level of SEM model fit can also be represented by CFI, TLI, RMSEA and SRMR. We first analyzed the saturated SEM model that includes all of the possible links from left to right between different constructs shown in Figure 20, and then we removed the most insignificant path line (with the highest *p* value) and re-ran the model. We used this method to trim one path at a time until all remained path lines were statistically significant. Next, we used modification indices to improve the model fit. The modification index is the chi-square value, with 1 degree of freedom, by which model fit would improve if a particular path was added back. Modification index bigger than 3.84 indicates that the model fit would be significantly improved, and the *p* value for the added parameter would be $< 0.05$ [291,292]. We added back the paths with modification index larger than 3.84 one at a time (from high to low modification index) to improve the model fit. Finally, we checked the statistical significance of each trimmed path by adding them back to make sure that all trimmed paths are not statistically significant and all statistically significant paths are kept.

We also tested measurement invariance (which tests whether the survey items were interpreted in a similar manner by male and female students) and performed gender moderation analysis using multi-group SEM (which tests whether the regression pathways were different across gender). Results showed that strong measurement invariance holds for our model, and regression pathways among the constructs do not have differences across gender. Therefore, we concluded that our SEM model can be interpreted similarly for men and women, and any gender differences can be modeled using a separate gender variable (1 for male and 0 for female) as an exogenous variable as in Figure 20. If there are statistically significant paths from gender to any of the constructs in the model, it implies that women and men did not have the same average value for those constructs controlling for all constructs to the left. This is the gender mediation SEM model, which we will discuss in more detail in the results section. In addition, we also tested measurement invariance between students in engineering school and school of arts and sciences, and our results show that measurement invariance also holds for our model. (see Supplemental Material for detailed results of testing measurement invariance and multi-group SEM analysis).

## 9.3 Results

### 9.3.1 Gender Differences in Motivational Characteristics and Grades

Table 33 shows the descriptive statistics of students' motivational characteristics in physics 1 and physics 2. For all of the *t*-tests conducted, the difference in means was considered significant if the *p*-value was below 0.05/28 = 0.0018, which was determined using a Bonferroni correction because 28 *t*-tests were performed [293]. We note that female students had significantly lower scores in all of the six constructs in both physics 1 and physics 2, and the effect sizes are all in the medium range [240]. These results indicate that, in the current learning environment, female students reported less benefit from peer interaction and also felt a lower sense of belonging than male students. Moreover, female students' average scores pertaining to perceived recognition and overall physics identity indicate that on average, female students did not think their instructors/TAs see them as a physics person, and they did not see themselves as a physics person either. Furthermore, the gender differences in students' perceived recognition increased from physics 1 to physics 2.

When we compared students' perception of the inclusiveness of the learning environment and motivational characteristics in the two courses, we found that, from physics 1 to physics 2, there was no statistically significant change in students' perceived recognition and overall physics identity. Even though both male and female students' self-efficacy and male students' interest, peer interaction and sense of belonging significantly decreased from physics 1 to physics 2, the effect sizes are relatively small compared with the effect sizes of the gender differences in these constructs.

Table 34 shows students' high school GPA, SAT math scores, and grades in physics 1 and physics 2. We note that even though female students had significantly lower grades than male students in both physics 1 and physics 2, there was no statistically significant gender difference in SAT math scores, and female students even had a higher average high school GPA than male students.

**Table 33 Descriptive statistics of female and male students' motivational characteristics in physics 1 and physics 2. N = 233 for female students N = 464 for male students. Cohen suggested that typically values of *d* = 0.2, 0.5 and 0.8 represent small, medium and large effect sizes. The minus sign indicates male students have higher a perceived recognition average score in physics 2 than in physics 1. **p*-value < 0.0018.**

| Gender | **Self-efficacy** | | **Statistics** | | **Interest** | | **Statistics** | |
|---|---|---|---|---|---|---|---|---|
| | 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* | | |

| Gender | **Perceived Recognition** | | **Statistics** | | **Peer Interaction** | | **Statistics** | |
|---|---|---|---|---|---|---|---|---|
| | physics 1 | physics 2 | *p* value | Cohen's *d* | physics 1 | physics 2 | *p* value | Cohen's *d* |
| Male | 2.53 | 2.55 | 0.711 | -0.02 | 3.09 | 2.95 | 0.001 | 0.22* |
| Female | 2.21 | 2.14 | 0.360 | 0.09 | 2.79 | 2.64 | 0.022 | 0.22 |
| *p* value | <0.001 | <0.001 | | | <0.001 | <0.001 | | |
| Cohen's *d* | 0.44* | 0.55* | | | 0.49* | 0.46* | | |

| Gender | **Sense of Belonging** | | **Statistics** | | **Overall Physics Identity** | | **Statistics** | |
|---|---|---|---|---|---|---|---|---|
| | physics 1 | physics 2 | *p* value | Cohen's *d* | physics 1 | physics 2 | *p* value | Cohen's *d* |
| Male | 3.92 | 3.75 | 0.001 | 0.22* | 2.75 | 2.68 | 0.190 | 0.09 |
| Female | 3.54 | 3.37 | 0.037 | 0.19 | 2.29 | 2.16 | 0.100 | 0.16 |
| *p* value | <0.001 | <0.001 | | | <0.001 | <0.001 | | |
| Cohen's *d* | 0.49* | 0.45* | | | 0.57* | 0.63* | | |

**Table 34 Descriptive statistics of students' high school GPA, SAT math, and grades in physics 1(Grade 1) and physics 2 (Grade 2). N = 233 for female students N = 464 for male students. Cohen suggested that typically values of *d* = 0.2, 0.5 and 0.8 represent small, medium and large effect sizes. The minus sign indicates female students have higher scores than male students. **p*-value < 0.0018.**

| Grades (Score Range) | Mean | | p value | Cohen's d |
|---|---|---|---|---|
| | Male | Female | | |
| High School GPA (0-5) | 4.20 | 4.34 | < 0.001 | -0.34* |
| SAT Math (400-800) | 713 | 706 | 0.130 | 0.13 |
| Grade 1 (0-4) | 2.93 | 2.74 | 0.001 | 0.28* |
| Grade 2 (0-4) | 2.73 | 2.48 | < 0.001 | 0.31* |

### 9.3.2 Perception of the Inclusiveness of the Learning Environment Mediation Models Using SEM

In this section, we show the predictive relationships among the constructs using Structural Equation Modeling (SEM). We ran the full SEM model in which perceived recognition, peer interaction and sense of belonging constitute the perception of the inclusiveness of the learning environment to study how these inclusiveness of the learning environment constructs predict students' motivational beliefs and grades in physics 2 after controlling for students' gender, high school GPA, SAT math scores, and their motivational beliefs and grades in physics 1 (see Supplemental Material for results of other SEM models with only one or two inclusiveness of the learning environment constructs if one is interested in comparing some other SEM models). The results of the SEM model are presented visually in Figure 21. The model fit indices suggest a good fit to the data: CFI = 0.929 (>0.90), TLI = 0.919 (>0.90), RMSEA = 0.053 (<0.08) and SRMR = 0.040 (<0.08).

As shown in Figure 21, students' course outcomes at the end of physics 2 are statistically significantly predicted by the perception of the inclusiveness of the learning environment. In particular, students' self-efficacy in physics 2 is predicted by all three inclusiveness of the learning environment constructs, interest in physics 2 is predicted by peer interaction and sense of belonging, Grade 2 is predicted by perceived recognition and sense of belonging, and overall physics identity is predicted by perceived recognition. Figure 21 shows that Self-efficacy 1 directly predicts Grade 1, while the direct effect of Self-efficacy 2 on Grade 2 is not statistically significant after controlling for Grade 1, sense of belonging and perceived recognition. We note that the regression coefficient from sense of belonging to Self-efficacy 2 is 0.37, which is almost as large as the effect of Self-efficacy 1 on Self-efficacy 2 ($\beta = 0.39$). In addition, consistent with Godwin et al. and Authors' prior work [58,81], Figure 21 shows that overall physics identity is mainly predicted by self-efficacy, interest and perceived recognition.

As shown in Figure 21, gender directly predicts students' motivational beliefs in physics 1 and the three inclusiveness of the learning environment constructs, which are consistent with the descriptive statistics shown in Table 33, which show that there were statistically significant gender differences disadvantaging women in these constructs. Although there were also significant gender differences in students' grades, self-efficacy, interest and overall physics identity in physics 2 as shown in Table 33 and Table 34, gender does not directly predict these constructs in the SEM model. Thus, Figure 21 reveals that the gender differences in these outcome constructs were partially mediated by the different components of the perception of the inclusiveness of the learning environment. We note that gender predicts high school GPA with a negative regression coefficient ($\beta$ = -0.17), which means that female students actually had a higher average high school GPA than male students. This is consistent with the results shown in Table 34.

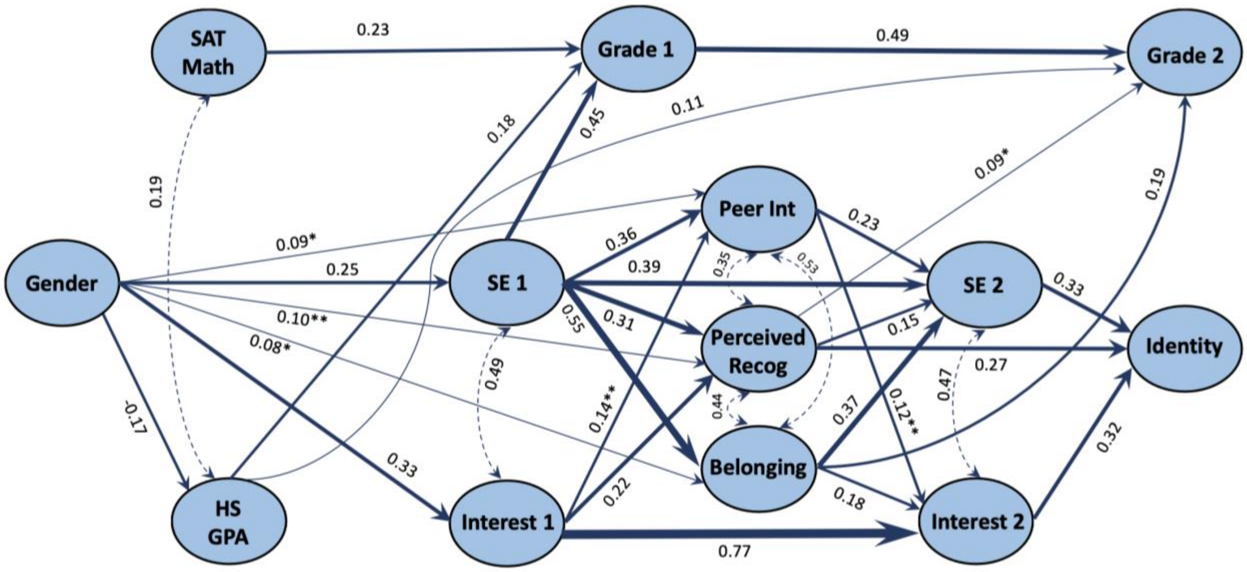



**Figure 21 Schematic diagram of the path analysis part of the SEM between gender and overall physics identity mediated through SAT Math, high school GPA (HS GPA), grade, self-efficacy (SE) and interest (in physics 1 and physics 2), perceived recognition (Perceived Recog), peer interaction (Peer Int) and sense of belonging. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.01 \leq p < 0.05$ indicated by * and $0.001 \leq p < 0.01$ indicated by **. Other regression lines show relations with $p < 0.001$.**

In addition to the direct paths shown in Figure 21, we also calculated the indirect effects of the controlled variables and the inclusiveness of the learning environment constructs on different outcome constructs (Table 35). An indirect path is measured by finding the product of each path ($\beta$) from a given predictor to an outcome construct. The total indirect effect of a given predictor on an outcome construct was found by adding all of the indirect paths together. For example, there are two indirect paths from sense of belonging to overall physics identity. The first path goes from sense of belonging $\rightarrow$ self-efficacy physics 2 $\rightarrow$ identity ($0.37 \times 0.33 = 0.12$). The second path goes from sense of belonging → interest physics 2 → identity (0.18×0.32 = 0.06). Therefore, the total indirect effect from sense of belonging to overall physics identity is 0.12 + 0.06 = 0.18. A summary of all direct and indirect effects can be found in Table 35. As we can see from Table 35, even though perceived recognition is the only inclusiveness of the learning environment construct that directly predicts overall physics identity, peer interaction and sense of belonging both have indirect effects on students' overall physics identity.

**Table 35 Regression coefficients for direct paths and indirect paths from the controlled variables and the inclusiveness of learning environment constructs to the outcome constructs. The paths that have a total regression coefficient equal to zero are not shown in the table.**

| Outcome | Predictor | Direct Path | Total Indirect Path | Total Path |
|---|---|---|---|---|
| Self-efficacy physics 2 | Self-efficacy physics 1 | 0.39 | 0.33 | 0.72 |
| | Interest physics 1 | 0.00 | 0.07 | 0.07 |
| | Peer Interaction | 0.23 | 0.00 | 0.23 |
| | Perceived Recognition | 0.15 | 0.00 | 0.15 |
| | Sense of belonging | 0.37 | 0.00 | 0.37 |
| Interest physics 2 | Self-efficacy physics 1 | 0.00 | 0.14 | 0.14 |
| | Interest physics 1 | 0.77 | 0.02 | 0.79 |
| | Peer Interaction | 0.12 | 0.00 | 0.12 |
| | Sense of belonging | 0.18 | 0.00 | 0.18 |
| Overall Physics Identity | Self-efficacy physics 1 | 0.00 | 0.37 | 0.37 |
| | Interest physics 1 | 0.00 | 0.34 | 0.34 |
| | Peer Interaction | 0.00 | 0.11 | 0.11 |
| | Perceived Recognition | 0.27 | 0.05 | 0.32 |
| | Sense of belonging | 0.00 | 0.18 | 0.18 |
| Grade physics 2 | SAT Math | 0.00 | 0.11 | 0.11 |
| | High School GPA | 0.11 | 0.09 | 0.20 |
| | Grade physics 1 | 0.49 | 0.00 | 0.49 |
| | Self-efficacy physics 1 | 0.00 | 0.35 | 0.35 |
| | Interest physics 1 | 0.00 | 0.02 | 0.02 |
| | Perceived Recognition | 0.09 | 0.00 | 0.09 |
| | Sense of belonging | 0.19 | 0.00 | 0.19 |

To further understand how much variance in students' course outcomes is explained by our model, we calculated the coefficient of determination $R^2$ (fraction of variance explained) for each of the four outcome constructs. According to our results, the $R^2$ value of Grade 2 is 0.40, which means that the model explains 40% of the variance in Grade 2. For the motivational outcomes, there is 87% of the variance in Self-efficacy 2, 82% of the variance in Interest 2, and 63% of the variance in Overall Physics Identity explained by the model. Thus, our model explains students' course outcomes reasonably well.

## 9.4 Discussion

Prior studies have shown that students' motivational beliefs such as self-efficacy, interest and identity can influence their persistence and retention in STEM fields such as physics [20,37,41,44,45,48,63,73]. According to prior quantitative studies [38,135,136], female students usually report lower physics self-efficacy, interest, and overall physics identity than their male peers, which can partially explain the underrepresentation of women in physics. Even though there have been many efforts to understand the factors that impact students' motivational beliefs, no prior study has investigated the effect of students' perception of the inclusiveness of the learning environment in a course on their course outcomes (including both motivational beliefs and academic performance at the end of the course). In this study, we focused on female and male students' physics motivational beliefs and academic performance in a calculus-based introductory physics sequence and investigated the role played by students' perception of the inclusiveness of the learning environment (including sense of belonging, peer interaction and perceived recognition) in predicting their motivational beliefs (including self-efficacy, interest and overall physics identity) and academic performance (measured by grades) at the end of this course sequence.

We found that there were statistically significant gender differences disadvantaging women in all motivational beliefs studied, which is consistent with prior studies [38,135,136]. Moreover, we found that, in the current learning environment, female students felt a lower sense of belonging and perceived recognition by their instructors/TAs than male students, and they also reported less benefit from peer interaction than men. However, we did not find gender difference in students' SAT math scores and female students actually had a somewhat higher average high school GPA than male students. Although female students had lower average grades in both physics 1 and physics 2, the gender differences in students' motivational beliefs and perception of the inclusiveness of the learning environment are much more pronounced than the gender differences in their course grades. This result is consistent with a prior study finding that in introductory physics courses, female students had significantly lower physics self-efficacy than their equally performing male peers [38]. Therefore, our study suggests that the largest gender differences were in students' motivational beliefs, rather than in actual physics performance. There could be several possible reasons for this phenomenon. For example, students' motivational beliefs can be impacted by how society sees the connection between gender and STEM achievement [25,26,29,262]. One stereotype about physics is that only students who are very smart or have natural ability can do well in physics. This may make people think of physics as a masculine subject because, due to societal stereotypes, the words "brilliant" or "genius" are usually associated with men [156]. These negative stereotypes may deteriorate female students' physics self-efficacy and identity, and they may think they need to put more effort than men to succeed in physics [25]. In addition, some prior studies have shown that instructors may teach and interact with women and men in different ways, which may lead to gender disparity in course outcomes [85]. For example, if instructors do not call on female students to answer questions or do not express the same expectation and recognition as they do for male students, female students' motivational beliefs such as perceived recognition and sense of belonging may be negatively impacted.

Another major contribution of this study is that we found that students' perception of the inclusiveness of the learning environment in an introductory physics course significantly predicts their physics motivational beliefs and grades even after controlling for their gender, motivational beliefs and grades in a previous course as well as their high school preparation. Moreover, we found that although there were large gender differences in students' self-efficacy, interest, overall physics identity and grades at the end of the course sequence, gender does not directly predict these outcome constructs, which means that the gender differences in these constructs were partially mediated by the perception of the inclusiveness of the learning environment. This finding indicates that we may be able to reduce the gender differences in students' course outcomes by developing an inclusive learning environment. However, our results show that female students perceived a lower level of inclusiveness of the learning environment than male students and the current learning environment is not helping to reduce the gender differences in students' course outcomes. Thus, the instructor's focus on equity and inclusion, and approaches to appropriately recognizing students in poorly gender-balanced classrooms become even more vital in supporting women and promoting learning for all students in the classroom [81].

Our findings suggest that an inclusive learning environment is very important for equitable outcomes in physics. As educators and instructors, we should make intentional efforts to develop an inclusive learning environment that emphasizes recognizing students for making progress, promoting positive peer interactions, and providing a space where all students can feel that they belong in which students from all demographic groups can equally excel. In addition, instructors should strive to reduce or eliminate the effects of gender, prior preparation, and prior motivational beliefs on student perception of the inclusiveness of the learning environment so that students can equally benefit from the learning environment regardless of their gender and motivational beliefs coming into the course.

In this study, we focused on female and male students' motivational beliefs and academic performance in a two-term college calculus-based introductory physics sequence. One possible limitation of this study is that we only included students who took physics 1 and physics 2 in the same school year, which means that they passed physics 1 with one attempt and students who failed physics 1 were not included in this study. Our data show that the fail rate of physics 1 was 13% in the years studied. Future studies will investigate students' perception of the inclusiveness of the learning environment in each course and how it predicts students' motivational beliefs and academic performance at end of the course, which can be compared with the results of the current study. Another limitation of the current study is that it only focuses on the underrepresentation of female students and not on other underrepresented demographic groups. In future studies, we intend to carry out similar investigations accounting for intersectional perspectives, e.g., with female and male students from different ethnic/racial groups and how their perceptions of the inclusiveness of learning environment predict their course outcomes. It would be also valuable to investigate the inclusiveness of the learning environment in other courses, such as algebra-based physics courses, where women are the majority group, often making up 60% or more of the classroom. Even though the general societal stereotype threat still exists even for female students in algebra-based physics courses, if their motivational beliefs can be protected by a classroom where they have many female peers, this may itself be a useful finding. In addition, the data from this study was collected from one research university in the US. Similar studies in different types of institutions and in other countries would also be helpful for developing a deeper understanding of the relationships between students' perception of the inclusiveness of the learning environment and their course outcomes.

## 9.5 Conclusions

Prior studies have shown that students' motivational beliefs are important factors that can influence students' engagement, performance, and retention in STEM disciplines [34,35,278-281]. In this study, we found that female students' motivational beliefs and grades in the college calculus-based introductory physics courses were statistically significantly lower than male students. Our study shows that students' perception of the inclusiveness of the learning environment plays an important role in predicting their grades and motivational beliefs and explaining the gender differences in these course outcomes. We found possible signatures of non-inclusive learning environment in that female students perceived significantly lower level of inclusiveness. Thus, the instructor's focus on equity and inclusion, and approaches to recognizing students is vital in supporting women and promoting learning for all students in the classroom. Our study can be valuable for formulating guidelines for creating an inclusive learning environment in which all students can excel.

# 10.0 How Students' Perception of The Inclusiveness of Learning Environment Mediates Gender Differences in Their Physics Motivational Beliefs and Force Concept Inventory Scores Controlling for High School Performance

## 10.1 Introduction

Prior studies have shown that women are often underrepresented in many science, technology, engineering, and mathematics (STEM) courses and disciplines [1-7,20-23,36-38,63,72,181,185]. For example, even though women earn approximately 60% of all bachelor's degrees in the US, only 20% of the physics undergraduate degrees are earned by women [13]. In addition, several studies have also reported gender disparity in students' performance in some STEM disciplines [14,15,188]. Prior research suggests that individuals' performance in STEM can be influenced by their motivational beliefs such as self-efficacy, interest and identity in that domain [18-23,34,37,41,43-45,48,63,73,255-257,259]. Students from underrepresented groups in STEM such as women may not have enough encouragement and role models to help them develop strong motivational beliefs in STEM. In addition, the societal stereotypes and biases in STEM may further undermine their motivational beliefs and could lead to withdrawal from STEM courses, majors or careers [24-30,32,74,193,262]. Therefore, investigation of how student perception of the inclusiveness of the learning environment impacts students' motivational beliefs and performance is important to understanding the underrepresentation, e.g., of women and other marginalized students in STEM, and can help in developing guidelines for developing an inclusive and equitable learning environment and promoting diversity and equity in STEM fields.

By equity in learning, we mean that not only should all students have equitable opportunities and access to resources, but they should also have an equitable and inclusive learning environment with appropriate support and mentoring so that they can engage in learning in a meaningful and enjoyable manner and the learning outcomes should be equitable. By equitable learning outcomes, we mean that students from all demographic groups (e.g., regardless of their gender identity or race/ethnicity) who have the pre-requisites to enroll in the course have comparable learning outcomes, which is consistent with Rodriguez et al.'s equity of parity model [33]. The STEM learning outcomes include student performance in courses as well as their STEM motivational beliefs because regardless of the performance, the motivational beliefs can influence students' short and long-term retention in STEM disciplines. We note that equitable opportunity, equitable and inclusive learning environment and equitable outcomes are strongly entangled with each other. For example, if the learning environment is not equitable and inclusive, the learning outcomes are unlikely to be equitable based upon the equity of parity model [33].

### 10.1.1 Students' Motivational Beliefs in Physics and Other STEM Fields

Prior research suggests that students' self-efficacy is an important motivational belief for them to excel in a domain [20,36-38]. Self-efficacy is one's belief in one's ability to succeed in a specific area or accomplish a task [39,40]. Studies have shown than students' engagement and performance can be influenced by their self-efficacy [41-44]. For example, students who have high self-efficacy tend to see difficulties as challenges and believe that productive struggles can help them improve, so they often choose to take harder courses and ask to do more challenging problems than students with low self-efficacy, who usually see difficulties as threats and obstacles to success [45].

Another motivational characteristic is interest, which refers to students' curiosity, enjoyment and engagement in a specific area [46,47]. Studies have shown that interest can also influence students' learning [41,47-51]. For example, one study showed that students' performance can be improved by connecting physics courses to students' daily lives or using evidence-based curricula to make the courses more interesting [52].

According to Eccles's Expectancy-Value Theory (EVT) [53,54], interest and self-efficacy are correlated with each other and together influence students' learning outcomes and career choices. In the EVT, expectancy refers to students' belief in their ability to succeed in a given task [54], which is closely related to self-efficacy. Value refers to the subjective task value for students, which can be differentiated into four components: intrinsic value, attainment value, utility value, and cost [54]. Intrinsic value refers to students' interest in the task and the enjoyment they experience from doing the task. Attainment value reflects how important students themselves feel it is for them to develop mastery and do a good job in the field [54]. Utility value pertains to students' perception of whether the task can help them achieve some other goals, e.g., help in career [54]. The last value component is cost, which refers to the assessments of how much effort and time will be taken to engage in the task as well as the amount of opportunity cost and stress caused by the task [54]. In the EVT, students' learning goals, academic engagement and performance, and persistence in a field are impacted by their expectancy of success and the four components of value [54].

In addition, students' identity in a specific field such as physics is another important motivational belief that influences their career decisions and outcome expectations [21,22,55-58,63,101]. Students' physics identity is related to whether they see themselves as a physics person or someone who can excel in physics [21,22,55,58,63,101]. Some studies have found that female students often report lower physics identity than male students [63-65]. This gender difference in physics identity has been shown to be related to societal biases and stereotypes about who belongs in physics and can excel in it [66-68]. In physics and other STEM fields, these stereotypes can negatively influence women's experiences, which may lower their sense of belonging and identity and thus lead to withdrawal from STEM fields [24,69,70]. Thus, studying students' STEM identity may help us understand the gender difference in participation in STEM. In Carlone and Johnson's science identity framework [22], students' science identity includes three components: competence, performance, and recognition. Hazari et al. adapted this model to physics and added interest to this model [21]. They studied whether the gender difference in students' physics identity was mediated by interest, competency belief, and perceived recognition from other people [78,79]. These two studies reveal that individuals' identity in science is not only impacted by their own motivational characteristics but also by their perceived recognition from others.

Several studies have shown that female students did not feel that they were recognized in STEM disciplines even before they entered college [66,83,183]. One stereotypical view of science is that it is for students who are very smart or have a natural gift in science [66]. In general, due to societal stereotypes, being brilliant or exceptionally smart is usually associated with boys [156]. One investigation showed that the gendered notions of brilliance are endorsed by children as young as 6 years old and have an immediate effect on their interest and identity in science [83]. These stereotypes and biases also exist in the university context [81,85]. For example, one study showed that science faculty participants rated men as more competent and were willing to offer higher starting salary and more mentoring to a male applicant than an identical female applicant even though only the names were different in the hypothetical information they were provided [85]. For female students, the experiences of not being recognized as a science person and the gender-based biases may accumulate over time and become a detriment to their science identity.

Moreover, students' interest and self-efficacy are also connected to their interaction with other people and recognition by them [40,47]. In the four-phase model of interest development, Hidi and Renninger pointed out that external factors such as group work and tutoring can trigger and maintain people's interest [47,49,86]. In addition, according to Bandura's social cognitive theory, an individual's self-efficacy can be shaped by verbal encouragement from others [151,153]. Kalender et al.'s physics identity framework [65,81] showed that students' perceived recognition not only strongly predicts their physics identity, but also predicts their physics interest and self-efficacy.

In addition to perceived recognition, students' sense of belonging and their interaction with peers have also been shown to be important aspects of the inclusiveness of the learning environment [2,69,70,195,263-265]. For example, if students have a high sense of belonging in class, they may interact with others more and with more positive attitudes, and they may also develop a higher perceived recognition [178]. However, there were very few quantitative studies about the effect of the inclusiveness of the learning environment on students' motivational beliefs and academic performance and what role is played by each component of the learning environment. Thus, further study is needed to develop a better understanding of how to develop an inclusive and equitable learning environment.

### 10.1.2 Factors That May Contribute to the Gender Difference on Concept Inventories in Physics

Since this investigation focuses on physics, we note that students' physics conceptual understanding is an important learning outcome. However, prior studies showed that female students often have lower average scores than male students on physics concept inventories [211-215]. Several factors have been proposed to explain the gender difference in students' performance on these physics concept inventories. For example, studies have shown that differences in background and preparation of men and women due to a variety of reasons can account for part of the gender gap [217,218]. In addition, some studies suggested that more interactive teaching methods may help reduce the gender gap in students' conceptual understanding [2,217,221]; however, this effect has not been consistently reproduced in other studies [14,181,222]. In particular, a study shows that in an inequitable and non-inclusive learning environment, female students may benefit less from interactive learning because they do not feel safe to express themselves, and thus the gender gap may be even larger at the end of the course than in a traditional lecture-based course [138]. Prior studies show that factors such as students' motivational beliefs can also contribute to the gender difference in students' performance in physics [180,223]. However, there are very few studies focusing on the effect of students' perception of the inclusiveness of the learning environment on their physics conceptual understanding. Therefore, in this study, we focus on how perception of the inclusiveness of learning environment predicts both students' motivational beliefs and their conceptual understanding in a college level introductory calculus-based physics course. In this study, students' conceptual understanding was measured by the Force Concept Inventory (FCI), which is one of the most commonly used concept inventories in physics for introductory mechanics [224].

### 10.1.3 The Present Study and Theoretical Model

Inspired by the above studies, we conducted a study focusing on students' physics motivational beliefs and conceptual understanding in a calculus-based introductory physics course at a large public university. This course includes topics such as kinematics, forces, energy and work, rotational motion, gravitation, and oscillations and waves. We investigated how students' perception of the inclusiveness of the learning environment (including students' sense of belonging, perceived effectiveness of peer interaction, and perceived recognition) predicts their self-efficacy, interest, identity and FCI scores at the end of the course after controlling for students' self-efficacy, interest, and FCI scores at the beginning of the course as well as their high school GPA and SAT math scores. For convenience, perceived effectiveness of peer interaction is shortened to peer interaction in the rest of the paper. We note that the learning environment here is not only the classroom environment but also includes students' experiences outside the class. For example, students may work together on their homework after class, and they could also ask for help during TAs'/instructors' office hours or communicate with the instructor/TA via email about various issues pertaining to the course. As shown in Figure 22, the thirteen constructs are divided into three groups: what we control for, students' perception of the inclusiveness of the learning environment, and outcomes. Students' gender, SAT math, high school GPA (HS GPA), and their self-efficacy, interest and FCI scores at the beginning of the course (Pre SE, Pre Interest, and Pre FCI) are constructs that we control for. Outcomes include students' self-efficacy, interest, FCI scores and identity at the end of the course (Post SE, Post Interest, and Post FCI). Perceived recognition (Perceived Recog), peer interaction (Peer Int) and sense of belonging (Belonging) constitute the perception of the inclusiveness of learning environment.

In our study, students' peer interaction, perceived recognition, sense of belonging and identity were measured at the end of the course because only after the course can students answer these survey questions based on their real experience in the course such as their interaction with peers, TAs and instructors. It is expected that students' responses to the motivational survey in pre- and post-survey are correlated because they are students' responses to the same questions pertaining to the same motivational construct at two different time points. However, if students' motivational beliefs changed from pre to post, we want to study whether the inclusiveness of the learning environment helps to explain the changes and what role is played by each construct in the inclusiveness of learning environment.

In this study, we first investigated how students' self-efficacy, interest, and FCI scores changed from the beginning to the end of the course and whether there were gender differences in the constructs studied. Then, we used Structural Equation Modeling (SEM) to study the effect of inclusiveness of learning environment on students' self-efficacy, interest, identity and FCI scores at the end of the course after controlling for gender, high school GPA, SAT math scores, and their motivational beliefs and FCI scores at the beginning of the course. To better understand the role played by each inclusiveness of learning environment construct, we first considered a model with perceived recognition as the only inclusiveness of learning environment construct to analyze how much variance in the outcome constructs is explained by the model. Then, we added peer interaction and sense of belonging into this model one by one to investigate whether adding these constructs helps to explain extra variance in the outcome constructs.

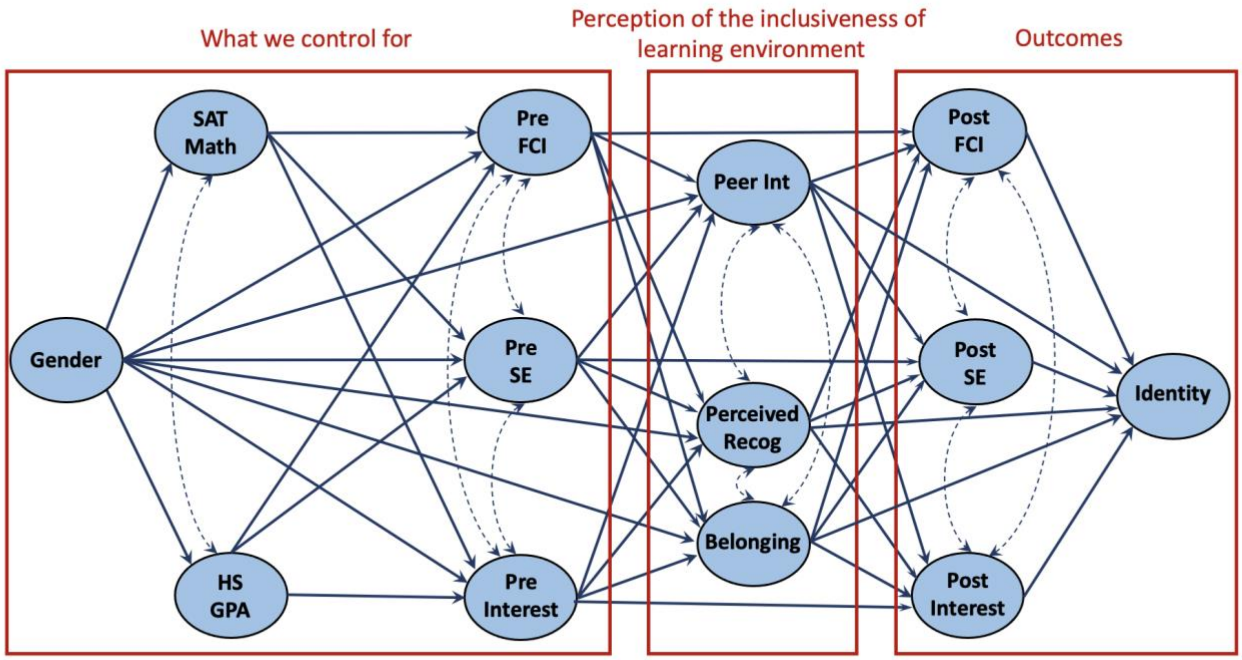



**Figure 22 Schematic representation of the theoretical model in which the relation between gender and physics identity is mediated through SAT Math scores, high school GPA (HS GPA), and FCI scores as well as peer interaction (Peer Int), perceived recognition (Recog), sense of belonging, self-efficacy (SE), and interest. The solid lines represent regression paths, and the dashed lines represent covariances. From left to right, all possible regression paths were considered, but only some of the paths are shown here for clarity.**

## 10.2 Research Questions

In this study, we used quantitative methods to investigate how students' perception of the inclusiveness of the learning environment predicts physics motivational beliefs and performance on the Force Concept Inventory (FCI) in a calculus-based introductory physics sequence at a large state-related university in the US. This course is mandatory for students majoring in engineering, physical science, and mathematics in their first year at the university. Specifically, we address the following research questions:

**RQ1.** Are there gender differences in students' FCI scores and motivational beliefs and do they change from pre to post?

**RQ2.** How do the components of students' perception of the inclusiveness of learning environment (including sense of belong, peer interaction and perceived recognition) predict students' self-efficacy, interest, identity, and FCI scores at the end of the course after controlling for students' gender, high school GPA, SAT math scores, and their self-efficacy, interest and FCI scores at the beginning of the course?

**RQ3.** Does gender moderate the relationship between any pairs of constructs in the models (i.e., does the strength of relationship given by the standardized regression coefficients between any two constructs in the models differ for women and men)?

**RQ4.** If gender does not moderate any path in the model, how does gender mediate

a. the factors that were controlled for?

b. the inclusiveness of learning environment constructs after controlling for students' high school GPA, SAT math scores, and their self-efficacy, interest and FCI scores at the beginning of the course?

c. the learning outcomes after controlling for everything in the model?

**RQ5.** What unique role is played by each of the three components we have included in the inclusiveness of learning environment in predicting the outcome constructs?

**RQ6.** Based on the aspects of students' perception of the inclusiveness of the learning environment that explain most of the variance in the outcome constructs, which model is most productive for providing guidelines for creating an inclusive environment?

## 10.3 Methodology

### 10.3.1 Participants and Data Sources

In this study, we collected the motivational survey data at the beginning and end of a college calculus-based introductory physics course in two consecutive school years at a large research university in the US. This course is generally mandatory and taken by engineering, physical science and mathematics majors in the first semester of their first year of undergraduate studies. This course is a traditional lecture-based course (4 hours per week) with recitations (1 hour per week), in which students typically work on physics problems with the help of a teaching assistant (TA). This course mainly includes mechanics topics such as kinematics, forces, energy and work, rotational motion, gravitation, and oscillations and waves.

Students' motivational beliefs were measured using a validated motivational survey. The Force Concept Inventory (FCI) [224] was used to measure students' conceptual understanding of introductory mechanics. Both the motivational survey and conceptual test were administered to students in the first and last recitation class of the semester. The demographic data of students—such as gender—were provided by the university. Students' SAT math scores and high school GPA were also obtained from the university records. Students' names and IDs were de-identified by an honest broker who provided each student with a unique new ID. Thus, researchers could analyze students' data without having access to students' identifying information. In this study, we analyzed data from 1045 students (382 female students and 663 male students) who completed the motivational survey and FCI test at both the beginning and end of the course (matched students from pre to post). We recognize that gender identity is not a binary construct. However, students' gender information was collected by the university, which offered binary options. For our analysis, we use the binary gender data. Fewer than 1% of the participants did not provide this information and therefore were not included in this analysis.

### 10.3.2 Survey Instruments

In this study, our analysis includes six motivational constructs—physics self-efficacy, interest, peer interaction, perceived recognition, sense of belonging, and identity. The questions for each construct are listed in Table 36. The survey questions were adapted from existing motivational research [92-94,162-164,267] and were re-validated in our prior work [37,95,96,127,165]. The validation and refinement of the survey involved use of one-on-one student interviews with both introductory and advanced students [37,76,96,268], exploratory and confirmatory factor analysis (EFA and CFA) [97], Pearson correlation between different constructs and Cronbach's alpha [99,100]. As shown in Table 36, all of the CFA item loadings are above 0.6 and most of them are above 0.7, which means that our construct extracts sufficient variance from the items [239].

Physics self-efficacy represents students' belief about whether they can perform well in physics. In our survey, we had four items for self-efficacy (Cronbach's alpha = 0.69 for pre-self-efficacy and Cronbach's alpha = 0.80 for post-self-efficacy [99]). These items had the response scale "NO!, no, yes, YES!", which is a 4-point Likert scale (1-4). We also had four items for physics interest (Cronbach's alpha = 0.75 for pre-interest, Cronbach's alpha = 0.82 for post-interest). 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. By choosing the four options, students will get a score from 1 to 4 accordingly. For example, if a student finds physics very boring, he/she will get one point for this item. The more interest a student has in physics, the higher score the student will get for this item. There is one item for physics identity in this survey (I see myself a physics person). Physics identity corresponds to students' belief about whether they designate themselves as a physics person [21]. This item involved a four-point Likert response on the scale: "Strongly disagree, Disagree, Agree, and Strongly agree" and they correspond to 1 to 4 points, respectively [102].

In addition, perceived recognition, peer interaction and sense of belonging are the other three motivational constructs in our study. Unlike self-efficacy, interest and identity, these three constructs are directly related to students' experience in the course. Perceived recognition included three items which represent whether a student thinks other people see them as a physics person [21,63,101] (Cronbach alpha's = 0.86). Peer interaction, including four items, represents whether students have a productive and enjoyable experience when working with peers (Cronbach alpha = 0.91). Sense of belonging is about students' feelings of whether they belonged in the physics class [195], and it included five items that were scored on a 5-point Likert scale: "not at all true, a little true, somewhat true, mostly true and completely true" (Cronbach alpha = 0.86). Two sense of belonging items ("I feel like an outsider in this class" and "Sometimes I worry that I do not belong in this physics class") were reverse coded, which means that a higher score in these two items represents a lower sense of belonging. A student's score for each construct is the average score of all items in this construct.

**Table 36 Survey items for each of the motivational constructs. The Cronbach's alphas and CFA item loadings (Lambda and *p*-values of the significance test for each item loading) shown here were calculated with postdata. †The response options for this question are "Never, Once a month, Once a week, Every day". ‡The response options for this question are "Very boring, boring, interesting, Very interesting".**

| Construct and Item | Lambda | *p* value |
|---|---|---|
| Physics Identity | | |
| I see myself as a physics person. | 1.000 | <0.001 |
| Physics Self-Efficacy (Cronbach's alpha = 0.80) | | |
| I am able to help my classmates with physics in the laboratory or in recitation. | 0.714 | <0.001 |
| I understand concepts I have studied in physics. | 0.721 | <0.001 |
| If I study, I will do well on a physics test. | 0.725 | <0.001 |
| If I encounter a setback in a physics exam, I can overcome it. | 0.663 | <0.001 |
| Physics Interest (Cronbach's alpha = 0.82) | | |
| I wonder about how physics works † | 0.647 | <0.001 |
| In general, I find physics ‡ | 0.796 | <0.001 |
| I want to know everything I can about physics. | 0.801 | <0.001 |
| I am curious about recent physics discoveries. | 0.698 | <0.001 |
| Physics Perceived Recognition (Cronbach's alpha = 0.86) | | |
| My family sees me as a physics person. | 0.900 | <0.001 |
| My friends see me as a physics person. | 0.892 | <0.001 |
| My physics TA and/or instructor see me as a physics person. | 0.677 | <0.001 |
| Physics Sense of Belonging (Cronbach's alpha = 0.86) | | |
| I feel like I belong in this class. | 0.822 | <0.001 |
| I feel like an outsider in this class. | 0.699 | <0.001 |
| I feel comfortable in this class. | 0.825 | <0.001 |
| I feel like I can be myself in this class. | 0.608 | <0.001 |
| Sometimes I worry that I do not belong in this physics class. | 0.730 | <0.001 |
| Physics Peer Interaction (Cronbach's alpha = 0.91)<br>My experience and interaction with other students in this class… | | |
| made me feel more relaxed about learning physics. | 0.700 | <0.001 |
| increased my confidence in my ability to do physics. | 0.916 | <0.001 |
| increased my confidence that I can succeed in physics. | 0.931 | <0.001 |
| increased my confidence in my ability to handle difficult physics problems. | 0.841 | <0.001 |

### 10.3.3 Quantitative Analysis of Survey Data

In this study, we first used a *t*-test [103,240] to compare students' motivational beliefs and FCI scores at the beginning and end of the course and also conducted an analysis of gender differences using descriptive statistics. Then, we used Structural Equation Modeling (SEM) [87] to study how students' perception of the inclusiveness of learning environment predicted their motivational beliefs and FCI scores at the end of the course after controlling for students' gender, high school GPA and SAT math as well as their self-efficacy, interest and FCI scores at the beginning of the course. The SEM includes two parts: confirmatory factor analysis (CFA) and path analysis.

To validate the items on our survey, we performed the CFA for each construct. The model fit is good if the fit parameters are above certain thresholds. 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 SRMR $< 0.06$ are considered as a good fit [98]. In our study, CFI = 0.940, TLI = 0.931, RMSEA = 0.051 and SRMR = 0.041, which represents a good fit. This result provides quantitative support for us to organize the motivational constructs as proposed.

Before performing the path analysis, we calculated the Pearson correlation coefficients pairwise between each pair of constructs [100]. As shown in Table 37, there are relatively strong correlations among students' motivational beliefs, while the correlation between motivational beliefs and SAT math or high school GPA are relatively small. We note that in Table 37, there are three very strong correlations. The correlation coefficient between pre-interest and post-interest is 0.87, which means that students' interest at the beginning and end of the course are highly correlated. The correlation coefficient between physics identity and perceived recognition is 0.83, which is consistent with Godwin et al. and Kalender et al.'s prior work [58,81] finding that perceived recognition is the largest predictor of physics identity. Another large correlation coefficient is between students' post-self-efficacy and sense of belonging, which is 0.82. According to prior work done by Kalender et al., these two constructs are indeed strongly correlated with each other even though they are separate constructs [183].

**Table 37 Pearson correlation coefficients of the constructs in the mediation model. *p* values are indicated by $^{**}$ for $0.001 \le p < 0.01$, $^{*}$ for $0.01 \le p < 0.05$, and $^{ns}$ for $p > 0.05$. All the other correlation coefficients have $p < 0.001$.**

| Observed Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. SAT math | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 2. HS GPA | 0.25 | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 3. Physics identity | $0.10^{**}$ | $-0.07^{*}$ | -- | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 4. Pre-FCI | 0.38 | 0.12 | 0.40 | -- | -- | -- | -- | -- | -- | -- | -- | -- |
| 5. Pre-self-efficacy | 0.18 | $0.08^{*}$ | 0.54 | 0.37 | -- | -- | -- | -- | -- | -- | -- | -- |
| 6. Pre-interest | $-0.01^{ns}$ | $-0.10^{**}$ | 0.60 | 0.31 | 0.65 | -- | -- | -- | -- | -- | -- | -- |
| 7. Post-FCI | 0.33 | 0.12 | 0.35 | 0.77 | 0.34 | 0.26 | -- | -- | -- | -- | -- | -- |
| 8. Post-self-efficacy | 0.20 | $0.03^{ns}$ | 0.67 | 0.48 | 0.60 | 0.43 | 0.44 | -- | -- | -- | -- | -- |
| 9. Post-interest | $-0.01^{ns}$ | -0.15 | 0.70 | 0.33 | 0.48 | 0.87 | 0.31 | 0.61 | -- | -- | -- | -- |
| 10. Perceived Recognition | 0.14 | $-0.02^{ns}$ | 0.83 | 0.42 | 0.51 | 0.59 | 0.40 | 0.68 | 0.67 | -- | -- | -- |
| 11. Peer Interaction | 0.16 | $0.05^{ns}$ | 0.50 | 0.25 | 0.36 | 0.29 | 0.22 | 0.67 | 0.44 | 0.47 | -- | -- |
| 12. Sense of Belonging | 0.23 | $0.05^{ns}$ | 0.61 | 0.40 | 0.45 | 0.37 | 0.40 | 0.82 | 0.54 | 0.61 | 0.68 | -- |

To analyze the relations among the constructs, we performed the full SEM. Apart from CFA, the path analysis in SEM gives regression coefficients $\beta$ for paths between each pair of constructs and the value of each $\beta$ is a measure of the strength of that relationship. Compared with a multiple regression model, a major 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. The level of SEM model fit can also be represented by CFI, TLI, RMSEA and SRMR. We first analyzed the saturated SEM model that includes all of the 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. Before performing gender mediation analysis, we first tested the gender moderation relations between each pair of constructs using multi-group 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 (see Appendix K for detailed multi-group SEM analysis results).

One advantage of SEM is that it shows not only the direct regression relation between two constructs but also all the indirect relations mediated through other constructs, which allowed us to calculate the total regression effect by adding the direct and indirect regression coefficients up. In this study, we first considered a model with perceived recognition as the only inclusiveness of learning environment construct to investigate how students' motivational outcomes and FCI scores at the end of the course are predicted by it. Then, we added peer interaction or sense of belonging as additional constructs in the inclusiveness of learning environment. Finally, our model included all the inclusiveness of learning environment constructs. We analyzed the variance in each outcome construct denoting students' motivational beliefs and FCI scores explained by each model to understand the unique role played by each inclusiveness of learning environment component and to determine if all three components are productive.

## 10.4 Results

### 10.4.1 Gender Differences in Students' Motivational Beliefs and FCI Scores

Table 38 shows the descriptive statistics of students' physics interest and self-efficacy as well as FCI scores at the beginning and end of the course. As shown in Table 38, female students had significantly lower average interest, self-efficacy, and FCI scores than male students, and the effect size given by Cohen's *d* [240] of gender difference in self-efficacy increased from 0.32 to 0.50 by the end of the course. In addition, Table 38 shows that both male and female students' interest and self-efficacy dropped generally from pre to post, and female students' interest and self-efficacy dropped ($d$ = -0.26 for interest and $d$ = -0.49 for self-efficacy) even more than male students' ($d$ = -0.20 for interest and $d$ = -0.27 for self-efficacy). Even though both female and male students' FCI scores increased by the end of the course, the gender difference is maintained.

Table 39 shows the descriptive statistics of students' perception of the inclusiveness of the learning environment (including peer interaction, perceived recognition, and sense of belonging) and physics identity. As shown in Table 39, female students had significantly lower average scores in all of the four constructs than male students, and the effect sizes are all in the medium range [104]. These results indicate that, in the current learning environment, female students reported less benefit from peer interaction and also felt a lower sense of belonging than male students. Moreover, female students' average scores pertaining to perceived recognition and physics identity indicate 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. In Appendix L, we report the percentages of students who selected each choice for each survey item, which show consistent results with the descriptive statistics shown in Table 38 and Table 39.

Table 40 shows the descriptive statistics of students' high school GPA and SAT math scores. As shown in Table 40, there was no statistically significant gender difference in students' SAT math scores, and female students had a higher average high school GPA than male students.

**Table 38 Descriptive statistics of pre- and post-interest, self-efficacy (SE), and FCI scores for female and male students. Cohen suggested that typically values of *d* ~ 0.2, 0.5 and 0.8 represent small, medium and large effect sizes [240]. Hake suggested that values of *g* < 0.3, 0.3 < *g* < 0.7, and *g* > 0.7 represent small, medium and large normalized gains [229]. A minus sign indicates that students' average score decreased from pre to post.**

| Gender | Pre-Interest (1-4) | Post-Interest (1-4) | Statistics | | Pre-SE (1-4) | Post-SE (1-4) | Statistics | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Mean | Cohen's *d* | *p* value | Mean | Mean | Cohen's *d* | *p* value |
| Male | 3.19 | 3.07 | -0.20 | <0.001 | 3.12 | 2.98 | -0.27 | <0.001 |
| Female | 2.89 | 2.73 | -0.26 | <0.001 | 2.96 | 2.70 | -0.49 | <0.001 |
| *p* value | <0.001 | <0.001 | | | <0.001 | <0.001 | | |
| Cohen's *d* | 0.55 | 0.55 | | | 0.32 | 0.50 | | |

| Gender | Pre-FCI | Post-FCI | Statistics | | |
|---|---|---|---|---|---|
| | Mean | Mean | Normalized gain (*g*) | Cohen's *d* | *p* value |
| Male | 62% | 73% | 0.29 | 0.50 | <0.001 |
| Female | 47% | 60% | 0.25 | 0.66 | <0.001 |
| *p* value | <0.001 | <0.001 | | | |
| Cohen's *d* | 0.73 | 0.64 | | | |

**Table 39 Descriptive statistics of peer interaction, perceived recognition, sense of belonging and physics identity for female and male students.**

| Gender | Peer Interaction (1-4) | Perceived recognition (1-4) | Sense of belonging (1-5) | Physics identity (1-4) |
|---|---|---|---|---|
| Male | 2.97 | 2.58 | 3.73 | 2.62 |
| Female | 2.70 | 2.26 | 3.36 | 2.19 |
| p value | <0.001 | <0.001 | <0.001 | <0.001 |
| Cohen's d | 0.42 | 0.44 | 0.43 | 0.52 |

**Table 40 Descriptive statistics of female and male students' high school GPA and SAT math scores. A minus sign indicates that female students have a higher average score than male students.**

| Grades (Score Range) | Mean | | p value | Cohen's d |
|---|---|---|---|---|
| | Male | Female | | |
| High School GPA (0-5) | 4.10 | 4.25 | < 0.001 | -0.39 |
| SAT Math (400-800) | 701 | 694 | 0.120 | 0.11 |

### 10.4.2 SEM Path Models

In this section, we describe results of the structural equation modeling (SEM) carried out to investigate how students' perception of the inclusiveness of the learning environment predicts their motivational beliefs and FCI scores at the end of the course. Because many studies have shown that perceived recognition, e.g., by instructors, is a strong predictor of students' motivational beliefs such as identity [81,269-271], all of the models shown in the main text of this paper include perceived recognition as one of the inclusiveness of learning environment constructs (see Appendix M for results of other SEM models). We first considered a model in which perceived recognition was the only inclusiveness of learning environment construct. Then we added peer interaction or sense of belonging to the inclusiveness of learning environment one by one to analyze how each helped to predict students' self-efficacy, interest, identity and FCI scores at the end of the course. Finally, we included all three constructs in our model and studied how these constructs mediated the outcomes together and what role was played by each of them.

#### 10.4.2.1 Model 1: Perceived Recognition

In our first model (Model 1), perceived recognition is the only inclusiveness of learning environment construct. The path analysis results of the SEM model are presented visually in Figure 23. The model fit indices suggest a good fit to the data: CFI = 0.939 (>0.90), TLI = 0.927 (>0.90), RMSEA = 0.050 (<0.08) and SRMR = 0.042 (<0.08). The solid lines represent regression paths and the numbers on the lines are regression coefficients ($\beta$ values), which represent the strength of the regression relations. For clarity, we present only the regression paths with $p$ value < 0.01. As shown in Figure 23, perceived recognition directly predicts students' FCI scores, self-efficacy, interest and physics identity at the end of the course. The direct effect of perceived recognition on post-self-efficacy ($\beta = 0.43$) is even larger than that of pre-self-efficacy ($\beta = 0.32$). In addition, we note that even though pre-self-efficacy directly predicts post-self-efficacy, there is also an indirect path from pre-self-efficacy to post-self-efficacy mediated through perceived recognition. The regression coefficient of the indirect path can be calculated by multiplying the regression coefficients from pre-self-efficacy to perceived recognition ($\beta$ = 0.20) and the regression coefficient from perceived recognition to post-self-efficacy ($\beta$ = 0.43), which gives us

$0.20 \times 0.43 = 0.09$. Similarly, the direct effect from pre-interest to post-interest is $\beta = 0.73$, and the indirect effect is $0.39 \times 0.24 = 0.09$. Consistent with Godwin et al. and Kalender et al.'s prior work [58,81], Figure 23 shows that identity is mainly predicted by self-efficacy, interest and perceived recognition, and perceived recognition is the largest predictor. In addition, perceived recognition also predicts students' post-FCI scores even after controlling for their pre-FCI scores, high school GPA and SAT math scores. We note that gender directly predicts high school GPA with a negative regression coefficient ($\beta = -0.19$), which means that female students on average had a somewhat higher high school GPA than male students. This is consistent with the results shown in Table 40.

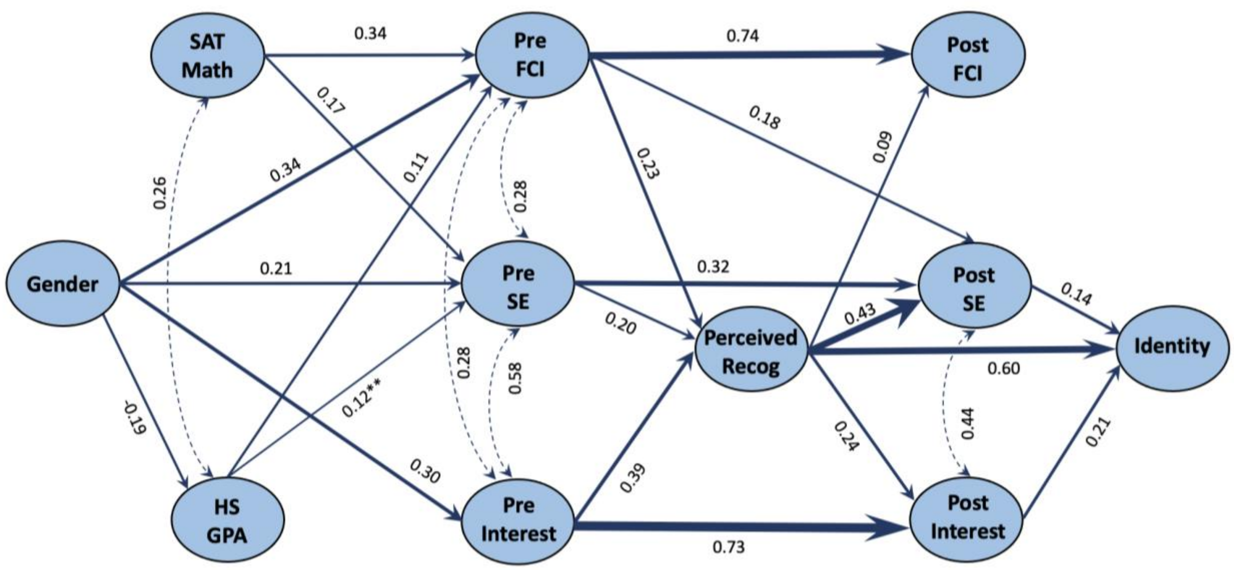



**Figure 23 Schematic diagram of the path analysis part of the structural equation modeling (Model 1) between gender and physics identity through SAT Math scores, high school GPA (HS GPA), and FCI scores as well as perceived recognition (Recog), self-efficacy (SE), and interest. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. All the other regression lines show relations with $p < 0.001$.**

#### 10.4.2.2 Model 2: Perceived Recognition and Peer Interaction

In the second model (Model 2), we include both perceived recognition and peer interaction in the perception of the inclusiveness of the learning environment. The results of the SEM model are presented visually in Figure 24. This model also fits the data very well. CFI = 0.948 (>0.90), TLI= 0.939 (>0.90), RMSEA = 0.045 (<0.08) and SRMR = 0.040 (<0.08). The results show that students' peer interaction directly predicts their post-self-efficacy ($\beta$ = 0.40) and post-interest ($\beta$ = 0.16), and it also mediates the effect from pre-self-efficacy to post-self-efficacy with indirect regression coefficient $0.32 \times 0.40 = 0.13$. We note that the direct effects of perceived recognition on post-self-efficacy and post-interest are weaker in Model 2. This is because the regression coefficient from a predictor to an outcome represents the expected change in the outcome as a result of change in the predictor in standard deviation units while controlling for the correlated effects of other predictors [244]. Since there is a shared variance between peer interaction and perceived recognition, after peer interaction was added to the model, the correlated effect of peer interaction is controlled for when estimating the regression coefficients from perceived recognition to post-self-efficacy and post-interest, so the regression coefficients decreased. On the other hand, the direct effects of perceived recognition on post-FCI scores and physics identity did not change much after adding peer interaction because peer interaction does not predict these outcome constructs directly. Similarly, the regression coefficients from perceived recognition, post-self-efficacy, and post-interest to identity are also similar to those in Model 1.

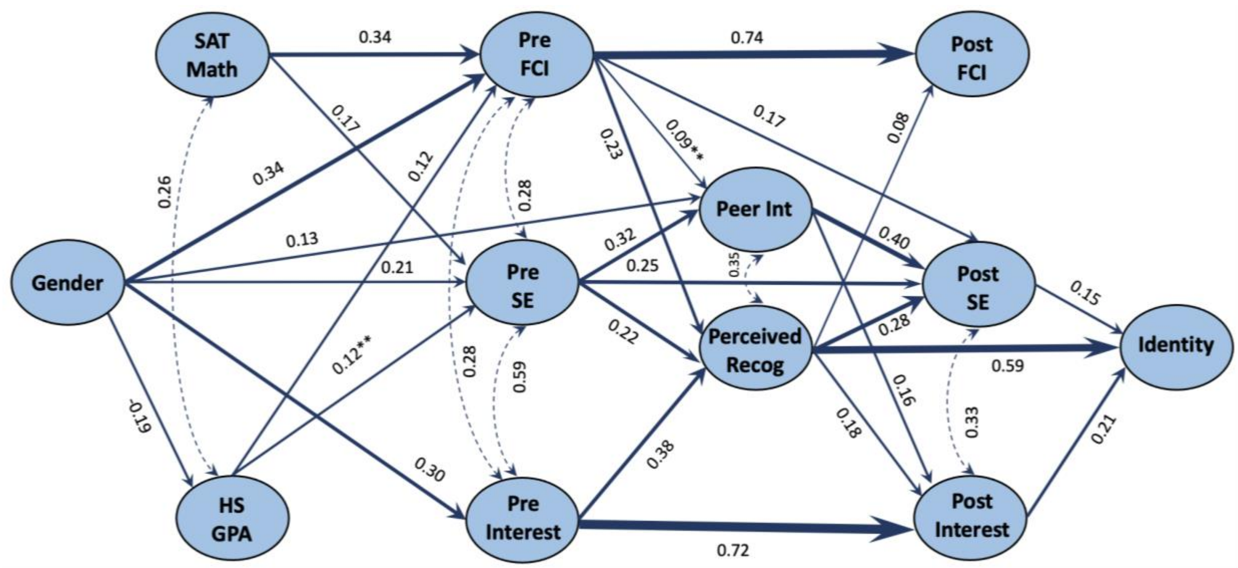



**Figure 24 Schematic diagram of the path analysis part of the structural equation modeling (Model 2) between gender and physics identity through SAT Math scores, high school GPA (HS GPA), and FCI scores as well as peer interaction (Int), perceived recognition (Recog), self-efficacy (SE), and interest. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. All the other regression lines show relations with $p < 0.001$.**

### 10.4.2.3 Model 3: Perceived Recognition and Sense of Belonging

We next analyzed a SEM model (Model 3) which includes only perceived recognition and sense of belonging as the inclusiveness of learning environment constructs. The results of the SEM model are presented visually in Figure 25. The model also fits the data well (CFI = 0.928 (>0.90), TLI = 0.918 (>0.90), RMSEA = 0.050 (<0.08) and SRMR = 0.044 (<0.08)). As shown in Figure 25, students' sense of belonging directly predicts their post-FCI scores, post-self-efficacy, and post-interest. Similarly, because there is a correlation between sense of belonging and perceived recognition, the correlated effect of sense of belonging was controlled for when estimating the regression coefficients from perceived recognition to the outcome constructs, and thus the direct effects of perceived recognition on post-FCI scores, post-self-efficacy, and post-interest became weaker or insignificant compared with those in Model 1. Since sense of belonging does not directly predict physics identity, the regression coefficients from perceived recognition, post-self-efficacy, and post-interest to identity are also similar to those in Models 1 and 2.

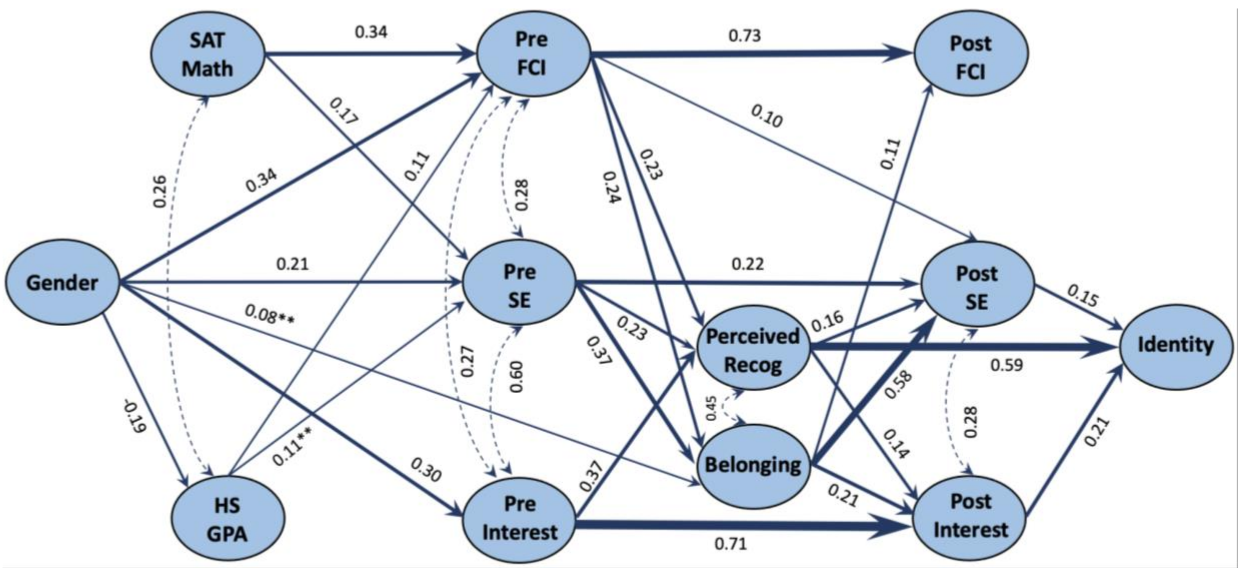



**Figure 25 Schematic diagram of the path analysis part of the structural equation modeling (Model 3) between gender and physics identity through SAT Math scores, high school GPA (HS GPA), and FCI scores as well as perceived recognition (Recog), sense of belonging, self-efficacy (SE), and interest. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. All the other regression lines show relations with $p < 0.001$.**

#### 10.4.2.4 Model 4: Perceived Recognition, Peer Interaction, and Sense of Belonging

Finally, we consider a SEM model (Model 4) which includes all three inclusiveness of learning environment constructs. Figure 26 shows the results visually. The model also fits the data very well (CFI = 0.937 (>0.90), TLI = 0.929 (>0.90), RMSEA = 0.046 (<0.08) and SRMR = 0.043 (<0.08)). As shown in Figure 26, students' perception of the inclusiveness of the learning environment significantly predicts their motivational beliefs and FCI scores at the end of the course. In particular, post-self-efficacy is directly predicted by all three inclusiveness of learning environment constructs, and sense of belonging is the largest predictor. Post-interest is predicted by perceived recognition and sense of belonging, and post-FCI is predicted by sense of belonging. Similar to Models 1-3, students' physics identity is directly predicted by perceived recognition, post-self-efficacy and post-interest, and perceived recognition is the largest predictor.

Although Table 38 and Table 39 show that there were large gender differences disadvantaging women in students' FCI scores, self-efficacy, interest, and identity at the end of the course, we note that gender does not directly predict these constructs in any of the models discussed. Thus, our results reveal that the gender differences in these outcome constructs were mediated through the different constructs of the model including components of students' perception of the inclusiveness of the learning environment.

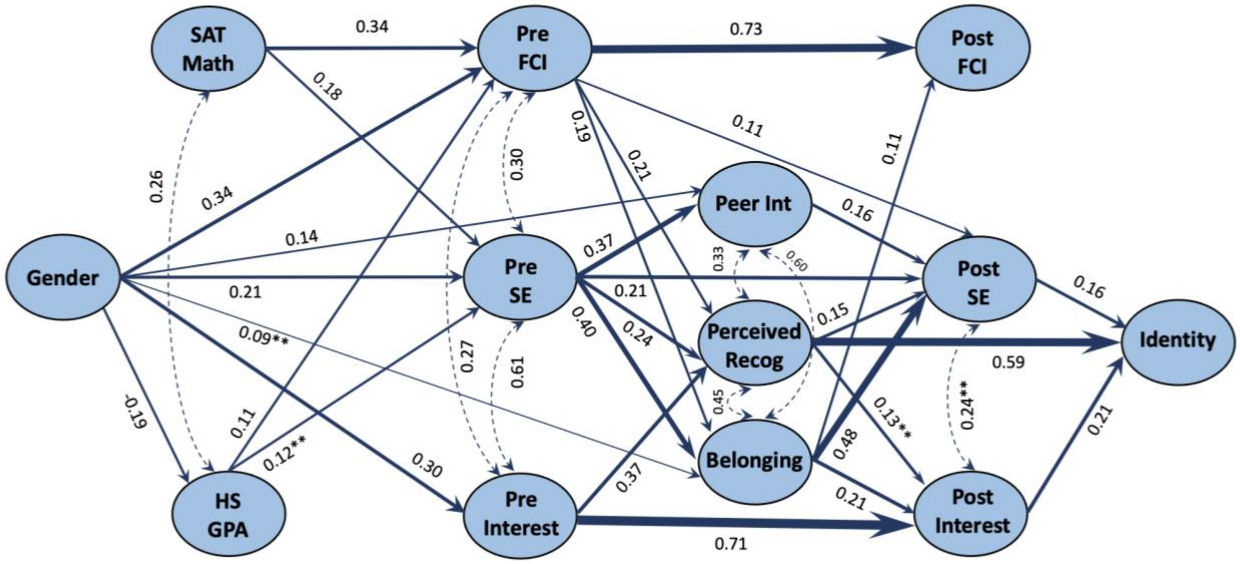



**Figure 26 Schematic diagram of the path analysis part of the structural equation modeling (Model 4) between gender and physics identity through SAT Math scores, high school GPA (HS GPA), and FCI scores as well as peer interaction (Int), perceived recognition (Recog), sense of belonging, self-efficacy (SE), and interest. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. All the other regression lines show relations with $p < 0.001$.**

### 10.4.2.5 Direct and Indirect Paths in Model 4

Model 4 shows that the three components of students' perception of the inclusiveness of the learning environment not only directly predict the outcome constructs but also mediate the indirect effect of pre-motivational beliefs and FCI scores on post-motivational beliefs and FCI scores. To summarize how the outcome constructs were predicted by different predictors through both direct and indirect paths, we calculated the regression coefficient for each path in Model 4. The results are shown in Table 41. For example, there are three different indirect paths from pre-self-efficacy to post-self-efficacy mediated through peer interaction, perceived recognition, and sense of belonging, respectively. The indirect effect of pre-self-efficacy on post-self-efficacy can be calculated by adding these three paths together ($\beta = 0.37 \times 0.16 + 0.24 \times 0.15 + 0.40 \times 0.48 = 0.29$), which is larger than the direct effect of pre-self-efficacy on post-self-efficacy ($\beta$ = 0.21). We note that the direct effect of students' sense of belonging on post-self-efficacy ($\beta$ = 0.48) is almost the same as the total effect of pre-self-efficacy on post-self-efficacy ($\beta$ = 0.50). In addition, we found that even though students' post-interest is mainly predicted by their pre-interest, it is also predicted by their sense of belonging ($\beta$ = 0.21) and perceived recognition ($\beta$ = 0.13). Similarly, even though post-FCI is mainly predicted by pre-FCI, it is also predicted by sense of belonging with $\beta$ = 0.11. Even though Figure 26 shows that perceived recognition is the only inclusiveness of learning environment construct that predicts physics identity, Table 41 shows that students' sense of belonging also indirectly predicts their physics identity with $\beta$ = 0.12.

**Table 41 Regression coefficients ($\beta$) of direct and indirect paths for the four outcome constructs predicted by various predictors in Model 4.**

| Outcome | Predictor | Direct | Indirect | Total |
|---|---|---|---|---|
| Post FCI | SAT Math | 0.00 | 0.26 | 0.26 |
| | High School GPA | 0.00 | 0.09 | 0.09 |
| | Pre-FCI | 0.73 | 0.02 | 0.75 |
| | Pre-self-efficacy | 0.00 | 0.04 | 0.04 |
| | Pre-interest | 0.00 | 0.00 | 0.00 |
| | Peer Interaction | 0.00 | 0.00 | 0.00 |
| | Perceived Recognition | 0.00 | 0.00 | 0.00 |
| | Belonging | 0.11 | 0.00 | 0.11 |
| Post Self-efficacy | SAT Math | 0.00 | 0.17 | 0.17 |
| | High School GPA | 0.00 | 0.09 | 0.09 |
| | Pre-FCI | 0.11 | 0.12 | 0.23 |
| | Pre-self-efficacy | 0.21 | 0.29 | 0.50 |
| | Pre-interest | 0.00 | 0.06 | 0.06 |
| | Peer Interaction | 0.16 | 0.00 | 0.16 |
| | Perceived Recognition | 0.15 | 0.00 | 0.15 |
| | Belonging | 0.48 | 0.00 | 0.48 |
| Post Interest | SAT Math | 0.00 | 0.04 | 0.04 |
| | High School GPA | 0.00 | 0.02 | 0.02 |
| | Pre-FCI | 0.00 | 0.07 | 0.07 |
| | Pre-self-efficacy | 0.00 | 0.12 | 0.12 |
| | Pre-interest | 0.71 | 0.05 | 0.76 |
| | Peer Interaction | 0.00 | 0.00 | 0.00 |
| | Perceived Recognition | 0.13 | 0.00 | 0.13 |
| | Belonging | 0.21 | 0.00 | 0.21 |
| Identity | SAT Math | 0.00 | 0.11 | 0.11 |
| | High School GPA | 0.00 | 0.05 | 0.05 |
| | Pre-FCI | 0.00 | 0.18 | 0.18 |
| | Pre-self-efficacy | 0.00 | 0.25 | 0.25 |
| | Pre-interest | 0.00 | 0.39 | 0.39 |
| | Peer Interaction | 0.00 | 0.03 | 0.03 |
| | Perceived Recognition | 0.59 | 0.05 | 0.64 |
| | Belonging | 0.00 | 0.12 | 0.12 |

### 10.4.2.6 Variance Explained by Each Model

To further understand the role played by each inclusiveness of learning environment construct in explaining the outcome constructs, we calculated the coefficients of determination $R^2$ (fraction of variance explained) for each construct in different SEM models with different combinations of the three inclusiveness of learning environment constructs. The results are shown in Table 42. Even though in the main text we only discussed four models that include perceived recognition, Table 42 also show the $R^2$ values for the other three models in Appendix M that do not include perceived recognition.

We found that in all models, input constructs have relatively small $R^2$ values. For example, both pre-self-efficacy and pre-interest have $R^2$ = 0.09, which means that only 9% of the variance in these constructs are explained by the models. This is because input constructs are what we are controlling for and thus they are only explained by very few predictors. On the other hand, $R^2$ values of all outcome constructs are reasonably high, which means that our models have explained much of the variance in them. For example, models that include any of the three inclusiveness of learning environment constructs can explain around 80% of the variance in post-interest and 60% of the variance in post-FCI. We note that the model only including sense of belonging can explain 75% of variance in post-self-efficacy and adding perceived recognition and peer interaction only slightly improves the $R^2$ value. Moreover, the models including sense of belonging always explain more variance in post-self-efficacy than the models without sense of belonging do. These results are consistent with the finding discussed earlier that sense of belonging is the major predictor of post-self-efficacy. Similarly, the model that only includes perceived recognition explains 74% of the variance in identity and adding peer interaction or sense of belonging does not help explain the variance in identity further. These results show that both sense of belonging and perceived recognition play unique roles in explaining students' motivational outcomes, while peer interaction covaries with sense of belonging and perceived recognition and uniquely explains very small percentages of the variance in the outcome constructs. However, this does not mean that student perception of peer interaction is not important. Figure 26 shows that peer interaction directly predicts post-self-efficacy even after controlling for the correlated effects of sense of belonging and perceived recognition. In addition, the correlations among these three inclusiveness of learning environment constructs suggest that an instructor can potentially improve students' sense of belonging and perceived recognition by helping students interact meaningfully with peers. Thus, we believe the model including all three inclusiveness of learning environment constructs is most productive.

**Table 42 Coefficient of determination ($R^2$) for various constructs in different models with different combinations of perceived recognition (Recog), peer interaction, and sense of belonging (Bel). All $R^2$ values are significant with *p* values < 0.001.**

| Construct | Models | | | | | | |
|---|---|---|---|---|---|---|---|
| | Recog | Peer | Bel | Peer+Recog | Peer+Bel | Recog+Bel | Peer+Recog+Bel |
| SAT Math | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| High school GPA | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| Pre-FCI | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| Pre-self-efficacy | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| Pre-interest | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| Peer interaction | / | 0.17 | / | 0.17 | 0.17 | / | 0.18 |
| Perceived recognition | 0.43 | / | / | 0.43 | / | 0.43 | 0.43 |
| Sense of belonging | / | / | 0.28 | / | 0.29 | 0.29 | 0.28 |
| Post-FCI | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 |
| Post-self-efficacy | 0.57 | 0.67 | 0.75 | 0.69 | 0.78 | 0.77 | 0.79 |
| Post-interest | 0.80 | 0.79 | 0.79 | 0.80 | 0.79 | 0.80 | 0.81 |
| Identity | 0.74 | 0.61 | 0.61 | 0.74 | 0.61 | 0.74 | 0.74 |

### 10.5 Summary and Discussion

In this study, we focused on students' physics motivational beliefs and FCI scores in a college calculus-based introductory physics course. We studied how students' perception of the inclusiveness of the learning environment—including peer interaction, perceived recognition, and sense of belonging—predicts students' physics self-efficacy, interest, identity and FCI scores at the end of the course after controlling for their gender, high school GPA, SAT math, and pre-self-efficacy, pre-interest and pre-FCI scores at the beginning of the course.

Our results show that the inclusiveness of the learning environment statistically significantly predicts students' motivational beliefs and FCI scores at the end of the course. Moreover, even though we found that there are statistically significant gender differences disadvantaging women in self-efficacy, interest, identity and FCI scores at the end of the course, gender does not directly predict these outcome constructs. This means that the gender differences in these learning outcomes were mediated by students' perception of the inclusiveness of the learning environment. Thus, in addition to being driven by prior differences, which often result from inequities including societal stereotypes and biases about who belongs in physics and lack of role models, students' self-efficacy, interest, identity and FCI scores are also influenced by the inclusiveness of learning environment [294]. However, our results show that the current learning environment is not helping to reduce the gender difference, and instead, the gender difference in students' self-efficacy increased by the end of the course. We note that, in the current learning environment, female students also reported less benefit from peer interaction, felt a lower sense of belonging and felt less recognized as a physics person than male students, which may all contribute to the gender differences in students' learning outcomes at the end of the course. For example, in a male-dominated classroom environment, a woman may experience a lower level of sense of belonging and higher level of anxiety with lower self-efficacy than men [255]. In addition, non-supportive instructional pedagogies, lack of recognition from instructors and TAs and lack of positive interactions with peers can further decrease women's self-efficacy in physics. Thus, the instructor's focus on equity and inclusion, and approaches to recognizing students in poorly gender-balanced classrooms, become even more vital in supporting women's self-efficacy and promoting learning for all students in the classroom [81].

Our findings also suggest that students' perception of the inclusiveness of the learning environment plays a very important role in explaining their motivational beliefs and performance at the end of the course. In particular, we found that perceived recognition uniquely contributed most to explaining physics identity, and sense of belonging uniquely contributed most to explaining self-efficacy. We note that even though peer interaction co-varies with sense of belonging and perceived recognition and peer interaction uniquely explains very small percentages of the variance in the outcome constructs, this does not mean that effective peer interaction is not important. Many instructors may not know how to implement strategies to improve students' sense of belonging. The co-variation between the inclusiveness of learning environment constructs suggests a possibility that students' sense of belonging and perceived recognition may possibly be shaped by helping students interact meaningfully with peers (which in turn can improve student outcomes). Actually, prior studies have pointed out that the learning environment is more of an interconnected ecological system than the simple sum of its parts [295]. Thus, we believe the model including all three inclusiveness of learning environment constructs is most productive.

By comparing students' responses to the survey in pre and post, we found that both male and female students' self-efficacy and interest statistically significantly dropped from pre to post. And female students' motivational beliefs dropped even more than male students', which may partially explain that the gender difference in students' self-efficacy increased by the end of the course. These results indicate that the current learning environment is not helping students improve their physics motivational beliefs and, on the contrary, may contribute to decreasing them in such a way that the gender gap increases.

Therefore, instructors must make intentional efforts to help students improve their physics motivational beliefs and performance within the equity of parity framework discussed earlier (i.e., regardless of the initial value at the beginning of the course, instructors should strive to ensure that at the end of the course, all demographic groups have similar high levels of motivational beliefs and performance). As noted, the perception of the inclusiveness of the learning environment directly predicts students' motivational outcomes and post-FCI scores, so it is reasonable to expect that a more inclusive and equitable learning environment will help. Instructors should strive to reduce the effects of prior preparation and prior motivational beliefs so that all students can equally benefit from the learning environment. If we could eliminate the gender difference in sense of belonging, perceived recognition and peer interaction by creating a learning environment, in which all students feel safe to engage in collaboration and discussions with peers and instructor, and provide appropriate scaffolding support commensurate with students' prior knowledge, the gender difference in students' motivational beliefs and FCI scores may also decrease.

Evidence-based instructional strategies may be helpful for instructors to improve the inclusiveness of the learning environment and support underrepresented students. For example, instructors can provide students with opportunities to engage in different types of interaction, such as setting up study groups or assigning collaborative tasks [296]. However, instructors need to keep in mind how the lack of role models and societal stereotypes and biases about who belongs in physics and can excel in it impact the stereotyped groups and avoid letting a small group of students dominate the discussion so that all students' voices can be heard and valued. Another stereotype about physics is that it requires a natural ability to excel [11,83]. Studies have shown that the idea of ability being fixed and unchangeable can increase students' concerns about belonging, especially for students from underrepresented groups such as women in physics who have few role models [249,250]. Thus, it is critical to build a learning environment which emphasizes that abilities are malleable and can be changed through deliberate practice and effort [146]. Instructors can also show students non-stereotypical role models from diverse demographic groups, personalities and interest in different contexts since this has been shown to increase students' sense of belonging [251,252]. In addition, instructors can explicitly recognize students by directly acknowledging their work and expressing faith in their ability, and 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 valued [145]. However, 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 [11,83].

In this study, we discussed how the inclusiveness of the learning environment impacts female and male students' motivational beliefs in the introductory calculus-based physics course. In the future studies, we intend to carry out similar investigations accounting for intersectional perspectives, e.g., with female and male students from different ethnic/racial groups and how their perceptions of the inclusiveness of learning environment predict their learning outcomes. In addition, it would be valuable to investigate the inclusiveness of the learning environment in other courses, such as algebra-based physics courses, where women are the majority group, often making up 60% or more of the class. It would also be useful to investigate students' physics motivational beliefs in advanced physics courses beyond the first year, which are typically taken by physics majors.

# 11.0 How Engineering Identity of First-year Female and Male Engineering Majors is Predicted by Their Physics Self-efficacy and Identity

## 11.1 Introduction and Theoretical Framework

Due to the increasing demand in the work force for engineers, many studies have focused on issues surrounding students' retention and persistence in engineering [297-303]. According to a recent report, the overall 4-year graduation rate of students in the US who enter an undergraduate engineering program remains below 40% in the last 10 years [304]. In particular, 20% of students are lost within the first year alone [304]. Another study shows that only 42% of seniors enrolled in undergraduate engineering programs definitely intend to pursue a career in engineering upon graduation [305]. Moreover, the retention issue is even more severe for students from underrepresented groups such as women [302,304,306]. Studies show that less than 30% of all engineering degrees are awarded to women [297], and women have been found to leave engineering at an earlier stage than men [307,308]. Many factors have been shown to affect undergraduate students' choices to persist in engineering, for example, students' prior preparation, quality of teaching, sociocultural and motivational factors [17,34,69,70,265,294,299,309-312]. In particular, motivational factors such as engineering identity have been shown to be significant indicators of students' retention in engineering, and also influence their short-term and long-term career goals [269,313-315].

In prior research, engineering identity has been studied from several different perspectives [316,317]. For example, some studies consider engineering identity as the combination of multiple identities such as academic, social and occupational identities [313,318,319]. Some other studies identified several cognitive, affective, and performance variables to comprise engineering identity [124,300,320,321]. Another widely used definition of engineering identity is how students see themselves with respect to engineering or whether they see themselves as an engineer based on their perceptions and navigation of engineering related experiences [322-324], which is also the most relevant definition to our study. However, studies have shown that many students have very few direct experiences with engineering before they enter college [325]. Thus, due to the interdisciplinary nature of engineering, students' experiences in other engineering related domains such as math and science may play a very important role in the development of students' engineering identity [124]. For example, studies have shown that doing well in math and science courses in high school has a positive impact on students' choice of and persistence in an engineering major and longer-term career goals [299,301]. Therefore, studying students' motivational beliefs in engineering related domains, e.g., physics, and how they interact with engineering identity may help us develop a better understanding of students' attrition and retention in engineering majors.

Introductory physics courses usually serve as a prerequisite for many engineering courses, and thus for most students who enrolled in an undergraduate engineering program, physics is mandatory in their first year. A study shows that students' grades in introductory physics courses predict their performance in later engineering courses [326]. Moreover, physics is not only important for engineering students' knowledge building but may also affect their attitudes and self-beliefs about being an engineer. For example, studies have shown that students' physics motivational beliefs such as self-efficacy and interest can influence their engineering career agency [58]. However, physics is also one of the most stereotyped domains in the sense that it is a traditionally male-dominated field and has a masculine culture and a masculine public image

[69,70]. In addition, physics is perceived by many people to depend largely on the innate qualities of "brilliance" or "genius", which are also typically attributed to men [11,83,156]. These societal stereotypes not only impact female students' physics motivational beliefs but may also dissuade them from pursuing study in physics-related disciplines such as engineering. A prior study shows that in an undergraduate engineering program, students' self-efficacy in physics showed a larger gender difference than their self-efficacy in mathematics, engineering, and chemistry [135]. In addition, physics was the only science subject for which female engineering students had a lower average score than male engineering students [135,327]. Therefore, the gender difference in physics motivational beliefs may partially explain the underrepresentation of women in engineering disciplines and studying the relationship between students' physics and engineering motivational beliefs may provide new insights into how to improve the recruitment, retention and diversity within engineering.

In this study, we investigated first-year undergraduate engineering students' engineering identity and physics motivational beliefs (including physics identity, self-efficacy, interest and perceived recognition) in a calculus-based introductory physics course. In particular, we focus on how students' motivational beliefs in physics and engineering change from the beginning to the end of the course and the predictive relationships among these motivational constructs. This course is usually taken by engineering students in the first semester of their first-year of undergraduate study, and they must pass this course before they declare a specific major, e.g., electrical engineering, within the engineering school. Thus, students' experiences in this course are not only important for the development of their physics and engineering motivational beliefs but may also influence their choice of majors. Even though there are several studies focusing on students' physics and engineering motivational beliefs [58,124,320], very few studies have investigated how male and female students' engineering identity and physics motivational beliefs change in an introductory physics course, and how students' physics motivational beliefs predict their engineering identity at the end of the course.

As noted, students' identity in engineering or physics is related to whether they see themselves as an engineer or a physics person, and these identities have been shown to influence students' career decisions and outcome expectations [21,22,300,322]. The other three motivational beliefs considered in this study (physics self-efficacy, interest and perceived recognition) have been shown to be the predictors of students' physics identity and also very important to students' engagement, performance and retention [21,81,136,260,261]. In particular, self-efficacy is defined as students' beliefs in their capability to succeed in a certain situation, task, or particular domain [39,81,82]. Studies suggest that 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 [41,44,45]. Another motivational belief is interest, which is defined by positive emotions accompanied by curiosity and engagement in a particular topic [47]. Interest has also been shown to influence students' learning [41,47,48]. For example, one study suggested that making science courses more relevant to students' lives and transforming curricula to promote interest in learning can improve students' achievement [52]. Perceived recognition (also called external identity) in physics refers to students' perception about whether other people see them as a physics person [79]. Some quantitative studies focusing on the relation between students' physics identity and other motivational beliefs show that physics perceived recognition is actually the strongest predictor of physics identity as compared to physics interest and self-efficacy [58,81].

In this study, we first investigated how students' engineering and physics motivational beliefs change from the beginning to the end of the introductory physics course (i.e., from pre to post) using descriptive statistics. Then, we performed structural equation modeling (SEM) using the post data to study the predictive relationships among these motivational constructs. We focus on the predictive relationships at the end of the course because most students take this course in the first semester in college, and they may feel uncertainty and anxiety during the transition from high school to college. Thus, students' motivational beliefs may be more stable after they have been on campus for a semester. In addition, since students' perceived recognition is related to their interaction with TAs and instructors, only after the course can students answer these survey questions based on their real experience in the course. We adapt the physics identity model from Hazari et al.'s (with Godwin, Lock, and Potvin) [57,58,125] and Kalender et al.'s prior work [81] as shown in Figure 27, in which students' physics identity is predicted by their interest, recognition, and performance/competence or competency belief, which is very closely tied to self-efficacy. In this study, we add the engineering identity construct and focus on how students' physics motivational beliefs predict their engineering identity. As shown in Figure 28 (a), we first considered a model (Model 1) in which there are only covariances between each pair of 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 (Figure 28 (b)), which is similar to the model in Kalender et al.'s prior work, in which competency belief and interest are predicted by perceived recognition (see Figure 27 (b)) [81].

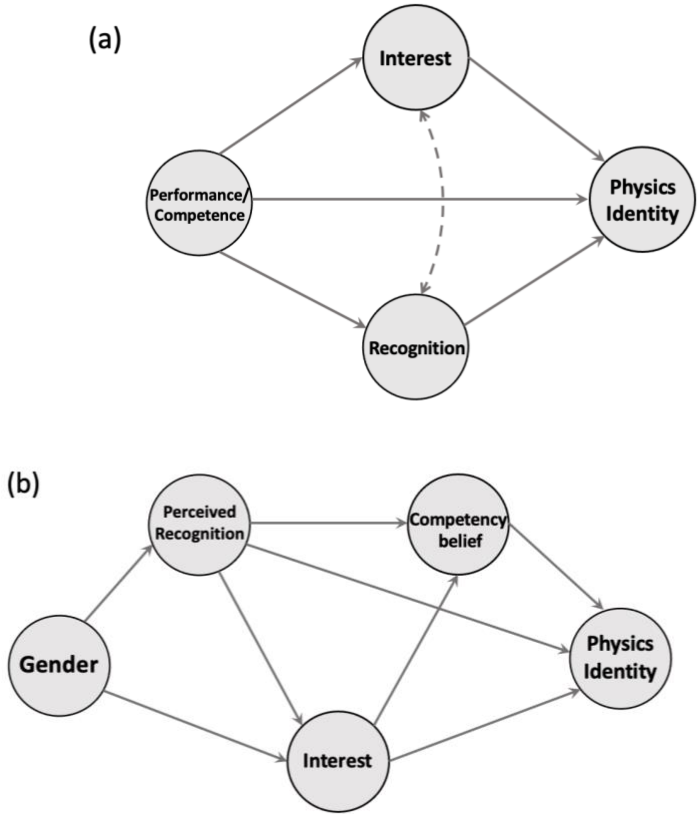



**Figure 27 Schematic representation of the physics identity models of prior studies. (a) shows the model used in Hazari et al.'s (with Godwin, Lock, and Potvin) prior studies [57,58,125] (b) shows the model used in Kalender et al.'s prior study [81]**

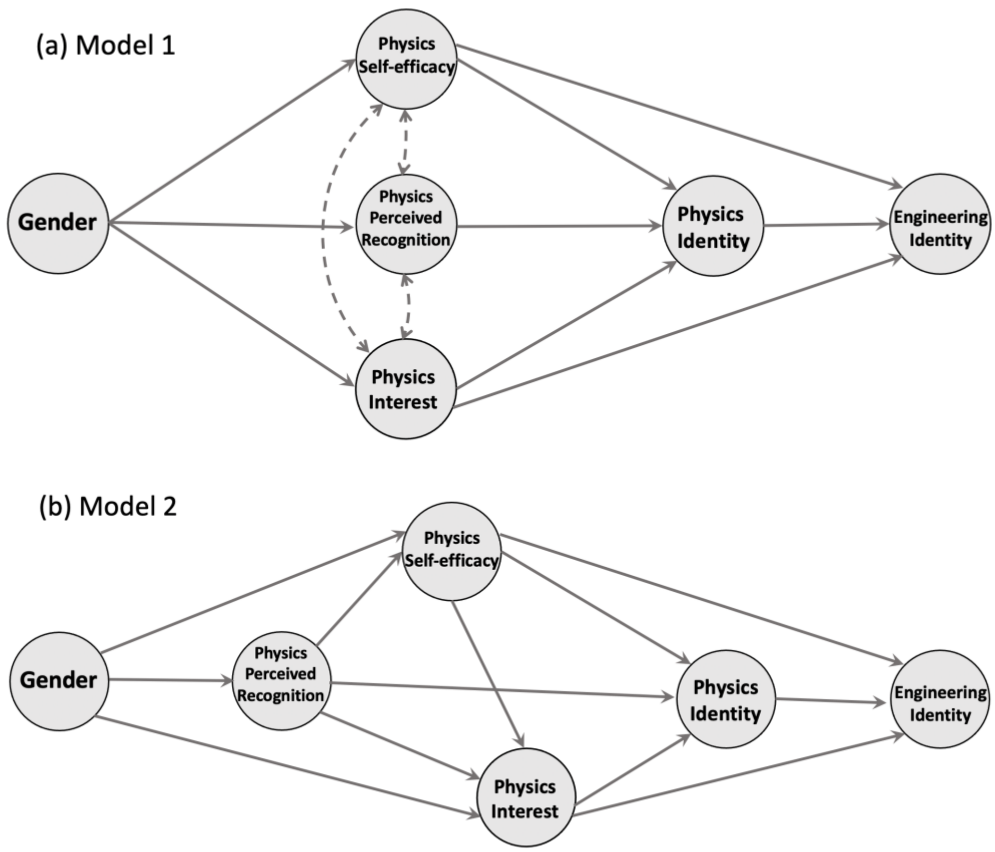



**Figure 28 Schematic representation of the path analysis part of the SEM models that shows how the relationship between gender and engineering identity is mediated by physics self-efficacy, physics perceived recognition, physics interest and physics identity. (a) In Model 1, physics self-efficacy, perceived recognition and interest are correlated with each other. (b) In Model 2, physics perceived recognition predicts physics self-efficacy and interest, and physics self-efficacy predicts physics interest. The direct paths from gender to physics and engineering identity are not shown because they are not statistically significant in both models.**

## 11.2 Research Questions

Our research questions to investigate the relationship between physics motivational beliefs and engineering identity of undergraduate engineering students in the calculus-based introductory physics 1 course at a large research university in the US are as follows:

**RQ1.** How do male and female students' engineering identity and physics motivational beliefs (including physics identity, self-efficacy, interest, and perceived recognition) change from the beginning to the end of the course (i.e., from pre to post)?

**RQ2.** Are there gender differences in students' motivational beliefs and do they change from pre to post?

**RQ3.** How do students' physics motivational beliefs directly and indirectly predict their engineering identity?

## 11.3 Methodology

### 11.3.1 Participants

The motivational survey data used in this study were collected at the beginning and end of the semester from engineering students who took the calculus-based introductory physics 1 course at a large research university in the US. The data were collected from two consecutive fall semesters. The majority of these students were in the first semester of their first year in the undergraduate engineering program. This course consists of traditional lectures (4 hours per week) and recitations (1 hour per week), in which students typically work collaboratively on physics problems. The paper surveys were handed out and collected by TAs in the first and last recitation class of a semester. We named the data collected at the beginning of the semester as pre-data and that collected at the end of the semester as post-data. Finally, we combined the two semesters' data and put them into two categories, pre and post. The demographic data of students—such as gender—were provided by the university. Students' names and IDs were de-identified by an honest broker who generated a unique new ID for each student (which connected students' survey responses with their demographic information). Thus, researchers could analyze students' data without having access to students' identifying information.

In this study, we first investigated how students' physics and engineering motivational beliefs change from pre to post. However, because some motivational constructs were added to our survey at the end of the course in the first year of study, we do not have the pre-data for these constructs in that year. Thus, we first focus on 346 undergraduate engineering students (205 male students and 141 female students) who completed both the pre- and post-survey in the second year of study. We use Structural Equation Modeling (SEM) to study the predictive relationships among the motivational constructs at the end of the course [87]. Since we have complete post-data for both years of study, we performed SEM with the post-data collected from 761 engineering students (273 female students and 488 male students) in both years, which further improved the statistical power. Because students' gender information was obtained from the university, which offered binary options, we did the analysis with the binary gender data in this study.

### 11.3.2 Survey Instruments

In this study, we considered five motivational constructs—engineering identity and physics identity as well as physics self-efficacy, interest and perceived recognition. The survey items for each construct are listed in Table 43. The survey items were adapted from the existing motivational research [92-94,162-164] and have been revalidated in our prior work [37,65,95,96,127,128,165]. 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) [97], Pearson correlation between different constructs and Cronbach's alpha (which is a measure of the internal consistency of each construct with several items) [98-100].

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 motivational beliefs. Physics self-efficacy represents students' belief about whether they can excel in physics. We had four items for physics self-efficacy and these items had the response scale "NO!, no, yes, YES!" (Cronbach's $\alpha$ = 0.81), which have been shown to have good psychometric properties and a low cognitive load while reading [82,92]. We also had four items for physics interest (Cronbach's $\alpha$ = 0.82). 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 under physics interest were answered on the "NO!, no, yes, YES!" scale. Physics perceived recognition corresponds to whether a student thinks other people see them as a physics person [21,63,101], and it includes three items which correspond to family, friends and TA/instructor (Cronbach's $\alpha$ = 0.87). Physics identity corresponds to students' belief about whether they designate themselves as a physics person [21]. Engineering identity corresponds to whether they see themselves as an engineer [322,323]. The items for physics perceived recognition and both physics and engineering identity involved a four-point Likert response on the scale "strongly disagree, disagree, agree, and strongly agree" and they correspond to 1 to 4 points [102].

**Table 43 Survey questions for each of the motivational constructs, along with factor loadings of CFA using two years of post-data. 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. †The response options for this question are "Never, Once a month, Once a week, Every day". ‡The response options for this question are "very boring, boring, interesting, very interesting".**

| Construct and Item | Lambda |
|---|---|
| Engineering identity | |
| I see myself as an engineer. | 1.000 |
| Physics identity | |
| I see myself as a physics person. | 1.000 |
| Physics self-efficacy (Cronbach's $\alpha$ = 0.81) | |
| I am able to help my classmates with physics in the laboratory or in recitation. | 0.731 |
| I understand concepts I have studied in physics. | 0.736 |
| If I study, I will do well on a physics test. | 0.742 |
| If I encounter a setback in a physics exam, I can overcome it. | 0.682 |
| Physics interest (Cronbach's $\alpha$ = 0.82) | |
| I wonder about how physics works † | 0.650 |
| In general, I find physics ‡ | 0.781 |
| I want to know everything I can about physics. | 0.791 |
| I am curious about recent physics discoveries. | 0.707 |
| Physics perceived recognition (Cronbach's $\alpha$ = 0.87) | |
| My family sees me as a physics person. | 0.913 |
| My friends see me as a physics person. | 0.909 |
| My physics TA and/or instructor sees me as a physics person. | 0.692 |

### 11.3.3 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' pre- and post-scores for each motivational construct as well as conducted an analysis of gender differences using descriptive statistics. We performed Item Response Theory (IRT) analysis using the R software package "mirt" to check the response option distances for our survey constructs [105-107,166]. The results show that our scales had approximately equal distance between the levels, so the linearity assumption is reasonable and allowed us to calculate the traditional mean scores [105,107]. Furthermore, we estimated the IRT-based scores (which tend to produce trait estimates that are linearly related to the underlying trait being measured) for each construct, and the results are highly correlated with the mean scores (the correlation coefficients are > 0.98 for all constructs), which indicates that the use of mean scores is reasonable [105].

Next, we performed Structural Equation Modeling (SEM) [87] with the post-data to study the predictive relationships between students' physics motivational beliefs and engineering identity. The SEM includes two parts: confirmatory factor analysis (CFA) and path analysis. In CFA, the model fit is good if the fit parameters are above threshold. In particular, 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 [98]. In our study, CFI = 0.976, TLI = 0.967, RMSEA = 0.054 and SRMR = 0.033, which represent a good fit. Thus, there is additional quantitative support for dividing the constructs as proposed. Besides, as shown in Table 43, all factor loadings are higher than 0.5, which is considered acceptable, 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 path analysis part of SEM [108].

Before performing the path analysis, we calculated the Pearson correlation coefficients pairwise between the motivational constructs. As shown in Table 44, all correlation coefficients are above 0.2, and most of them are less than 0.8, which means that even though these motivational constructs have strong correlations with each other, the correlations are not so high that they could not be examined as separate constructs in SEM [109]. We note that the correlation coefficient between physics identity and perceived recognition is 0.84. This is consistent with Godwin et al.'s [58] and Kalender et al.'s [81] prior finding that students' physics perceived recognition (external identity) is the largest predictor of their physics identity (internal identity).

**Table 44 Pearson correlation coefficients of the constructs in the mediation model. ($p<0.001$)**

| Constructs | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1. Engineering identity | -- | -- | -- | -- | -- |
| 2. Physics identity | 0.34 | -- | -- | -- | -- |
| 3. Physics self-efficacy | 0.37 | 0.70 | -- | -- | -- |
| 4. Physics Interest | 0.32 | 0.71 | 0.64 | -- | -- |
| 5. Physics perceived recognition | 0.30 | 0.84 | 0.70 | 0.67 | -- |

To analyze the predictive relationships among the constructs, we performed the path analysis. Apart from CFA, the path analysis in SEM gives regression coefficients $\beta$ for paths between each pair of constructs and the value of each $\beta$ is a measure of the strength of that relationship. 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. The level of SEM model fit can also be represented by CFI, TLI, RMSEA and SRMR. 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. Before performing gender mediation analysis, we first tested the gender moderation relations between each pair of constructs using multi-group 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 of 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 (see Appendix N for detailed multi-group SEM analysis results). We fit the two SEM models (Model 1 and Model 2) shown in Figure 28 with our data and then compared the path analysis results (predictive relationships among the constructs) for these two models.

## 11.4 Results

### 11.4.1 Descriptive Statistics of Students' Motivational Beliefs at the Beginning and End of the Course

Here, we present the descriptive statistics of the students' pre- and post-motivational beliefs. As shown in Table 45, female students had significantly lower scores in all of the five motivational constructs. In particular, we note that the gender difference in students' physics identity is larger than that in engineering identity and students' physics identity is lower than engineering identity, which is expected because they are all engineering students. In addition, Table 45 shows that both male and female students' physics self-efficacy and physics identity deteriorated from pre to post. Moreover, female students' average scores on physics self-efficacy and physics interest decreased more than male students' did so that the gender differences in these two constructs became larger by the end of the course. Although students' average score on engineering identity also decreased from pre to post, this change is only statistically significant for male students.

We also conducted a one-way repeated measures MANOVA to analyze the changes in multiple dependent variables over time (from pre to post). The results show that female students' overall physics and engineering motivational beliefs decreased from pre to post ($F(5,128) = 7.103$ $p < 0.001$, Wilks' Lambda = 0.78, partial eta squared ($\eta_p^2$) = 0.217). Partial eta squared values indicate effect sizes in one-way MANOVA with $\eta_p^2 \sim 0.01$ generally considered a small effect size, $\eta_p^2 \sim 0.06$ a medium effect size and $\eta_p^2 \sim 0.14$ a large effect size [328]. Follow-up univariate tests show that female students' physics identity ($F(1,132) = 12.946$, $p < 0.001$, $\eta_p^2 = 0.089$), self-efficacy ($F(1,132) = 22.590$, $p < 0.001$, $\eta_p^2 = 0.146$), and interest ($F(1,132) = 11.215$, $p = 0.001$, $\eta_p^2 = 0.078$) statistically significantly decreased from pre to post. Similarly, male students' overall physics and engineering motivational beliefs also decreased from pre to post ($F(5,189) = 4.361$ $p = 0.001$, Wilks' Lambda = 0.90, $\eta_p^2 = 0.103$). Follow-up univariate tests show that male students' physics identity ($F(1,193) = 14.668$, $p < 0.001$, $\eta_p^2 = 0.071$), self-efficacy ($F(1,193) = 10.086$, $p = 0.002$, $\eta_p^2 = 0.050$), and engineering identity ($F(1,193) = 6.739$, $p = 0.01$, $\eta_p^2 = 0.034$) statistically significantly decreased from pre to post. Thus, the results of the one-way repeated measures MANOVA are consistent with the results shown in Table 45. In addition, we also report the percentages of students who selected each choice for each survey item in the pre and post-survey (see Appendix O), which are also consistent with the descriptive statistics shown in Table 45.

**Table 45. Descriptive statistics of female and male students' motivational beliefs at the beginning and end of the course. The sample size is 346 (205 male students and 141 female students). Cohen suggested that a typical value *d* ~ 0.2 be considered a small effect size, *d* ~ 0.5 represents a medium effect size and *d* ~ 0.8 a large effect size.**

| Gender | **Physics self-efficacy** | | **Statistics** | | **Physics interest** | | **Statistics** | |
|---|---|---|---|---|---|---|---|---|
| | pre | post | *p* value | Cohen's *d* | pre | post | *p* value | Cohen's *d* |
| Male | 3.09 | 3.00 | 0.046 | 0.20 | 3.05 | 3.00 | 0.394 | 0.08 |
| Female | 2.97 | 2.78 | <0.001 | 0.42 | 2.86 | 2.72 | 0.033 | 0.26 |
| *p* value | <0.001 | <0.001 | | | 0.001 | 0.003 | | |
| Cohen's *d* | 0.29 | 0.42 | | | 0.37 | 0.46 | | |

| Gender | **Physics perceived recognition** | | **Statistics** | | **Physics identity** | | **Statistics** | |
|---|---|---|---|---|---|---|---|---|
| | pre | post | *p* value | Cohen's *d* | pre | post | *p* value | Cohen's *d* |
| Male | 2.70 | 2.60 | 0.151 | 0.14 | 2.77 | 2.59 | 0.019 | 0.23 |
| Female | 2.44 | 2.36 | 0.310 | 0.12 | 2.42 | 2.23 | 0.043 | 0.24 |
| *p* value | 0.001 | 0.003 | | | <0.001 | <0.001 | | |
| Cohen's *d* | 0.38 | 0.33 | | | 0.48 | 0.43 | | |

| Gender | **Engineering identity** | | **Statistics** | |
|---|---|---|---|---|
| | pre | post | *p* value | Cohen's *d* |
| Male | 3.62 | 3.50 | 0.036 | 0.21 |
| Female | 3.45 | 3.33 | 0.114 | 0.19 |
| *p* value | 0.006 | 0.016 | | |
| Cohen's *d* | 0.31 | 0.26 | | |

### 11.4.2 SEM Models Mediated by Motivational Factors

In this section, we will show the predictive relationships among students' motivational beliefs using SEM models. We first considered a model (Model 1) in which there are only covariances between each pair of constructs: physics 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. The path analysis results are shown in Figure 29 (a). The model fit indices suggest a good fit to the data: CFI = 0.976 (> 0.90), TLI = 0.969 (> 0.90), RMSEA = 0.049 (< 0.08) and SRMR = 0.032 (< 0.08).

As shown in Figure 29 (a), there is a statistically significant regression line from gender to each of physics self-efficacy, perceived recognition, and interest, consistent with Table 45 showing that there are statistically significant gender differences in all of these three motivational constructs. However, the direct effects of gender on both physics identity and engineering identity are statistically insignificant ($p$ = 0.21 for physics identity and $p$ = 0.22 for engineering identity) even though female students' physics and engineering identities are statistically significantly lower than those of male students as shown in Table 45. This result indicates that the gender differences in students' physics and engineering identity are actually mediated by the other three physics motivational constructs. In addition, Figure 29 (a) shows that even though there is a strong covariance among physics self-efficacy, interest, and perceived recognition, students' physics perceived recognition is the strongest predictor of their physics identity ($\beta$ = 0.58). This result is consistent with Hazari et al.'s [21] and Kalender et al.'s prior work [81].

Next, we consider a model (Model 2) in which physics perceived recognition predicts physics self-efficacy and interest, and physics self-efficacy predicts interest. The path analysis results are shown in Figure 29 (b). This model also fits the data very well: CFI = 0.976 (> 0.90),

TLI = 0.969 (> 0.90), RMSEA = 0.049 (< 0.08) and SRMR = 0.032 (< 0.08). We note that the direct effect of gender on physics 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, the direct effects of gender on physics self-efficacy and interest are smaller in Model 2 compared with those in Model 1. This is because in Model 2, physics 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 regression coefficients from gender to physics self-efficacy and interest.

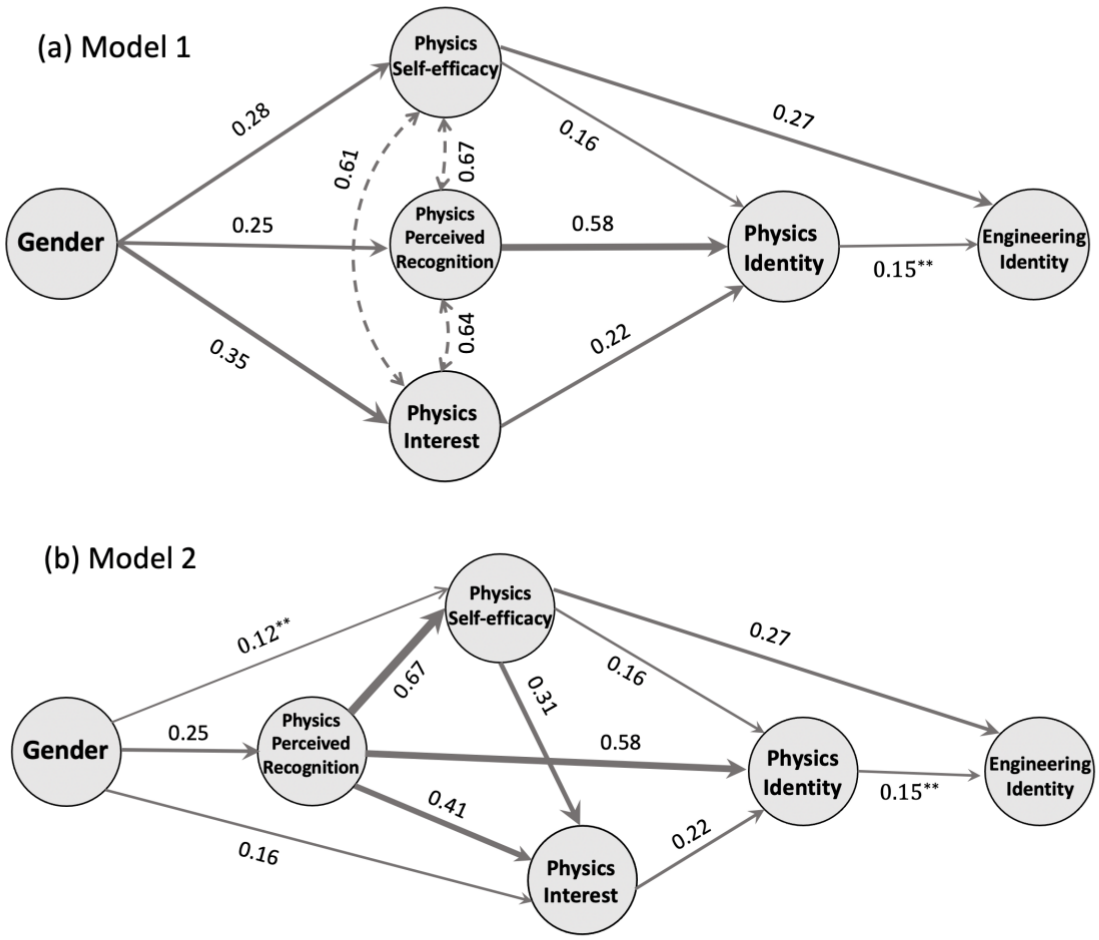



**Figure 29 Results of the path analysis part of the SEM models that show how the relationship between gender and engineering identity is mediated by physics self-efficacy, physics perceived recognition, physics interest and physics identity. (a) In Model 1, physics self-efficacy, perceived recognition and interest are correlated with each other. (b) In Model 2, physics perceived recognition predicts physics self-efficacy and interest, and physics self-efficacy predicts physics interest. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients ($\beta$ values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of $\beta$ with $0.001 \leqslant p < 0.1$ indicated by $^{**}$. All the other regression lines show relations with $p < 0.001$. For clarity, we have removed all statistically insignificant regression paths.**

To further understand the relationships among the motivational constructs in different models, we calculated the coefficients of determination R squared (fraction of variance explained) for each construct in each model (Table 46). We note that the $R^2$ of physics identity is 0.75 and the $R^2$ of engineering identity is 0.15 in both Model 1 and Model 2. This is because in both models, physics and engineering identity are predicted by the same constructs, even though the predictive relationships among the predictors are different in the two models. Similarly, since perceived recognition is only predicted by gender in both models, the $R^2$ of perceived recognition is the same across models. On the other hand, the $R^2$ of physics self-efficacy and interest are larger in Model 2 than in Model 1. This is because in Model 2, physics 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.

**Table 46 Coefficient of determination ($R^2$) for various constructs in different models. All $R^2$ values are statistically significant with *p* values < 0.001. In Model 1, there are only covariances between each pair of constructs: physics self-efficacy (SE), perceived recognition (Recog), and interest. In Models 2, the arrows indicate the direction of the predictive relationships.**

| Models | Constructs | $R^2$ |
|---|---|---|
| Model 1<br>SE + Recog + Interest | Physics perceived recognition | 0.06 |
| | Physics self-efficacy | 0.08 |
| | Physics interest | 0.12 |
| | Physics identity | 0.75 |
| | Engineering identity | 0.15 |
| Model 2<br>Recog → SE → Interest | Physics perceived recognition | 0.06 |
| | Physics self-efficacy | 0.49 |
| | Physics interest | 0.53 |
| | Physics identity | 0.75 |
| | Engineering identity | 0.15 |

## 11.5 Discussion

In this study, we investigated female and male undergraduate engineering students' engineering identity and physics motivational beliefs (including physics identity, self-efficacy, interest and perceived recognition) in a calculus-based introductory physics course. In particular, we focused on the predictive relationships among these engineering and physics motivational constructs, and how these constructs change from pre to post in the course.

Our results reveal that students' engineering identity is directly predicted by their physics identity and self-efficacy. Even though physics interest and perceived recognition do not have direct effects on engineering identity, they all indirectly predict engineering identity through physics identity. Since engineering is interdisciplinary in nature and physics is an important foundational discipline for engineering students, students' perception of their ability to do well in engineering may be influenced by their physics motivational beliefs, which can be influenced by their physics learning experiences. In addition, since physics is one of the disciplines that are believed to require intelligence for success [11,83], doing well in physics may boost students' self-beliefs in learning other subjects such as engineering, while a low self-efficacy in physics may lead students to doubt their ability. According to our previous interviews with students, some students chose to major in engineering because of their earlier experience with mechanics in high school, while some other engineering students considered changing their majors because of their negative experience in a previous physics course.

Another important finding is that even though there are statistically significant gender differences disadvantaging women in all engineering and physics motivational constructs, gender does not directly predict engineering and physics identity. This means that the gender differences in students' engineering and physics identity are mediated through the other three physics motivational beliefs (physics self-efficacy, interest, and perceived recognition). According to a prior study, among an undergraduate engineering population, the gender difference in physics self-efficacy is the largest compared with the gender differences in students' self-efficacy in other STEM disciplines such as chemistry and math [135]. Thus, the gender difference in physics motivational beliefs may help explain the underrepresentation of women in "physics-heavy" engineering disciplines. Our study further indicates that we may be able to reduce the gender gap in students' engineering identity by eliminating the gender differences in physics motivational beliefs. For example, we can create a more inclusive and equitable learning environment for physics learning so that all students feel recognized as people who can do well in physics and other related disciplines.

However, our results show that the two direct predictors of students' engineering identity—physics identity and self-efficacy—actually decreased by the end of the physics course for both male and female students. Moreover, female students' physics self-efficacy dropped even more than male students' did, and the gender difference in physics self-efficacy became larger at the end of the course. These findings may partially explain the result that students' average score on engineering identity also decreased from pre to post, although this change is only statistically significant for male students. Overall, our results indicate that the current learning environment didn't help students develop a stronger engineering identity, and the gender difference in engineering identity is also maintained.

Physics courses are very important for engineering students because not only are they the foundation for engineering courses, but students' physics motivational beliefs can also influence their attitudes and beliefs toward engineering as well as their choice of careers. Due to societal stereotypes, physics is one of the disciplines that have a masculine image and are believed to require a natural ability to excel [156]. Studies have shown that these stereotypes and biases can negatively impact female students' motivational beliefs in physics [70,185]. According to our study, these gender differences in physics motivational beliefs contribute to the gender difference in undergraduate engineering students' engineering identity. Thus, it is important to focus on the role played by physics courses in students' persistence and retention in engineering and engineering school should work with physics department to take effective measures to create an inclusive and equitable learning environment in which all students can develop a stronger identity in both physics and engineering. There are some research-based classroom interventions that have been shown to reduce gender gaps in students' performance in different types of classes (not necessarily focused on engineering students) [177-180]. However, to our knowledge, no intervention has investigated how engineering students' self-efficacy and identity are impacted by these interventions. Their impact on self-efficacy and identity of engineering students from different demographic groups should be studied in future studies. Appropriate interventions could particularly help underrepresented engineering students such as women in physics courses if they were designed well.

In this study, we used single item to measure students' physics identity and engineering identity. Even though these items are commonly used in studies involving physics and engineering identity [58,90,124,125,149,315,329], it would be helpful in future studies to develop more survey items for these identity constructs. Another limitation of the current study is that it only focuses on the underrepresentation of female students and not on other underrepresented demographic groups. In future studies, we intend to investigate motivational beliefs of students from other underrepresented groups such as ethnic/racial minority students. In addition, the data from this study was collected from one research university in the US. Similar studies in different types of institutions and in other countries would also be helpful for developing a deeper understanding of the relationships between students' physics motivational beliefs and their engineering identity.

## 11.6 Conclusion

Students' engineering identity is an important motivational belief that can influence students' retention in engineering as well as their short-term and long-term career goals. Introductory physics courses usually serve as a prerequisite for many engineering courses because they are foundation of many disciplines and contribute directly to engineering. In this study, we investigated how undergraduate engineering students' physics motivational beliefs predict their engineering identity in an introductory physics course. We find that students' engineering identity is directly predicted by their physics identity and self-efficacy and also indirectly predicted by their physics interest and perceived recognition (RQ3). However, our results show that both women and men's physics identity and self-efficacy decreased from the beginning to the end of the course (RQ1). In addition, there are statistically significant gender differences in all physics motivational beliefs and engineering identity, and the gender differences in physics self-efficacy and interest became larger at the end of the course (RQ2). Our results show that students' physics motivational beliefs play an important role in shaping their engineering identity; however, students' physics motivational beliefs decreased after the course, and current learning environment didn't help students develop a stronger engineering identity. Therefore, engineering school should reflect upon the role played by physics courses in undergraduate students' academic trajectory and retention in engineering and work with physics department to make intentional efforts to create an inclusive and equitable learning environment in which all students can develop a stronger identity in both physics and engineering.

# 12.0 Improving Student Understanding of Quantum Measurement Using a Research-Validated Multiple-Choice Question Sequence

## 12.1 Introduction

Prior studies have shown that learning quantum mechanics is challenging for students at all levels including advanced undergraduate and graduate students [330-350]. Our group has been involved in investigating student difficulties in learning quantum mechanics and developing curricula and pedagogies to reduce these difficulties [351-388]. Quantum measurement is one of the key concepts in quantum mechanics. The outcome of a quantum measurement is in general probabilistic and reflects the probabilistic nature of quantum mechanics. Quantum measurement is also fundamental to quantum computation since measurement is necessary to extract the outcome at the end of the computation. In prior studies, quantum measurement has been found to be one of the particularly difficult topics for students [331,340,362,363,372,389].

One source of difficulty of students in learning quantum measurement is the fact that quantum measurements are very different from those in classical mechanics, which students are more familiar with [369]. In classical mechanics, physical observables such as position and energy have well defined values even before the measurement. However, in a generic quantum state, physical observables such as position and energy are usually not well defined, and there are many possible outcomes of measuring an observable [390]. Also, in a quantum system, each observable corresponds to a Hermitian operator, which has a complete set of orthonormal eigenstates and a corresponding set of real eigenvalues. Any quantum state of this system can be expanded as a linear superposition of the complete set of eigenstates of the operator corresponding to any observable. In a measurement, when an observable is measured, the quantum system instantaneously collapses into an eigenstate of the operator corresponding to that observable measured and we obtain a corresponding eigenvalue as the measurement outcome. The probability of yielding (or collapsing into) each eigenstate is given by the absolute square of the projection of the quantum state before the measurement along the eigenstate. The eigenvalue spectrum of an operator can be discrete or continuous or a combination of both. For example, in a one-dimensional infinite potential well, the energy operator $\widehat{H}$ (Hamiltonian) has a discrete eigenvalue spectrum, while the position operator $\hat{x}$ has a continuous eigenvalue spectrum. A measurement of energy will collapse the system into one of the energy eigenstates with well-defined energy, and a measurement of position will collapse the system into an extremely peaked wavefunction (a position eigenstate) with well-defined position between x and *x+dx*.

Another source of difficulty in understanding quantum measurement comes from the fact that for a given quantum system, the probability of measuring each possible value of an observable may change with time depending upon the quantum state and the observable measured [341,391-393]. The time development of quantum states is governed by the time-dependent Schrödinger equation (TDSE) $i\hbar\frac{\partial}{\partial t}\Psi(x,t)=\widehat{H}\,\Psi(x,t)$. For a time-independent Hamiltonian, by solving the TDSE, the time dependence of a generic state is given by $\Psi(x,t)=e^{\frac{-i\widehat{H}t}{\hbar}}\Psi(x,0)$, where $e^{\frac{-i\widehat{H}t}{\hbar}}$ is the time-evolution operator. This equation shows that a quantum state at time $t$ is given by acting with the time-evolution operator on the quantum state at time $t$ = 0. If the initial state $\Psi(x,0)=\Psi_n(x)$ is an energy eigenstate, then $e^{\frac{-i\widehat{H}t}{\hbar}}\Psi(x,0)=e^{\frac{-i\widehat{H}t}{\hbar}}\Psi_n(x)=e^{\frac{-iE_nt}{\hbar}}\Psi_n(x)$, where $E_n$ is the energy corresponding to the energy eigenstate $\Psi_n(x)$. On the other hand, if $\Psi(x,0)$ is not an energy eigenstate, it can be expanded as a linear superposition of energy eigenstates, and the time dependence of this quantum state is given by multiplying each expansion term by a time-dependent phase factor $e^{\frac{-iE_n t}{\hbar}}$, where $E_n$ is the energy corresponding to the $n^{\text{th}}$ energy eigenstate.

As noted, measuring an observable will collapse the system into an eigenstate of the Hermitian operator corresponding to the observable measured. After measuring an observable, the time development of the collapsed quantum state is governed by the TDSE. If the measured observable is energy, the measurement of the observable will collapse the system into an energy eigenstate. Energy eigenstates evolve in time via an overall phase factor, so the probability or probability density of measuring any observable in an energy eigenstate will be time independent. Therefore, energy eigenstates are also called stationary states. On the other hand, if the measured observable corresponds to an operator that does not commute with $\hat{H}$, e.g., position, the outcome of the measurement will not be a stationary state. For example, a measurement of position will collapse the system into a position eigenstate, which can be expanded as a linear superposition of energy eigenstates. The time development of the position eigenstate can be obtained by multiplying each expansion term by a corresponding time-dependent phase factor. Since these phase factors are typically different for each term in the expansion, the state will evolve in time in a non-trivial manner, and the probability or probability density of measuring most observables after the position measurement will be time dependent. The only exception is the probability of measuring energy or observables whose corresponding operators commute with $\hat{H}$ since they are constants of motion. This is because the probability of measuring each energy only depends on the modulus of the expansion coefficients, which is independent of time. Therefore, the probability of measuring a particular value of energy is time independent even in a non-stationary state after a position measurement.

Our prior studies have shown that students have many common difficulties with quantum measurement after lecture-based instruction [331,361-363]. For example, some students have difficulties in distinguishing between eigenstates of operators corresponding to different observables [362,363]. Research shows that some students incorrectly think that the eigenstates of any Hermitian operator are stationary states, and some students do not relate the concept of stationary state with the special nature of the time evolution of that state [362,363]. In addition, some students incorrectly think that an operator acting on a state corresponds to a measurement of the corresponding observable [362,363]. The prior studies also show that many students have difficulties in identifying the probability of measuring different energies in a given state, especially when the quantum state is not explicitly given as a linear superposition of energy eigenstates, and they also have difficulties in identifying the probability density of measuring position given a wavefunction [362,363]. Some other common difficulties relate to the time development of a quantum state after a measurement [362,363]. For example, prior research shows that some students incorrectly think that a system will stay in a position eigenstate after measuring position, and some others think that a system will evolve back to the initial state a long time after a measurement of an observable.

In our prior studies, we developed, validated, and implemented quantum interactive learning tutorials (QuILTs) to help students learn many quantum mechanics topics including quantum measurement [363], and the implementation of these materials showed encouraging results [355,363,364,367,371-373,378-382,387]. A QuILT consists of learning sequences which are developed and validated based on cognitive task analysis from both the expert and student perspectives and extensive research on students' common difficulties in learning quantum mechanics. QuILTs use a guided inquiry-based approach to keep students engaged and build a good knowledge structure.

In addition to interactive learning tutorials, studies have shown that multiple-choice questions implemented in classes with peer instruction are effective for students' learning. This method was first popularized by Eric Mazur in the physics community [394]. In Mazur's approach, an instructor poses a multiple choice question, and students first think by themselves and answer the question anonymously using an electronic response system (clickers) [394]. Then, the students discuss their thoughts with their peers, during which they can compare their answers with each other and explain their reasoning. After the discussion, the instructor may ask for volunteers to share their discussion or give students feedback based on students' performance. In this process, students can have immediate feedback from their peers and instructor, and the instructor can obtain an understanding of students' common difficulties and the percentage of students who understand the concepts. This method using multiple-choice questions with peer instruction has been shown to be effective and relatively easy to incorporate in classes without the need to greatly restructure them [395]. Moreover, prior studies show that the use of multiple-choice questions in class can be more effective if the questions are carefully sequenced to systematically help students with a particular theme that they may be struggling with [363]. In our prior studies [360,384,386,396,397], we have developed and validated multiple-choice question sequences (MQS) for several topics in quantum mechanics such as the Stern–Gerlach experiment [384] and time-development of two-state quantum systems [391]; however, there has not been a MQS developed for quantum measurement. In this paper, we describe the development, validation and evaluation of a multiple-choice question sequence for helping students learn quantum measurement.

## 12.2 Methodology

### 12.2.1 Development and Validation of Research-Validated Multiple-Choice Question Sequence

To develop the quantum measurement MQS, we took inspiration from the learning objectives and inquiry-based guided learning sequences of the quantum measurement QuILT [363], which had been validated and implemented in previous years with students in junior-/senior-level quantum mechanics courses at a large research university in the US. We paid specific attention to when and how different concepts are invoked and build upon students' prior knowledge in the QuILT, and what were the common difficulties students had and how those difficulties were reduced after the implementation of the QuILT [363]. In addition, we took advantage of the cognitive task analysis from both the expert and student perspectives conducted when developing the QuILT and focused on what scaffolding supports are needed to reduce students' difficulties and help them develop a good knowledge structure of quantum measurement. After identifying and contemplating upon students' common difficulties with quantum measurement and scaffolding supports that have been shown to be effective as well as a proper order to invoke different concepts in the MQS, we drafted, discussed, and iterated a MQS among researchers and several other faculty members. While some questions in this MQS are adapted from the quantum measurement QuILT [363], we also created new questions inspired by students' written responses to the QuILT and its corresponding pre-test and post-test.

The MQS was designed to help students improve their conceptual understanding of quantum measurement, so we avoided complicated calculations to avoid causing cognitive overload for students. Since the MQS is designed to be used in class with peer instruction, we ensured that the number of the questions was such that it can be administered during limited class time while still covering the common difficulties students have in understanding quantum measurement. In addition, we carefully designed each alternative choice in each multiple-choice question and incorporated the common incorrect responses that we found in previous interviews and students' written responses to the QuILT and its corresponding pre- and post-tests. Thus, the MQS provides students opportunities to think about common difficulties, struggle productively, and get immediate feedback from their peers and instructors. Moreover, we balanced the difficulty level of the questions that build on each other and avoided using either too difficult or too easy questions. If a question is too difficult and most students choose the distractors, then students' inaccurate understanding may be reinforced during peer instruction. On the other hand, if a question is too easy, the learning potential of peer instruction might be limited [398,399]. Another feature of the quantum measurement MQS is that it includes both concrete and abstract questions. Concrete questions provide opportunities for students to apply their knowledge in a concrete context, which help them learn applications of the quantum measurement concepts in specific contexts. Concrete questions are usually followed by or integrated with abstract questions, which can help students generalize their understanding of the concepts and transfer their knowledge across contexts.

In the quantum measurement MQS, the questions are carefully sequenced to build on each other. For example, the same concept may be applied in different contexts or different concepts may be applied in similar contexts in two consecutive questions. Thus, students can compare and contrast the premise of consecutive questions to solidify their understanding of the concepts and build their knowledge structure. In addition to the class discussion using peer instruction mentioned earlier, we also added some discussion slides between the questions in the MQS, which can be used by instructors to review and emphasize the important concepts in the previous questions and lead general class discussion on some broader themes related to those questions. Some discussion slides that instructors can use with the MQS also include visual representation of quantum measurements to help students build physical intuition and resolve possible conflicts between their understanding and the correct conceptual model.

After the initial development of the quantum measurement MQS using the learning objectives adapted from the inquiry-based guided sequences in the QuILT [363] as well as empirical data from student responses to existing individually validated questions in previous years, we conducted hour-long individual interviews with five students to further validate the MQS. We asked students to answer the questions in the MQS and think aloud as they answered them so that we could understand their reasoning process. We did not disturb them when they thought aloud in order to not disrupt their thought processes. After each MQS, we first asked students for clarification of the points they may not have made, then we led discussions with them on each choice in the MQS question as appropriate. The MQS was iterated multiple times as the interviews were conducted. The feedback from students helped in fine-tuning and refining some new questions that were developed and integrated with the existing ones to construct the sequence of questions in the quantum measurement MQS and to ensure that the questions were unambiguously worded. In the interviews, students showed some common difficulties in understanding quantum measurements, which are consistent with the results of our prior studies [362,363]. Moreover, the interviews showed that after working through the whole MQS, students' difficulties with many concepts related to quantum measurement were reduced. The students also reported that they found the scaffolding provided by the sequenced questions and discussion slides helpful. The interviews also helped us to further improve the MQS by simplifying the wording to reduce students' cognitive load.

### 12.2.2 Learning Objectives and Structure of the Quantum Measurement MQS

The quantum measurement MQS in its final iteration includes four sections (see Appendix P). Each section focuses on one aspect of students' common difficulties in understanding quantum measurement. The first section of the MQS includes two questions (MQS 1.1 and 1.2). The learning objective of this section is to help students review basic concepts pertaining, e.g., to eigenvalues and eigenstates of energy or position operators. The second section of the MQS includes three questions (MQS 2.1-2.3), which help students learn about the postulates related to quantum measurement, e.g., a measurement of an observable collapses the system into an eigenstate of the operator corresponding to that observable and returns the eigenvalue corresponding to the eigenstate. In particular, the learning objectives include being able to identity the possible outcomes of the measurement of an observable and calculate the probability (for operators with discrete eigenvalue spectra) or probability density (for operators with continuous eigenvalue spectra) of measuring each outcome for a given quantum state. MQS 2.2 helps students to recognize that an operator acting on a state is not equivalent to a measurement of the corresponding observable in that state, which has been shown to be a common difficulty in our prior studies [362,363]. The third section of the MQS includes six questions (MQS 3.1-3.6), which aim to help students learn the time development of the quantum state after a measurement of an observable. In particular, MQS 3.1 and 3.2 help students to learn about the time development of a generic state and a stationary state (energy eigenstate) and the fact that one can always expand a generic state as a linear superposition of energy eigenstates. In MQS 3.3-3.5, we use both mathematical and pictorial representations to help students learn about how a quantum system evolves in time after a measurement of energy or position, respectively. MQS 3.6 helps students to learn about whether the probability (for operators with discrete eigenvalue spectra) or probability density (for operators with continuous eigenvalue spectra) of measuring an observable depends on time. This question also prepares students for the next section of the MQS, pertaining to consecutive measurement. In the last section (MQS 4.1-4.5), MQS 4.1 and 4.2 help students identify the possible outcomes and the corresponding probability densities of measuring position immediately after or a long time after a measurement of energy. MQS 4.3 and 4.4 help students identify the possible outcomes and the corresponding probability densities of measuring position immediately after or a long time after a measurement of position. MQS 4.5 helps students identify the possible outcomes and the corresponding probabilities of measuring energy immediately after or a long time after a measurement of position.

### 12.2.3 In Class Implementation

The quantum measurement MQS can be used in many different situations. For example, it can be integrated with lecture as an in-class activity to help students apply their newly learned knowledge. The MQS can also be given at the end of each class to summarize and review the important concepts covered in the class. In addition, it can be given after all lectures on quantum measurement to help reduce common difficulties, resolve possible conflicts between students' understanding and the correct conceptual model, make connections between different concepts, and help students build a good knowledge structure. In this study, the MQS was implemented by an instructor after traditional lecture instruction on quantum measurement in a junior-/senior- level quantum course in a large research university in the US. There were 25 students in this class, who are all physics majors. The implementation involved students answering the questions anonymously using an electronic response system (clickers) [394] followed by peer instruction and feedback from the instructor. We carefully considered effective strategies for in-class implementation including when the instructor should provide opportunities for class discussion to promote students' engagement and take advantage of productive struggles.

After the traditional lecture-based instruction on quantum measurement, students were administered a pre-test before the implementation of the quantum measurement MQS. After the implementation of the MQS, which took three class periods, students completed a post-test. The post-test was a slightly modified version of the pre-test, containing changes such as the initial wavefunction, but remaining conceptually similar. The pre-test and post-test were adapted from the quantum measurement QuILT developed and validated by Zhu et al. [363]. We adjusted the questions to make them align with the learning objectives introduced earlier. We revalidated the pre-test and post-test by obtaining feedback from two faculty members and interviewing 5 students to make sure that the questions can be interpreted as intended by students.

There are 4 questions in the pre-test and post-test (see Appendix Q). The first question is a multiple-choice question, and the rest of the questions are all open-ended questions that ask for both the answer and the corresponding reasoning. Question 2 includes two sub-questions (2a, 2b), and questions 3 and 4 each include five sub-questions (3a-3e, 4a-4e). Q1 tests students' understanding of what happens when the Hamiltonian or position operator acts on a generic state and whether students realize that any state can be expanded as a linear superposition of a complete set of energy eigenstates. Questions 2a and 2b test students' understanding of the possible outcomes and the corresponding probability/probability density of measuring energy/position in a given quantum state. Questions 3 and 4 test students' understanding of the time development of a quantum system after measuring energy/position and the possible outcomes and corresponding probability/probability densities of measuring energy/position in consecutive sequence.

Students' responses to the pre-test and post-test were graded by two researchers. Each sub-question in the pre-test and post-test was scored out of 1 or 2 points, with partial credit assigned to answers that were correct, but for which either incorrect justification or no justification was provided if reasoning was requested. The inter-rater reliability was better than 95%.

## 12.3 Results and Discussion

Table 47 compares students' performances on the pre-test (after traditional lecture-based instruction) and on the post-test (after students had engaged with the quantum measurement MQS). The normalized gain was calculated as $g$ = (post% – pre%)/(100% – pre%) [229,240]. Effect size was calculated as Cohen's $d = (\mu_{post} - \mu_{pre})/\sigma_{pooled}$, where $\mu_{pre}$ and $\mu_{post}$ are students' average correctness in the pre-test and post-test and $\sigma_{pooled}$ is the pooled standard deviation, which is the weighted average of standard deviations of pre- and post-test [240].

**Table 47 Comparison of the mean pre-test and mean post-test scores on each question, normalized gains, and effect sizes for students who engaged with the quantum measurement MQS (N=23 for the pre-test and N=25 for the post-test).**

| Question | Pre-test mean | Post-test mean | Normalized gain | Cohen's *d* |
|---|---|---|---|---|
| 1 | 81% | 84% | 0.15 | 0.12 |
| 2a | 93% | 98% | 0.69 | 0.26 |
| 2b | 43% | 84% | 0.72 | 1.17 |
| 3a | 86% | 93% | 0.54 | 0.27 |
| 3b | 91% | 92% | 0.08 | 0.02 |
| 3c | 63% | 74% | 0.30 | 0.24 |
| 3d | 57% | 72% | 0.36 | 0.39 |
| 3e | 48% | 70% | 0.43 | 0.48 |
| 4a | 13% | 64% | 0.59 | 1.20 |
| 4b | 65% | 92% | 0.77 | 0.76 |
| 4c | 48% | 94% | 0.89 | 1.22 |
| 4d | 0% | 42% | 0.42 | 1.20 |
| 4e | 33% | 62% | 0.44 | 0.27 |

As can be seen in Table 47, students' average correctness for questions 1, 2a, 3a, and 3b in the pre-test were higher than 80%, which indicates that after traditional lecture-based introduction, students had a relatively good understanding of single energy measurement and consecutive energy measurements. However, the average correctness for other questions were low in the pre-test. In particular, students' average correctness on questions 4a, 4d, and 4e were below 40% in the pre-test. Question 4a tests students' understanding of the wavefunction immediately after a position measurement, which is a delta function, and questions 4d and 4e assess students' understanding of the possible outcomes of an energy measurement immediately after or a long time after the measurement of position in 4a. Students' average correctness for the rest of the questions (2b, 3c, 3d, 3e, 4b, and 4c) was around 50% on the pre-test, which indicates that students also had difficulties in understanding the concepts assessed by these questions. In particular, question 2b assesses students' understanding of the possible outcomes and the corresponding probability densities of measuring position in a given quantum state. Questions 3c, 3d, and 3e assess students' understanding of the possible outcomes of energy/position measurements after a measurement of energy. Questions 4b and 4c assess students' understanding of the possible outcomes of a position measurement made immediately after or a long time after a measurement of position.

By comparing students' average correctness for different questions, we find that students had more difficulties in questions involving position measurement (such as 4b) than those involving energy measurement (such as 3b). Students also had more difficulties in identifying the possible outcomes of consecutive measurements (such as 3c) than of a single measurement (such as 2a and 3a). In addition, students had more difficulties on questions involving consecutive measurements with a time interval between them (such as 3c) than those involving measurements made in immediate succession (such as 3b). Moreover, they had more difficulties on questions involving measuring energy after a position measurement (such as 4d and 4e) than those involving measuring position after an energy measurement (such as 3d and 3e). Below, we discuss the difficulties students found in detail and compare student performance on the pre-test and post-test.

Table 47 shows an overall improvement in students' performance from the pre-test to post-test. In particular, the major improvements were on questions 2b, 4b, and 4c. We note that the average correctness was around 50% for questions 2b, 4b, and 4c on the pre-test, and it improved to around 90% on the post-test. Question 2b asks about the probability density of measuring position in state $\Psi(x,0) = \sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$ for one dimensional infinite square well (0<x<a). One common difficulty we found from student written responses is that some students could not distinguish between the eigenstates of energy and position operators. For example, some students wrote that the probability density of measuring position in the state $\Psi(x,0)$ is "$|\Psi_1(x)|^2$ or $|\Psi_2(x)|^2$". One possible reason for this difficulty could be that since wavefunctions are often discussed in the context of energy eigenstate or written as a linear superposition of energy eigenstates, students may assume that each term in the linear superposition corresponds to a possible measurement outcome regardless of the observable measured. In addition, some students incorrectly thought that the position operator has discrete eigenvalues. For example, some students wrote the possible outcomes of a position measurement as "$x_n$".

Questions 4b and 4c assess students' understanding of the possible outcomes of a position measurement made immediately after or a long time after the first measurement of position. We note that on the pre-test, some students wrote that the second measurement of position will always yield the same result as the first position measurement regardless of whether there is a time interval between the two measurements because "the first measurement collapsed the state" or "the position measurement is time independent". These types of answers reflect students' difficulties in distinguishing between stationary states and eigenstates of other operators, e.g., position, that do not commute with the Hamiltonian. Some other students noted that position eigenstates evolve with time, but they thought that the time development of position eigenstates is such that the probability density of measuring position is not affected. For example, one student who stated that the probability of measuring position is time-independent explained, "the time [phase] factor will be all that changed”. This reasoning also indicates that students may have difficulty in distinguishing between the time development of energy eigenstates and those of other operators such as position. As shown in Table 47, after the implementation of the MQS, students had significant improvement in their understanding of single position measurement and consecutive position measurements (evidenced by performance on questions 2b, 4b, and 4c).

As shown in Table 47, the average correctness for questions 4a, 4d, and 4e was very low on the pre-test. Question 4a asks students to identify the wavefunction after a measurement of position in state $\Psi(x,0) = \sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$ with outcome $x_0$, and questions 4d and 4e ask students to identify the possible outcomes of an energy measurement made immediately after or a long time after the position measurement in 4a. As we can see in Table 47, the correctness for question 4a was only 13% in the pre-test. The most common student response was that the wavefunction right after the position measurement is $\sqrt{\frac{2}{7}}\Psi_1(x_0) + \sqrt{\frac{5}{7}}\Psi_2(x_0)$, which is obtained by simply replacing the variable $x$ in the initial state $\Psi(x,0)$ with position $x_0$. However, this expression is actually the probability amplitude at $x = x_0$ before the position measurement. This result shows that some students had difficulties in recognizing that a position measurement will instantaneously collapse the wavefunction to a delta function in position. Another common incorrect answer to question 4a was “$\Psi_1(x_0)$ or $\Psi_2(x_0)$”, which also shows that some students had difficulties in distinguishing between energy eigenstates and position eigenstates as discussed earlier.

The analysis of students’ written responses shows that after the implementation of the MQS, many students were able to recognize that a delta function is the outcome for the state right after a position measurement. Question 4d asks students to identify the probability of measuring

$E_1$ immediately after the position measurement in 4a. This question requires students to expand the delta function as a linear superposition of energy eigenstates and then express the probability of measuring $E_1$ using the expansion coefficient for the ground state. We note that on the pre-test, no students provided the correct response. Many students wrote that the probability of measuring $E_1$ is $\frac{2}{7}$, obtained by simply squaring the coefficient of $\Psi_1(x)$ in the initial state $\Psi(x, 0)$, which is consistent with the difficulty in identifying the wavefunction after a position measurement. Table 47 shows that on the post-test, many students were able to express the probability of measuring $E_1$ by projecting the delta function along state $\Psi_1(x)$. For example, on the post-test, some students wrote statements such as "the system is in a superposition of an infinite number of energy eigenstates and the probability of measuring $E_1$ is $|\langle\Psi_1(x)|\delta(x - x_0)\rangle|^2$". Even though keeping the wavefunction with *x* in Dirac notation is not accurate, this example shows students' improvement at the conceptual level about probability of measuring energy after the implementation of the MQS.

Question 4e asks about the time dependence of the probability of measuring $E_1$ after a position measurement. This question is challenging for students because the system itself is not in a stationary state and will evolve in time in a non-trivial manner, but the probability of measuring each eigenenergy is time independent since the Hamiltonian is a constant of motion. Students need to recognize that any state can be expanded as a linear superposition of energy eigenstates, and each expansion coefficient evolves in time via a different phase factor, but the phase factor will cancel out when calculating the probability of measuring a specific energy (by taking the absolute square). A common difficulty that students had was that they thought that since the state is not a stationary state, the probability for measurement of a particular value of energy should also change with time. Moreover, some students stated that the probability of measuring $E_1$ will be $\frac{2}{7}$ with explanation such as this, "after a long time, the wavefunction will evolve back to $\Psi(x, 0) =$

$\sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$". In the post-test, many students recognized that the probability of measuring different energies is time independent. These results show that students' understanding of energy measurements after a position measurement improved after the implementation of the quantum measurement MQS.

In addition, Table 47 shows some improvement in students' performances on questions 3c, 3d, and 3e. The correctness of these questions on the pre-test was 50% - 60% and improves to around 70% on the post-test. These three questions ask about the possible outcomes of measuring energy or position immediately after or a long time after a measurement of energy. Question 3c asks about the possible outcomes of measuring energy a long time after an energy measurement that yields $E_1$. One common difficulty with this question was that some students thought that the system would evolve to a linear superposition of energy eigenstates after the first measurement of energy. For example, one student stated that the second energy measurement will yield a different value from the first one because "the wavefunction will no longer be collapsed, so you could measure $E_1$ or $E_2$ [for the second measurement]". Question 3d asks about the probability density of measuring position at $x = x_0$ immediately after an energy measurement that yields $E_1$. Even though many students correctly identified that after the energy measurement, the state will collapse to $\Psi_1(x)$, some of them still had difficulties in identifying the probability density of the position measurement. For example, one student stated, "the probability density is 100% at $x = x_0$ because the system has collapsed to $\Psi_1(x)$". This response again shows students' difficulty in distinguishing between eigenstates of different operators corresponding to observables (e.g., Hamiltonian and position).

In addition, we note that some students had difficulties in writing the probability density for a given wavefunction, which was reflected in students' responses to both questions 2d and 3d.

As examples, some students stated something similar to this student explanation, "the probability density of measuring position is 0 because position is a single point", and some put a position operator in front of the wavefunction, e.g., "$|\hat{x}\Psi_1(x)|^2$", or multiplied the wavefunction by $x$, e.g., "$|x\Psi_1(x)|^2$" to represent the probability density. Some students had difficulties in distinguishing between probability density and probability and used an integral such as $\int_0^{x_0}|\Psi_1(x)|^2dx$ to represent probability density. Question 3e asks whether the probability density of measuring position changes with time after an energy measurement. Students who incorrectly stated that the probability density will change with time usually had two types of reasoning: the state itself will evolve with time after the energy measurement, or since the position operator does not commute with the energy operator (Hamiltonian), the measurement of position is time dependent even in an energy eigenstate. In the post-test, many students correctly stated that in a stationary state, the probability/probability density of measuring any observable will be time independent.

Even though most students answered questions 1, 2a, 3a and 3b correctly on the pre-test as shown in Table 47, we still find some common difficulties among students who did not answer these questions correctly. For example, in our interviews, students who chose $\widehat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E_n}\psi_\mathrm{n}(\mathrm{x})$ as a correct statement in question 1 explained that acting with an operator on a state will collapse this state to an eigenstate of the operator corresponding to the observable, which is consistent with the findings of prior studies [362,363]. We also found that some students were confused about the normalization of quantum states after energy measurement. For example, question 3a asks about the normalized state after a measurement of energy in state $\Psi(x,0) = \sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$ that yields $E_1$, and some students incorrectly stated $\sqrt{\frac{2}{7}}\Psi_1(x)$. A summary of the student difficulties found in written pre-test and post-test is presented in Table 48.

**Table 48 Summary of conceptual difficulties addressed by the MQS. Below, specific examples of difficulties with quantum measurement along with the MQS questions that address them are listed. In the comments section, we include the relevant pre-/post-test question numbers and whether post-test showed 'some' improvement or 'major' improvement compared to the pre-test.**

| **Difficulties** | **MQS #** | **Comments** |
|---|---|---|
| Incorrectly thinking that an operator acting on a state corresponds to a measurement of the corresponding observable. | 1.1, 1.2, 2.2 | 1<br>Some improvement |
| Difficulties in distinguishing between measurement outcome, e.g., energies and probability of measuring it (related to the coefficients in the expansion of the wave function in terms of energy eigenstates). | 2.1, 2.3 | 2a<br>Some improvement |
| Difficulties in appropriately distinguishing between energy and position measurements or overgeneralizing some ideas, e.g., from energy to position measurement:<br>• Stating that the probability density of measuring position in state $\sqrt{\frac{2}{7}}\Psi_1(x)+\sqrt{\frac{5}{7}}\Psi_2(x)$ is "$\|\Psi_1(x)\|^2$ or $\|\Psi_2(x)\|^2$".<br>• Stating that the probability density of measuring position in state $\sqrt{\frac{2}{7}}\Psi_1(x)+\sqrt{\frac{5}{7}}\Psi_2(x)$ is "$\frac{2}{7}+\frac{5}{7}=1$".<br>• Stating that the probability of measuring ground state energy after a position measurement is 100% because the state has collapsed.<br>• Stating that a position measurement in a given state will yield discrete possible outcomes, e.g., "$x_n$". | 2.1, 2.3, 3.1, 4.1 | 2b, 3d<br>Major improvement |
| Not writing the correct probability density for measuring position for a given wavefunction.<br>• Using an incorrect integral to represent probability density, e.g., "$\int \hat{x}\Psi_1(x)dx$" or "$\int_0^a x_0\|\Psi_1(x_0)\|^2 dx$".<br>• Acting with the position operator on the wavefunction, e.g., "$\|\hat{x}\Psi_1(x)\|^2$" or multiplying the wavefunction by $x$, e.g., "$\|x\Psi_1(x)\|^2$" to represent probability density.<br>• Stating that the probability density is 0 because position is a single point. | 2.3, 4.1, 4.2 | 2b, 3d<br>Some improvement |
| Not writing the correct normalized state after an energy measurement, e.g., stating that the normalized state after a measurement of energy in state $\Psi(x,0)=\sqrt{\frac{2}{7}}\Psi_1(x)+\sqrt{\frac{5}{7}}\Psi_2(x)$ that yields $E_1$ is $\sqrt{\frac{2}{7}}\Psi_1(x)$ | 3.3, 4.5 | 3a<br>Some improvement |

| | | |
|---|---|---|
| Difficulties with measurements made in an energy eigenstate at future times<br>• Not recognizing that energy eigenstates evolve in time via a trivial overall time-dependent phase factor.<br>• Not recognizing that the probability/probability density of measuring any observable (with no explicit time-dependence) is time independent in an energy eigenstate. | 3.5, 3.6, 4.1, 4.2 | 3c, 3e<br>Some improvement |
| Not recognizing that the probability of measuring a particular value of energy is time independent regardless of state since energy is a constant of motion. | 3.6, 4.5 | 3c,4e<br>Some improvement |
| Stating that states will evolve back to the original state (before measurement) after a long time after a measurement of position or energy. | 3.3, 3.5 | 3c, 3e, 4c, 4e<br>Some improvement |
| Not recognizing that the wavefunction will collapse to a delta function after a position measurement.<br>• Stating that the wavefunction after a position measurement that yields $x_0$ is $\sqrt{\frac{2}{7}}\Psi_1(x_0)+\sqrt{\frac{5}{7}}\Psi_2(x_0)$.<br>• Stating that the wavefunction after a position measurement that yields $x_0$ is $\Psi_1(x_0)$ or $\Psi_2(x_0)$. | 3.3, 3.4, 4.5 | 4a<br>Major improvement |
| Not recognizing that consecutive position measurements made in immediate succession will yield the same outcomes. | 4.3 | 4b<br>Major improvement |
| Not recognizing that the probability density for position measurement in a non-stationary state changes with time. | 3.2, 3.6, 4.4 | 4c<br>Major improvement |
| Stating that a position eigenstate does not evolve in time (system is stuck in a position eigenstate because it is an eigenstate) or it evolves in time via a trivial overall time-dependent phase factor so that the probability density for position measurement will be time-independent if the system was in a position eigenstate at time $t = 0$ | 3.3, 3.4, 4.4 | 4c, 4e<br>Major improvement |
| Difficulties in recognizing the need to expand a generic wavefunction as a linear superposition of energy eigenstates in order to identify the probability of measuring each energy. | 4.5 | 4d, 4e<br>Major improvement |
| Difficulties in calculating the expansion coefficients of each energy eigenstate when a generic wavefunction is not explicitly written as a linear superposition of energy eigenstates. | | |

## 12.4 Conclusions and Summary

In this study, we developed and validated a multiple-choice question sequence (MQS) intended to improve students' understanding of quantum measurements. This MQS was developed based on students' common difficulties with quantum measurement, e.g., the difficulties in distinguishing between eigenstates of different operators, the difficulties with the time development of a quantum system after measuring an observable, and the difficulties in identifying the possible outcomes of consecutive measurements with or without time evolution. The development of the MQS including the learning objectives was guided by the learning objectives and inquiry-based guided learning sequences of the interactive tutorial (QuILT) that we previously developed for the same topic. The questions in the MQS were carefully sequenced to build on each other and help students organize, extend, and repair their knowledge structure and develop a better understanding of different issues pertaining to quantum measurement. There are 16 multiple-choice questions divided into four sections, each of which focuses on one aspect of students' common difficulties in understanding quantum measurement. Each section includes concrete questions followed by abstract questions, so students can apply their knowledge in concrete contexts as well as generalize their understanding of the concepts and transfer their knowledge across contexts.

In this study, the quantum measurement MQS was implemented in a junior-/senior-level quantum mechanics course in a large research university in the US. The implementation involved students answering the questions anonymously using an electronic response system (clickers) [394] followed by peer instruction and feedback from the instructor. We used a research-validated pre-/post-test to assess students' understanding before and after the implementation of the MQS. The results show that the average correctness for all questions improved. The major improvement was in students' understanding of single position measurement, consecutive position measurements, and energy measurements after a position measurement. There were also some improvements in questions pertaining to measurements of energy or position immediately after or a long time after an energy measurement. We note that, for some post-test questions, there is still room for further improvement, so we plan to add more checkpoints for discussion to further enhance students' understanding. In addition, in future studies, it would also be helpful to investigate the effectiveness of the quantum measurement MQS in other modes of classroom implementations.

## 13.0 Future Directions

The research presented here gives insight into the role played by female and male students' perception of the inclusiveness of the learning environment in predicting their motivational beliefs and academic performance in college calculus-based introductory physics courses. These findings can be very useful for instructors interested in making these courses more equitable and inclusive. However, more research is needed.

The present studies only focus on female students and not on other underrepresented demographic groups. In future studies, it would be useful to carry out similar investigations accounting for intersectional perspectives, e.g., with female and male students from different ethnic/racial groups, and how their perceptions of the inclusiveness of learning environment predict their course outcomes.

In addition, the research presented here was conducted at a large public research university in the US. Similar studies in different types of institutions such as small colleges and universities in the US and in other countries would also be helpful for developing a deeper understanding of the relationships between students' perception of the inclusiveness of the learning environment and their course outcomes. Moreover, the present studies were 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 a studio physics class and an active learning class, affect students' perception of the inclusiveness of the learning environment and their motivational and academic outcomes.

Future studies can also investigate students' motivational beliefs and academic performance in the classes in which there is an intentional focus on equity and inclusion or use research-based classroom interventions with control groups to further study their effect on the perception of the inclusiveness of the learning environment and on students' course outcomes.

## Appendix A Coefficient of Determination for Each Construct in Each Model

Here, we present the coefficient of determination $R^2$ for each construct in Models 1-4 (see Table 49), which represents the proportion of each construct's variance explained by each statistically equivalent SEM model [98] (see Supplemental Material [111] for results of $R^2$ in some other statistically equivalent SEM models). We note that these equivalent models have the same coefficient of determination $R^2$ for the physics identity but $R^2$ is in general different for each construct that mediates the relation between gender and physics identity. As shown in see Table 49, $R^2$ of identity is 0.74 in all of the models (as expected), which means that each equivalent model explains 74% of the variance in the physics identity. However, the $R^2$ of self-efficacy, interest, and perceived recognition are different in different models. In particular, if a motivational construct is only predicted by gender, its $R^2$ will be smaller than if it is also predicted by other motivational constructs. For example, see Table 49 shows that in Models 1 and 2, in which self-efficacy is only predicted by gender, the $R^2$ of self-efficacy is only around 0.08, while in Models 3 and 4, in which self-efficacy is also predicted by another motivational construct, $R^2$ of self-efficacy becomes 0.39 in Model 3 and 0.50 in Model 4. This is because when self-efficacy is predicted by more constructs, more variance in self-efficacy is explained by the model.

**Table 49 Coefficient of determination ($R^2$) for various constructs in different statistically equivalent SEM models. All $R^2$ values are statistically significant with $p$ < 0.001. In Model 1, there are only covariances between each pair of constructs: physics self-efficacy (SE), perceived recognition (Recog), and interest. In Models 2-4, the arrows indicate the direction of the predictive relationships**

| NO. | Models | Variables | $R^2$ |
|---|---|---|---|
| 1 | SE + Recog + Interest | Perceived Recog | 0.06 |
| | | Self-efficacy | 0.08 |
| | | Interest | 0.10 |
| | | Identity | 0.74 |
| 2 | SE → Interest → Recog | Perceived Recog | 0.59 |
| | | Self-efficacy | 0.08 |
| | | Interest | 0.41 |
| | | Identity | 0.74 |
| 3 | Interest → SE → Recog | Perceived Recog | 0.59 |
| | | Self-efficacy | 0.39 |
| | | Interest | 0.10 |
| | | Identity | 0.74 |
| 4 | Recog → SE → Interest | Perceived Recog | 0.06 |
| | | Self-efficacy | 0.50 |
| | | Interest | 0.52 |
| | | Identity | 0.74 |

## Appendix B Students' Pre- and Post-Motivational 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 post- motivational beliefs in the second year studied (Table 50).

As shown in Table 50 and Table 51, 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.

**Table 50 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* ~ 0.2 be considered a small effect size, *d* ~ 0.5 represents a medium effect size and *d* ~ 0.8 a large effect size.**

| Physics 1 Gender | **Self-efficacy** | | **Statistics** | | **Interest** | | **Statistics** | |
|---|---|---|---|---|---|---|---|---|
| | 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 51 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 | | |

## Appendix C Percentages of Students Who Selected Each Choice for Each Survey Item I

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 52) and male students (Table 53) 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 Table 52 and Table 53, 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 descriptive statics shown in Table 12, 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 Table 52 and Table 53, 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 12 showing that there were statistically significant gender differences in all motivational constructs studied.

**Table 52 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 (SE) and interest (Int) items have the response scale: 1= NO!, 2 = no, 3 = yes, and 4 = YES!, while the perceived recognition (Recog) and identity (Idt) items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.**

| Survey items | Physics 1 | | | | Physics 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | 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 53 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 (SE) and interest (Int) items have the response scale: 1= NO!, 2 = no, 3 = yes, and 4 = YES!, while the perceived recognition (Recog) and identity (Idt) items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.**

| Survey items | Physics 1 | | | | Physics 2 | | | |
|---|---|---|---|---|---|---|---|---|
| | 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% |

# Appendix D Pearson Correlation between the Constructs

Before performing structural equation modeling, we calculated the Pearson correlation coefficients pairwise between the constructs [100]. As shown in Table 54, these constructs have correlations with each other, but all of the correlations are lower than 0.6 except between pre and post-FCI [109]. In particular, in Table 54, the strongest correlation is between pre-FCI and post-FCI with correlation coefficient 0.81, which means that students' FCI scores at the end of the course are highly related to their scores at the beginning of the course. In addition, Table 54 shows that the correlation between high school GPA and students' sense of belonging is not statistically significant at either the beginning or end of the course. This may imply that students with a higher high school GPA do not necessarily have a higher sense of belonging in the physics course.

**Table 54 Pearson correlation coefficients of the constructs. p values are indicated by $^{**}$ for $0.001 \leq p < 0.01$ and $^{ns}$ for $p > 0.05$ (not statistically significant). All of the other correlation coefficients have $p < 0.001$.**

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. SAT math | -- | -- | -- | -- | -- | -- | -- |
| 2. HS GPA | 0.32 | -- | -- | -- | -- | -- | -- |
| 3. Pre-belonging | 0.19 | $0.00^{ns}$ | -- | -- | -- | -- | -- |
| 4. Post-belonging | 0.32 | $0.09^{ns}$ | 0.60 | -- | -- | -- | -- |
| 5. Pre-FCI | 0.37 | $0.12^{**}$ | 0.36 | 0.48 | -- | -- | -- |
| 6. Post-FCI | 0.38 | 0.16 | 0.36 | 0.48 | 0.81 | -- | -- |
| 7. Grade | 0.50 | 0.38 | 0.28 | 0.50 | 0.54 | 0.59 | -- |

## Appendix E Percentages of Students Who Selected Each Choice for Each Survey Item II

To better understand how students' sense of belonging changed from the beginning to the end of the course, we calculated the percentages of female (Table 55) and male students (Table 56) who selected each answer choice from a 5-point Likert scale for each survey item for sense of belonging. Students were given a score from 1 to 5 respectively with a higher score indicating a greater level of sense of belonging.

As shown in Table 55 and Table 56, for both female and male students, from the beginning (pre) to the end (post) of the course, the percentages of students who selected 5 decreased, while the percentages of students who selected 1 or 2 increased. These results are consistent with the descriptive statistics shown in Table 21, which show that both female and male students' sense of belonging statistically significantly decreased from pre to post. In addition, by comparing Table 55 and Table 56, we found that the percentages of female students who selected 2 were larger than those of male students, while the percentages of female students who selected 5 were smaller than those of male students. These findings are also consistent with Table 21 showing that there were statistically significant gender differences disadvantaging women in both pre- and post-sense of belonging.

**Table 55 Percentages of female students who selected each choice from a 5-point Likert scale for each survey item of sense of belonging. All four items have the response scale: 1 = not at all true, 2 = a little true, 3 = somewhat true, 4 = mostly true and 5 = completely true.**

| **Survey items** | **Pre** | | | | | **Post** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| **Bel 1** | 1% | 10% | 26% | 38% | 24% | 8% | 15% | 30% | 31% | 16% |
| **Bel 2** | 1% | 5% | 17% | 37% | 41% | 4% | 8% | 20% | 37% | 31% |
| **Bel 3** | 4% | 17% | 33% | 36% | 10% | 10% | 22% | 30% | 32% | 7% |
| **Bel 4** | 4% | 11% | 21% | 30% | 33% | 9% | 15% | 22% | 26% | 28% |

**Table 56 Percentages of male students who selected each choice from a 5-point Likert scale for each survey item of sense of belonging. All four items have the response scale: 1 = not at all true, 2 = a little true, 3 = somewhat true, 4 = mostly true and 5 = completely true.**

| **Survey items** | **Pre** | | | | | **Post** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| **Bel 1** | 1% | 4% | 27% | 37% | 31% | 4% | 8% | 30% | 34% | 23% |
| **Bel 2** | 1% | 3% | 8% | 31% | 56% | 3% | 8% | 13% | 33% | 43% |
| **Bel 3** | 2% | 10% | 32% | 40% | 16% | 5% | 17% | 33% | 29% | 16% |
| **Bel 4** | 3% | 3% | 13% | 36% | 45% | 5% | 10% | 18% | 31% | 36% |

## Appendix F Full SEM Model

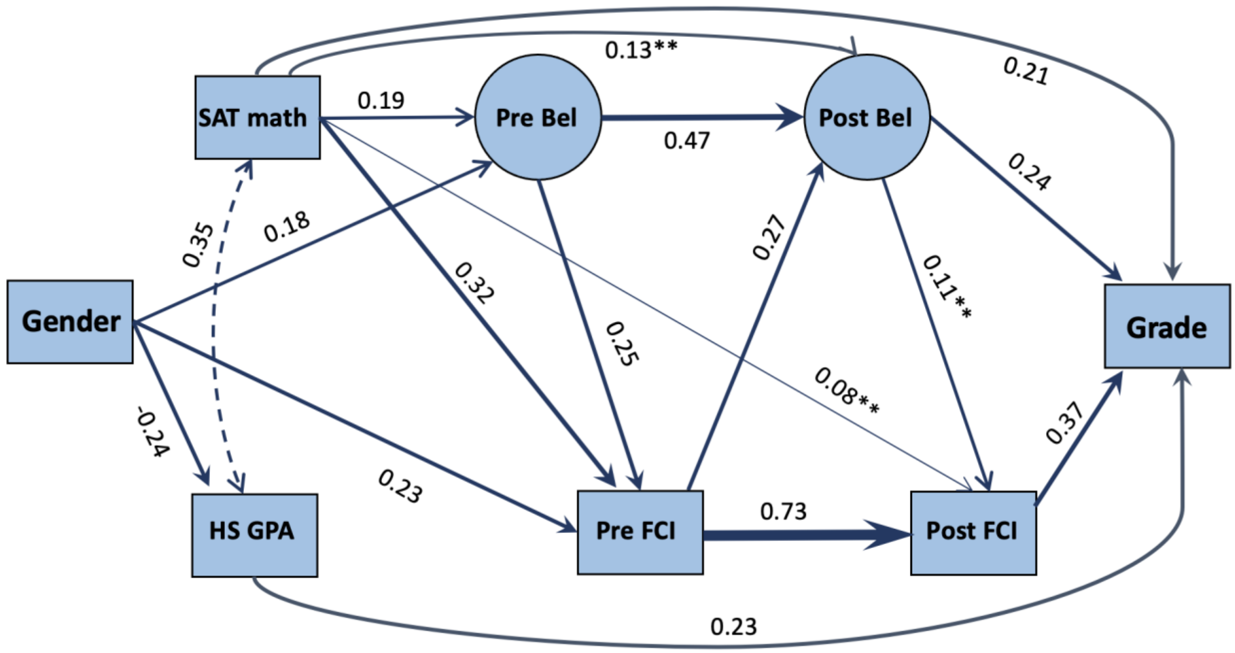



**Figure 30 Results of the path analysis part of the full SEM model. Bel represents sense of belonging. The solid lines represent regression paths and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. All the other regression lines show relations with $p < 0.001$.**

## Appendix G Moderation Analysis I

We did a moderation analysis to test whether gender moderates the relationship between any two constructs in the models (i.e., Does the strength of relationships given by the standardized regression coefficients between any two constructs in the models differ for women and men?). We used the R [110] software package "lavaan" to conduct multi-group SEM. We initially tested for measurement invariance. In other words, we looked at whether the factor loadings, intercepts, and residual variances of the observed variables are equal in the model for the latent constructs so we can confidently perform multi-group analysis. The analysis involved introducing certain constraints in steps and testing the model differences from the previous step. In each step, we compared the model to both the previous step and the freely estimated model, i.e., the model in which all parameters are freely estimated for each gender group. First, to test for "weak" or "metric" measurement invariance, we ran the model in which only factor loadings were fixed to equality across both gender groups, but intercept and errors were allowed to differ. The model was not statistically significantly different from the freely estimated model according to a likelihood ratio test, so weak measurement invariance holds (Chi-square difference ($\Delta\chi^2$) = 21.324, degree of freedom difference ($\Delta dof$) =21, and non-significant $p$ = 0.439). Next, we tested for "strong" or "scalar" measurement invariance by fixing both factor loadings and intercepts to equality across gender groups. This model was not statistically significantly different from either the metric invariance model ($\Delta\chi^2$ = 25.314, $\Delta dof$ = 21, $p$ = 0.234) or the freely estimated model ($\Delta\chi^2$= 46.637, $\Delta dof$ = 42, $p$ = 0.288), so strong measurement invariance holds. Finally, to test for "strict" measurement invariance we fixed factor loadings, intercepts, and residual variances to equality. In this step, there was a statistically significant difference from the previous models, therefore "strict invariance" did not hold when we compared to scalar measurement ($\Delta\chi^2$ = 64.732, $\Delta dof$ = 27, $p$ <0.001). However, strict invariance is unlikely to hold in most situations. Therefore, since strong measurement invariance holds for this model, we continued on to perform other group comparisons.

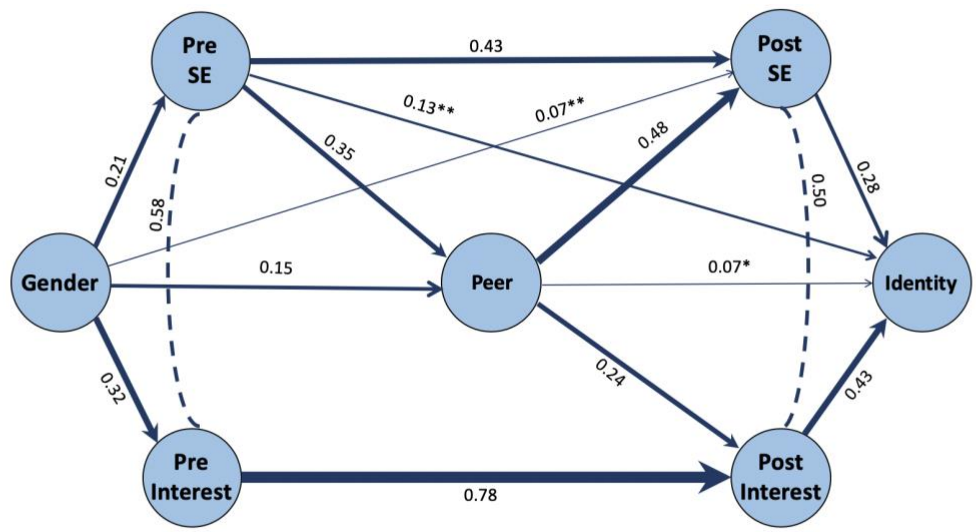
Next, we ran a multi-group SEM in which all regression estimates were fixed to equality for female and male students in addition to the factor loadings and intercepts, and we compared this model with freely estimated model. There was no statistically significant difference between the two models, so we reported the model where regression pathways are equal for men and women. The model fit parameters for this case were acceptable (RMSEA = 0.051, SRMR = 0.057, CFI = 0.931, TLI = 0.928). The multi-group SEM results suggest that regression pathways among the constructs did not show differences across gender when we compared to freely estimated model ($\Delta\chi^2$ = 77.059, $\Delta dof$ = 62, $p$ = 0.094) or to the scalar model ($\Delta\chi^2$ = 30.422, $\Delta dof$ = 20, $p$ = 0.063). However, the means of the latent variables showed the same gender differences that have been reported in the mediation models. That is, there were large gender differences in students' pre-self-efficacy and interest and slight differences in peer interaction, perceived recognition and sense of belonging.

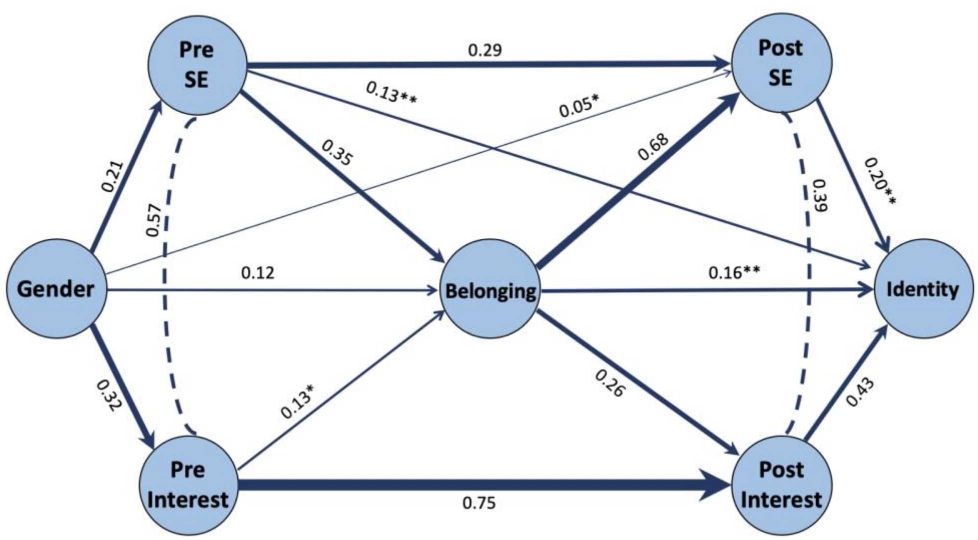
## Appendix H Paired-Sample t-test

**Table 57 Paired-sample *t*-test for matched pairs (matched pair refers to students who took both pre and post surveys) of pre- and post-self-efficacy and interest for female and male students.**

| Gender | Pre-Interest (1-4) | Post-Interest (1-4) | Statistics | | Pre-SE (1-4) | Post-SE (1-4) | Statistics | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Mean | p value | Cohen's d | Mean | Mean | p value | Cohen's d |
| Male (N=662) | 3.19 | 3.07 | <0.001 | 0.24 | 3.12 | 2.98 | <0.001 | 0.25 |
| Female (382) | 2.89 | 2.72 | <0.001 | 0.33 | 2.96 | 2.70 | <0.001 | 0.45 |

## Appendix I SEM Results of Other Models I



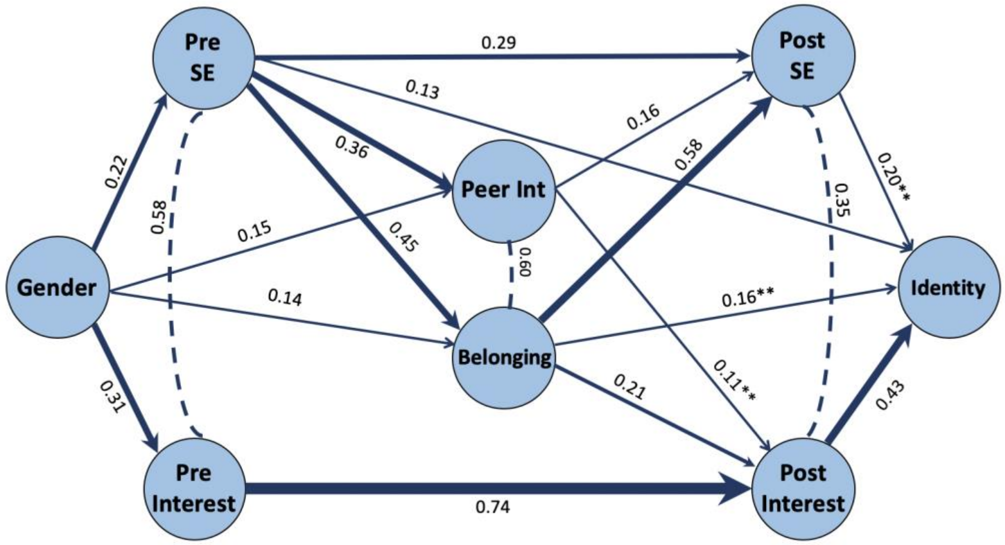



**Figure 31 Schematic diagram of the path analysis part of the structural equation modeling of perception of learning environment models including only peer interaction, only sense of belonging and both peer interaction and sense of belonging. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with 0.01< $p$ < 0.05 indicated by * and 0.001 < $p$ < 0.01 indicated by **. Other regression lines show relations with $p$ < 0.001.**

## Appendix J SEM Results of the Model Including Pre-Identity

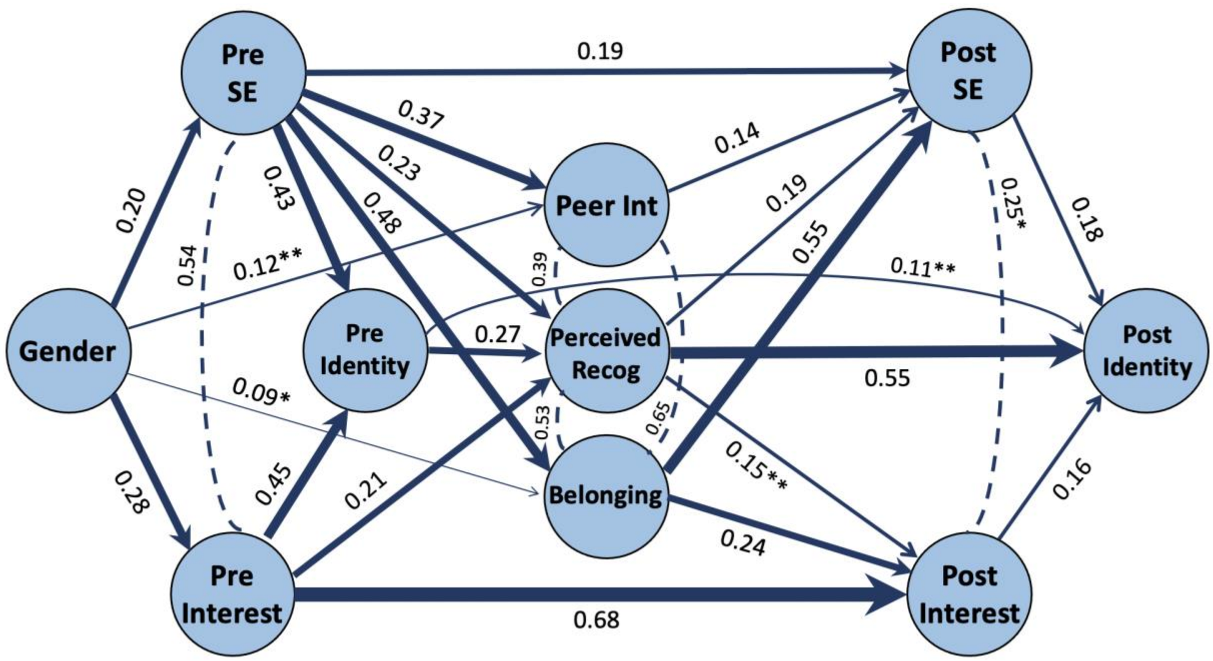



**Figure 32 Schematic diagram of the path analysis part of the SEM model including pre-identity. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with 0.01< $p$ < 0.05 indicated by * and 0.001 < $p$ < 0.01 indicated by **. Other regression lines show relations with $p$ < 0.001.**

## Appendix K Moderation Analysis II

We conducted a moderation analysis to test whether gender moderates the relationship between any two constructs in the models (i.e., do the strength of relationships given by the standardized regression coefficients between any two constructs in the models differ for women and men). We used the R [110] software package "lavaan" to conduct multi-group SEM. We initially tested for measurement invariance. In other words, we looked at whether the factor loadings, intercepts, and residual variances of the observed variables are equal in the model where we measured the latent constructs so we can confidently perform multi-group analysis. The analysis involved introducing certain constraints in steps and testing the model differences from the previous step. In each step, we compared the model to both the previous step and the freely estimated model, that is, the model where all parameters are freely estimated for each gender group. First, to test for "weak" or "metric" measurement invariance, we ran the model where only factor loadings were fixed to equality across both gender groups, but intercept and errors were allowed to differ. The model was not statistically significantly different from the freely estimated model according to a likelihood ratio test, so weak measurement invariance holds (Chi-square difference ($\Delta\chi^2$) = 25.001, degree of freedom difference ($\Delta dof$) =21, and non-significant $p$ = 0.2471). Next, we tested for "strong" or "scalar" measurement invariance by fixing both factor loadings and intercepts to equality across gender groups. This model was not statistically significantly different from either the metric invariance model ($\Delta\chi^2$ = 27.924, $\Delta dof$ = 21, $p$ = 0.1423) or the freely estimated model ($\Delta\chi^2$ = 52.925, $\Delta dof$ = 42, $p$ = 0.1203), so strong measurement invariance holds. Finally, to test for "strict" measurement invariance we fixed factor loadings, intercepts, and residual variances to equality. In this step, there was a statistically significant difference from the scalar measurement model ($\Delta\chi^2$ = 60.908, $\Delta dof$ = 27, $p$ <0.001), therefore “strict invariance” did not hold. However, strict invariance is unlikely to hold in most situations. Therefore, since strong measurement invariance holds for this model, we continued on to perform other group comparisons.

Next, we ran a multi-group SEM in which all regression estimates were fixed to equality for female and male students in addition to the factor loadings and intercepts, and we compared this model with the freely estimated model. There was no statistically significant difference between the two models, so we reported the model where regression pathways are equal for men and women. The model fit parameters for this case were acceptable (RMSEA = 0.056, SRMR = 0.058, CFI = 0.914, TLI = 0.910). The multi-group SEM results suggest that regression pathways among the constructs do not have differences across gender when we compared to the freely estimated model ($\Delta\chi^2$ = 93.438, $\Delta dof$ = 74, $p$ = 0.063) or to the scalar model ($\Delta\chi^2$ = 40.513, $\Delta dof$ = 32, $p$ = 0.1437).

## Appendix L Percentages of Students Who Selected each Choice for Each Survey Item III

In the main text, we discussed how students' motivational beliefs change from the beginning (pre) to the end (post) of the course by comparing their average scores on the pre- and post-motivational constructs. Here, we present the percentages of female and male students who selected each answer choice from a Likert scale for each survey item (Table 58-Table 61). The survey items for sense of belonging were scored on a 5-point Likert scale, while the survey items for all the other motivational constructs were scored on a 4-point Likert scale. For all survey items, higher scores indicate greater levels of motivational beliefs.

As shown in Table 58, for both female and male students, the percentages of students who selected 4 decreased from pre to post for all self-efficacy items, while the percentages of students who selected 1 or 2 mostly increased. Table 59 shows similar shifts in students' responses to the survey items under interest. These results are consistent with the descriptive statics shown in Table 38, which show that both male and female students' self-efficacy and interest statistically significantly decreased from pre to post.

In addition, by comparing percentages of female and male students who selected each answer choice, we found that for most survey items, 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 (for sense of belonging is 5) were smaller than those of male students. These findings are also consistent with Table 38 and Table 39 showing that there were statistically significant gender differences in all motivational constructs studied.

**Table 58 Percentages of female and male students who selected each choice from a 4-point Likert scale for each survey item of self-efficacy (SE) in the pre- and post-survey, which have the response scale: 1= NO!, 2 = no, 3 = yes, and 4 = YES!.**

| | Survey items | **Pre** | | | | **Post** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **Female** | SE1 | 7% | 29% | 54% | 10% | 9% | 30% | 55% | 6% |
| | SE2 | 1% | 11% | 75% | 12% | 4% | 16% | 71% | 8% |
| | SE3 | 1% | 4% | 64% | 31% | 5% | 28% | 54% | 13% |
| | SE4 | 1% | 10% | 70% | 19% | 5% | 24% | 61% | 10% |
| **Male** | SE1 | 3% | 25% | 60% | 12% | 4% | 22% | 63% | 11% |
| | SE2 | 1% | 8% | 71% | 21% | 1% | 10% | 70% | 19% |
| | SE3 | 1% | 3% | 55% | 41% | 2% | 14% | 56% | 28% |
| | SE4 | 0% | 7% | 69% | 24% | 2% | 16% | 66% | 16% |

**Table 59. Percentages of female and male students who selected each choice from a 4-point Likert scale for each survey item of interest in the pre- and post-survey. Interest1 has the response scale: 1 = Never, 2 = Once a month, 3 = Once a week, 4 = Every day". Interest2 has the response scale: 1 = Very boring, 2 = boring, 3 = interesting, 4 = Very interesting. The other two items have the response scale: 1= NO!, 2 = no, 3 = yes, and 4 = YES!.**

| | Survey items | **Pre** 1 | **Pre** 2 | **Pre** 3 | **Pre** 4 | **Post** 1 | **Post** 2 | **Post** 3 | **Post** 4 |
|---|---|---|---|---|---|---|---|---|---|
| **Female** | Interest1 | 8% | 36% | 41% | 15% | 8% | 20% | 47% | 25% |
| | Interest2 | 3% | 14% | 64% | 20% | 6% | 19% | 62% | 13% |
| | Interest3 | 2% | 26% | 55% | 17% | 6% | 41% | 42% | 10% |
| | Interest4 | 1% | 23% | 57% | 19% | 7% | 30% | 50% | 12% |
| **Male** | Interest1 | 4% | 22% | 43% | 31% | 3% | 12% | 41% | 44% |
| | Interest2 | 1% | 5% | 62% | 32% | 2% | 8% | 61% | 28% |
| | Interest3 | 1% | 12% | 55% | 32% | 3% | 22% | 49% | 25% |
| | Interest4 | 1% | 13% | 64% | 23% | 3% | 20% | 52% | 25% |

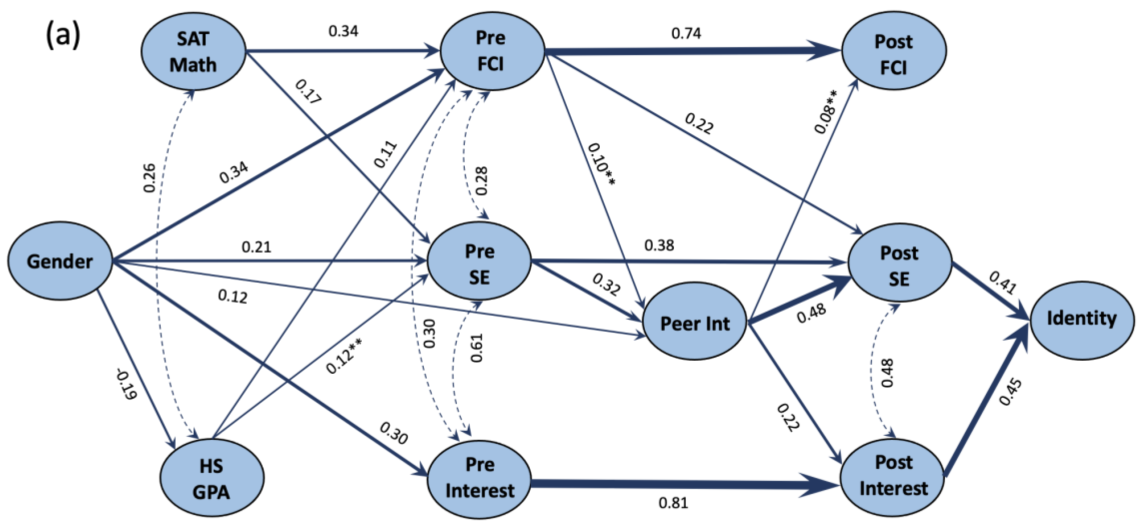
**Table 60 Percentages of female and male students who selected each choice from a 4-point Likert scale for each survey item of peer interaction, perceived recognition, and physics identity. All items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.**

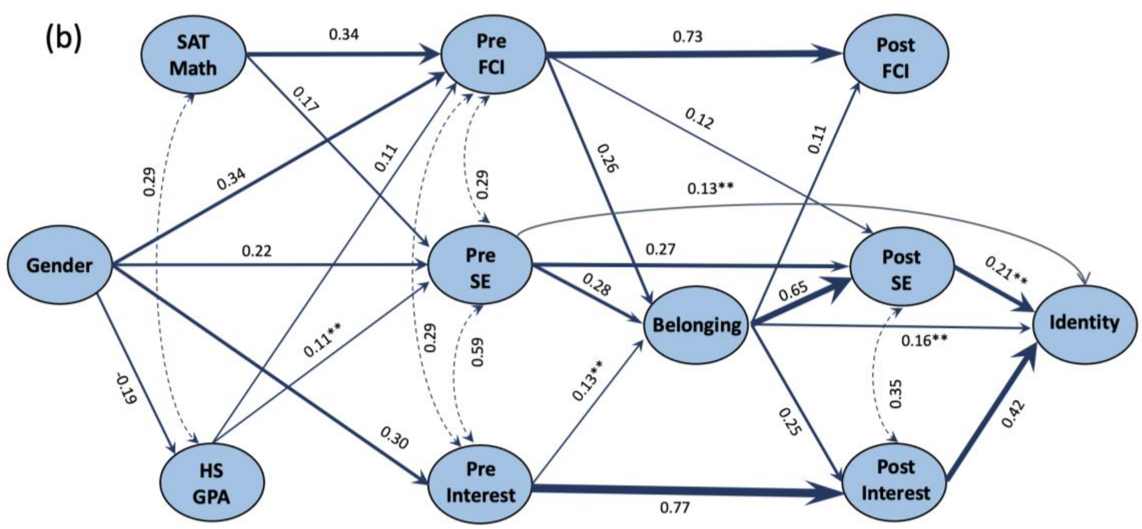
| Survey items | Female | | | | Male | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **Peer1** | 6% | 17% | 60% | 17% | 2% | 17% | 55% | 26% |
| **Peer2** | 8% | 26% | 57% | 10% | 3% | 17% | 58% | 22% |
| **Peer3** | 8% | 29% | 53% | 10% | 3% | 20% | 55% | 22% |
| **Peer4** | 8% | 33% | 50% | 8% | 4% | 22% | 56% | 19% |
| **Recognition1** | 18% | 38% | 35% | 8% | 9% | 30% | 44% | 17% |
| **Recognition2** | 16% | 40% | 36% | 8% | 9% | 32% | 44% | 15% |
| **Recognition3** | 21% | 48% | 28% | 2% | 12% | 42% | 40% | 6% |
| **Identity1** | 21% | 45% | 28% | 6% | 9% | 35% | 42% | 15% |

**Table 61 Percentages of female and male students who selected each choice from a 5-point Likert scale for each survey item of sense of belonging. All items have the response scale: 1 = not at all true, 2 = a little true, 3 = somewhat true, 4 = mostly true, and 5 = completely true.**

| **Survey items** | **Female** | | | | | **Male** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| **Belonging1** | 10% | 15% | 30% | 30% | 14% | 4% | 9% | 28% | 35% | 24% |
| **Belonging2** | 4% | 8% | 21% | 37% | 31% | 2% | 5% | 11% | 34% | 48% |
| **Belonging3** | 11% | 22% | 30% | 29% | 8% | 5% | 16% | 31% | 33% | 15% |
| **Belonging4** | 6% | 18% | 30% | 35% | 12% | 4% | 12% | 32% | 36% | 17% |
| **Belonging5** | 8% | 14% | 28% | 24% | 26% | 5% | 9% | 18% | 30% | 39% |

## Appendix M SEM Results of Other Models II



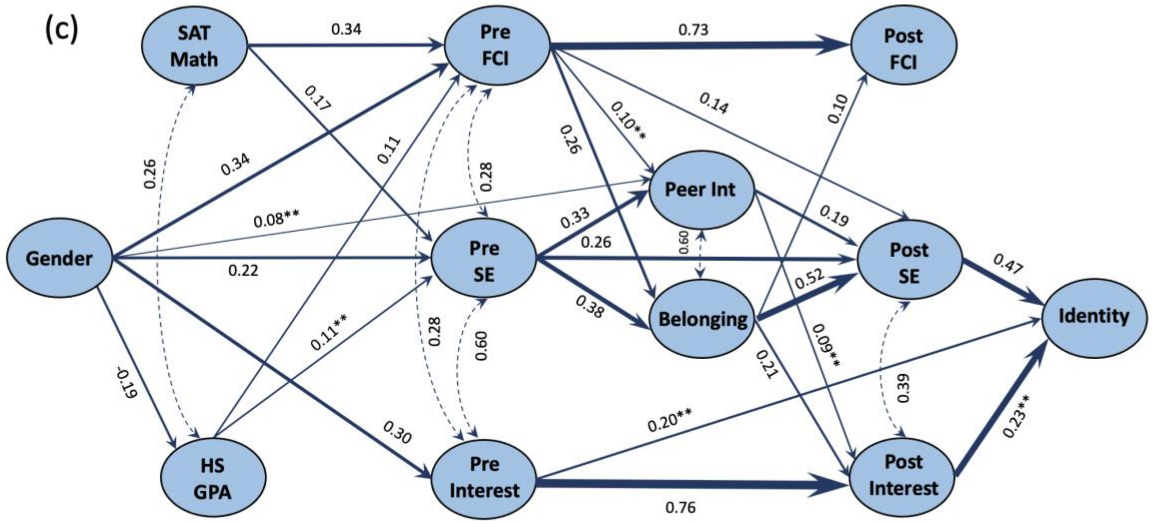



**Figure 33 Results of the path analysis part of the SEM models including (a) only peer interaction, (b) only sense of belonging, and (c) both peer interaction and sense of belonging. The solid lines represent regression paths, and the dashed lines represent residual covariances. The regression line thickness corresponds to the magnitude of $\beta$ value (standardized regression coefficient) with $0.001 \leq p < 0.01$ indicated by **. Other regression lines show relations with $p < 0.001$.**

## Appendix N Multi-Group SEM Analysis

We conducted a multi-group analysis to examine whether the survey items were interpreted in a conceptually similar manner by female and male students, and whether the strength of relationships given by the standardized regression coefficients between any two constructs in the models differ for women and men.

We first tested for measurement invariance. In other words, we looked at whether the factor loadings, intercepts, and residual variances of the items are equal across gender in the model. Since Model 1 and Model 2 include the same motivational constructs, measurement invariance tests are the same for these two models. To test measurement invariance, we ran a set of increasingly constrained models and tested the differences between these models. First, we examined the configural invariance model, in which the number of constructs and the correspondence between constructs and items are the same across gender groups, but all parameters can vary freely in each group. The result indicated that configural invariance holds (CFI = 0.975 > 0.90, TLI = 0.967 > 0.90, RMSEA = 0.051 < 0.08, SRMR = 0.038 < 0.08). Second, to test for "weak" measurement invariance, we ran the model in which the item loadings were constrained to be equal across gender groups, but intercepts and residual variances were allowed to vary between groups. According to a likelihood ratio test, there was no statistically significant difference between the weak invariance model and the configural invariance model, so the weak measurement invariance holds (Chi-square difference $\Delta\chi^2$ = 4.936, degree of freedom difference $\Delta dof$ = 8, $p$ = 0.764). The third step is testing for "strong" measurement invariance. We ran the model in which both the item loadings and intercepts were constrained to be equal across gender groups, but the residual variances were allowed to differ. A likelihood ratio test shows that there was no statistically significant difference between the strong invariance model and the weak invariance model ($\Delta\chi^2$ = 7.935, $\Delta dof$ = 8, $p$ = 0.440) or the configural invariance model ($\Delta\chi^2$ = 12.872, $\Delta dof$ = 16, $p$ = 0.682), so strong measurement invariance holds. Finally, to test for “strict” measurement invariance, we ran the model in which the item loadings, intercepts, and residual variances were constrained to be equal across gender groups. This model was statistically significantly different from the strong invariance model ($\Delta\chi^2$ = 18.378, $\Delta dof$ = 9, $p$ =0.031), therefore “strict invariance” did not hold. However, strict invariance is unlikely to hold in most situations. Therefore, since strong measurement invariance holds for this model, we proceeded on to test for structural invariance.

We tested for structural invariance to examine whether the regression coefficients among the motivational constructs are equal across gender. Since the regression relationships among the constructs are different in Model 1 and Model 2, we conducted the structural invariance test for Model 1 and Model 2 separately. We first ran a multi-group SEM for Model 1, in which all regression coefficients were constrained to be equal across gender groups in addition to the item loadings and intercepts. The model fit parameters for this model indicate a good fit (CFI = 0.975, TLI = 0.972, RMSEA = 0.047, SRMR = 0.050). According to the results of likelihood ratio tests, this model was not statistically significantly different from either the configural invariance model ($\Delta\chi^2$ = 26.599, $\Delta dof$ = 24, $p$ = 0.324) or the strong invariance model ($\Delta\chi^2$ = 13.728, $\Delta dof$ = 8, $p$ = 0.089). Thus, the regression pathways among the constructs do not have statistically significant differences across gender for Model 1. Then, we ran a multi-group SEM for Model 2, in which all regression coefficients were constrained to be equal across gender groups in addition to the item loadings and intercepts. This model also fits the data very well (CFI = 0.975, TLI = 0.972, RMSEA = 0.047, SRMR = 0.048). Similarly, there was no statistically significant difference between this model and the configural invariance model ($\Delta\chi^2$ = 25.754, $\Delta dof$ = 24, $p$ = 0.366) or the strong invariance model ($\Delta\chi^2$ = 12.882, $\Delta dof$ = 8, $p$ = 0.116). Thus, structural invariance also holds for Model 2.

## Appendix O Percentages of Students Who Selected Each Choice for Each Survey Item IV

In the main text, we investigated how students' motivational beliefs change from the beginning to the end of the course by comparing their average scores on the motivational constructs in the pre- and post-survey. Here, we present the percentages of female (Table 62) and male students (Table 63) 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, physics identity, and engineering identity.

As shown in Table 62 and Table 63, for both female and male students, the percentages of students who selected 3 or 4 for most survey items under self-efficacy and physics identity decreased from pre to post, while the percentages of students who selected 1 or 2 increased. These results are consistent with the descriptive statics shown in Table 45, which show that both male and female students' self-efficacy and physics identity statistically significantly decreased from the beginning to the end of the course. In addition, by comparing Table 62 and Table 63, we found that for most survey items, 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 45 showing that there were statistically significant gender differences in all motivational constructs studied.

**Table 62 Percentages of female students who selected each choice from a 4-point Likert scale for each survey item in the pre- and post-survey. The self-efficacy (SE) and interest (Int) items have the response scale: 1= NO!, 2 = no, 3 = yes, and 4 = YES!, while the perceived recognition (Recog), physics identity, and engineering identity items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.**

| Survey items | Pre | | | | Post | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **SE1** | 4% | 26% | 65% | 5% | 9% | 25% | 60% | 6% |
| **SE2** | 2% | 11% | 78% | 9% | 4% | 16% | 74% | 6% |
| **SE3** | 0% | 5% | 72% | 23% | 6% | 20% | 61% | 14% |
| **SE4** | 0% | 11% | 72% | 18% | 5% | 19% | 65% | 11% |
| | | | | | | | | |
| **Int1** | 6% | 38% | 44% | 12% | 8% | 22% | 46% | 24% |
| **Int2** | 1% | 9% | 70% | 19% | 4% | 22% | 63% | 11% |
| **Int3** | 0% | 25% | 62% | 13% | 5% | 48% | 39% | 9% |
| **Int4** | 1% | 27% | 54% | 18% | 6% | 29% | 54% | 11% |
| | | | | | | | | |
| **Recog1** | 11% | 39% | 42% | 7% | 13% | 38% | 41% | 8% |
| **Recog2** | 12% | 36% | 45% | 6% | 12% | 39% | 40% | 9% |
| **Recog3** | 7% | 50% | 42% | 1% | 15% | 54% | 29% | 2% |
| | | | | | | | | |
| **Physics identity** | 9% | 48% | 38% | 6% | 16% | 49% | 30% | 5% |
| | | | | | | | | |
| **Engineering identity** | 0% | 4% | 46% | 49% | 1% | 5% | 52% | 41% |

**Table 63 Percentages of male students who selected each choice from a 4-point Likert scale for each survey item in the pre- and post-survey. The self-efficacy (SE) and interest (Int) items have the response scale: 1= NO!, 2 = no, 3 = yes, and 4 = YES!, while the perceived recognition (Recog), physics identity, and engineering identity items have the response scale: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree.**

| Survey items | Pre | | | | Post | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **SE1** | 2% | 27% | 60% | 10% | 3% | 20% | 64% | 13% |
| **SE2** | 0% | 8% | 72% | 20% | 0% | 11% | 68% | 20% |
| **SE3** | 0% | 4% | 60% | 36% | 3% | 14% | 55% | 27% |
| **SE4** | 0% | 8% | 69% | 23% | 0% | 18% | 65% | 17% |
| **Int1** | 4% | 24% | 42% | 29% | 4% | 14% | 38% | 43% |
| **Int2** | 2% | 6% | 66% | 25% | 3% | 12% | 60% | 25% |
| **Int3** | 2% | 15% | 61% | 21% | 3% | 28% | 47% | 21% |
| **Int4** | 1% | 15% | 62% | 22% | 4% | 25% | 49% | 21% |
| **Recog1** | 4% | 31% | 49% | 16% | 9% | 26% | 49% | 15% |
| **Recog2** | 6% | 33% | 43% | 18% | 10% | 33% | 42% | 15% |
| **Recog3** | 5% | 35% | 54% | 5% | 13% | 35% | 46% | 6% |
| **Physics identity** | 3% | 30% | 52% | 14% | 11% | 32% | 45% | 12% |
| **Engineering identity** | 1% | 1% | 32% | 66% | 1% | 4% | 39% | 56% |

## Appendix P Quantum Measurement MQS

The multiple-choice questions in the sequence and notes to the instructors are reproduced below. The correct answers are in boldface.

Notes to Instructor

- In general, an operator acting on a state can be represented in Dirac notation as $\hat{Q}|\Psi\rangle$, where $\hat{Q}$ stands for a quantum operator, and $|\Psi\rangle$ stands for a quantum state.
- In this Multiple-choice Question Sequence (MQS), we write $\hat{Q}|\Psi\rangle$ in position representation as $\langle x|\hat{Q}|\Psi\rangle = \hat{Q}\left(x, -i\hbar\frac{\partial}{\partial x}\right)\Psi(\mathrm{x}) = \hat{Q}\Psi(\mathrm{x})$, where $\hat{Q}\left(x, -i\hbar\frac{\partial}{\partial x}\right)$ (shorten as $\hat{Q}$) is the operator $\hat{Q}$ in position representation, and $\Psi(\mathrm{x})$ is the wavefunction in position representation corresponding to state $|\Psi\rangle$.
- In order to go from the Dirac notation to position representation, one should take the scalar product with $\langle x|$

$$\hat{Q}|\Psi\rangle \longrightarrow \langle x|\hat{Q}|\Psi\rangle \overset{\text{def}}{=} \hat{Q}\left(x, -i\hbar\frac{\partial}{\partial x}\right)\Psi(x)$$

$$|\Psi\rangle \longrightarrow \langle x|\Psi\rangle \overset{\text{def}}{=} \Psi(x)$$

- In this Multiple-choice Question Sequence below, none of the operators explicitly depend on time.

**(MQS 1.1)**

Choose the following that is correct regarding the Hamiltonian operator $\hat{H}$ acting on a generic wavefunction $\Psi(x)$ which is not an eigenstate of $\hat{H}$.

A. $\hat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E}\Psi(\mathrm{x})$

B. $\hat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E_n}\psi_\mathrm{n}(\mathrm{x})$

C. $\hat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E_n}$

D. $\hat{\mathrm{H}}\Psi(\mathrm{x}) = \psi_\mathrm{n}(\mathrm{x})$

***E. None of the above***

**Class discussion for MQS 1.1**

- Some students may incorrectly think that an operator acting on a state corresponds to a measurement of the corresponding observable and that this process of measuring the observable is given, e.g., by equations such as $\hat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E_n}\psi_\mathrm{n}(\mathrm{x})$ for measurement of energy.
- Some students may incorrectly think that whenever an operator $\hat{\mathrm{Q}}$ corresponding to a physical observable Q acts on any generic state $\Psi(\mathrm{x})$, it will yield a corresponding eigenvalue and the same state back, i.e., $\hat{\mathrm{Q}}\,\Psi(\mathrm{x}) = \lambda\,\Psi(\mathrm{x})$ (e.g., $\hat{\mathrm{H}}\Psi(\mathrm{x}) = \mathrm{E}\Psi(\mathrm{x})$)
- But only when $\Psi_q(\mathrm{x})$ is an eigenstate of $\hat{\mathrm{Q}}$, we obtain $\hat{\mathrm{Q}}\,\Psi_q(\mathrm{x}) = q\Psi_q(\mathrm{x})$, where $q$ is the corresponding eigenvalue.
- A generic wave function $\Psi(\mathrm{x})$ can be expanded in terms of a complete set of eigenstates of any Hermitian operator corresponding to an observable, e.g., if $\hat{\mathrm{H}}\psi_\mathrm{n}(\mathrm{x}) = \mathrm{E_n}\psi_\mathrm{n}(\mathrm{x})$
    - $\Psi(\mathrm{x}) = \sum_\mathrm{n} \mathrm{C_n}\,\psi_\mathrm{n}(\mathrm{x})$, where $\mathrm{C_n} = \langle\psi_\mathrm{n}|\Psi\rangle$

Thus, $\hat{H}$ acting on a generic state $\Psi(x)$ can be represented by $\hat{H}\Psi(x) = \sum_n \hat{H}C_n\,\psi_n(x) = \sum_n E_n C_n\,\psi_n(x)$

- If the instructor has covered Dirac notation, they can also discuss:
  Because $\Psi(x) = \langle x|\Psi\rangle, |\Psi\rangle = \sum_n C_n|\psi_n\rangle$ , and $\hat{H}|\psi_n\rangle = E_n|\psi_n\rangle$

**(MQS 1.2)**

Choose the statement that is correct regarding the position operator $\hat{x}$ acting on a generic wavefunction $\Psi(x)$.

A. $\hat{x}\Psi(x) = x'$
B. $\hat{x}\Psi(x) = x'\delta(x - x')$
***C.*** $\boldsymbol{\hat{x}\Psi(x) = x\Psi(x)}$
D. $\hat{x}\Psi(x) = \delta(x - x')$
E. None of the above

**Class discussion for MQS 1.2**

This question is similar in spirit to the previous one about the Hamiltonian operator.
In the position representation, $\hat{x}\Psi(x) = x\Psi(x)$. This is because in position representation, $\hat{x} = x$. This does not mean that $\Psi(x)$ is an eigenstate of $\hat{x}$.

- If you have covered Dirac notation, you can also discuss:

Because $\Psi(x) = \langle x|\Psi\rangle, |\Psi\rangle = \int_{-\infty}^{\infty} C(x')\,|x'\rangle dx'$, $\hat{x}\,|x'\rangle = x'|x'\rangle$ and $\langle x'|\hat{x} = x'\langle x'|$

- $\hat{x}\delta(x - x') = \hat{x}\langle x|x'\rangle = \langle x|\hat{x}|x'\rangle = x'\langle x|x'\rangle = x'\delta(x - x')$ is an eigenvalue equation, in which position operator $\hat{x}$ is acting on a position eigenstate $\delta(x - x')$ (in the position representation) with eigenvalue $x'$. Here, $x'$ is a number, which is the eigenvalue corresponding to the position eigenstate $\delta(x - x')$.
- $\hat{x}\Psi(x) = \hat{x}\langle x|\Psi\rangle = \int_{-\infty}^{\infty} \hat{x}\langle x|x'\rangle\langle x'|\Psi\rangle\, dx' = \int_{-\infty}^{\infty} \hat{x}\delta(x - x')\Psi(x')\, dx'$

  $= \int_{-\infty}^{\infty} x\delta(x - x')\Psi(x')\, dx' = x\Psi(x)$.

  Please note that $\hat{x}\Psi(x) = x\Psi(x)$ is not an eigenvalue equation. In position representation, position operator $\hat{x}$ acting on any generic wave function $\Psi(x)$ or $\langle x|\Psi\rangle$ simply corresponds to multiplication by x.
- Or you can start with a generic expression of the position operator $\hat{x}$ acting on a state $|\Psi\rangle$, which is $\hat{x}|\Psi\rangle$. By multiplying it the bra $\langle x|$, we can write this expression in the position representation $\langle x|\hat{x}|\Psi\rangle$. Since $\langle x|\hat{x}|\Psi\rangle = x\langle x|\Psi\rangle = x\Psi(x)$, position operator $\hat{x}$ acting on any generic wave function $\Psi(x)$ or $\langle x|\Psi\rangle$ simply corresponds to multiplication by x, i.e., $\hat{x}\Psi(x) = x\Psi(x)$.

**Checkpoints**

- An operator (corresponding to an observable) acting on a state does NOT correspond to the measurement of the corresponding observable.
- What is the result of $\hat{x}$ acting on a position eigenstate?
- What is the result of $\hat{x}$ acting on a generic state?
- What is the result of $\hat{H}$ acting on an energy eigenstate?
- What is the result of $\hat{H}$ acting on a generic state?

**(MQS 2.1)**

Suppose at time $t = 0$, the initial wavefunction of a particle in a 1D infinite square well of width a (0<x<a) is $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$, where $\Psi_1(x)$ and $\Psi_2(x)$ are the ground state and first excited state wavefunctions. Choose all of the following statements that are correct for measurements on the system in this state at $t$ = 0.

1. A measurement of the energy can yield any energy $E_n$, where n=1,2,3…∞.
2. A measurement of the energy will yield ($E_1$ + $E_2$)/2.
3. A measurement of the position in a narrow range *dx* can yield many different values in this well (0<x<a).

***A.*** 1 only B. 2 only ***C. 3 only***
D. 2 and 3 only E. None of the above

**(MQS 2.2)**

Q is a generic observable (with corresponding Hermitian operator $\hat{\mathrm{Q}}$ which has eigenstates $\varphi_{\mathrm{q}}(\mathrm{x})$ and continuous eigenvalues q and eigenvalue equation $\hat{\mathrm{Q}}\varphi_{\mathrm{q}}(\mathrm{x}) = \mathrm{q}\varphi_{\mathrm{q}}(\mathrm{x})$). Choose all of the following statements that are correct about a measurement of the observable $Q$ on a generic state $\Psi(x)$ (which is not an eigenstate of the operator $\hat{Q}$).

1. The measurement of the observable $Q$ will collapse the wavefunction into one of the eigenstates $\varphi_{\mathrm{q}}(x)$ of operator $\hat{Q}$.
2. A measurement of an observable $Q$ must return one of the eigenvalues $q$ of the operator $\hat{Q}$.
3. The operator $\hat{Q}$ acting on state $\Psi(x)$ is equivalent to the measurement of the observable $Q$ . The measurement process is given by $\hat{Q}\Psi(x) = q\varphi_{\mathrm{q}}(x)$.

A) 1 only ***B) 1 and 2 only*** C) 1 and 3 only
D) 2 and 3 only E) All of the above

**(MQS 2.3)**

At time $t = 0$, the initial wavefunction of a particle is $\Psi(\mathrm{x}, 0) = (\frac{1}{5} - \frac{4}{5}i)\Psi_1(\mathrm{x}) + \frac{\sqrt{8}}{5}\Psi_2(\mathrm{x})$, where $\Psi_1(\mathrm{x})$ and $\Psi_2(\mathrm{x})$ are the ground state and first excited state wavefunctions. Choose all of the following statements that are correct for measurements on the system in this state at $t$ = 0.

1. If energy is measured, the probability of obtaining $\mathrm{E}_1$ is $\frac{1}{25}$ and $\mathrm{E}_2$ is $\frac{8}{25}$.
2. If position is measured, the probability density for measuring $\mathrm{x}_0$ is $(\frac{17}{25}|\Psi_1(\mathrm{x}_0)|^2 + \frac{8}{25}|\Psi_2(\mathrm{x}_0)|^2)$
3. If a generic observable D (with corresponding Hermitian operator $\hat{D}$ which has eigenstates $|\phi_{\mathrm{i}}\rangle$ and discrete eigenvalues $\mathrm{d}_i$ and eigenvalue equation $\hat{D}|\phi_{\mathrm{i}}\rangle = \mathrm{d}_i|\phi_{\mathrm{i}}\rangle$, where $i$ =1,2,3…) is measured, the probability of obtaining $\mathrm{d}_i$ is $|\langle\phi_{\mathrm{i}}|\Psi\rangle|^2$, where $|\Psi\rangle$ represents the quantum state corresponding to $\Psi(\mathrm{x})$.
4. If a generic observable Q (with corresponding Hermitian operator $\hat{Q}$ which has eigenstates $|q\rangle$ and continuous eigenvalues $q$ and eigenvalue equation $\hat{Q}|q\rangle = q|q\rangle$) is measured, the probability density for measuring $q$ is $|\langle q|\Psi\rangle|^2$, where $|\Psi\rangle$ represents the quantum state corresponding to $\Psi(\mathrm{x})$.

***A.*** 1 only B. 2 only ***C. 3 and 4 only***
D. 2 and 3 only E. None of the above

**Class discussion for MQS 2.3**

If you haven't discussed Dirac notation with the students, please feel free to replace the choice 3 and choice 4 with:

3. If a generic observable D (with corresponding Hermitian operator $\hat{D}$ which has eigenstates $\phi_i(x)$ and discrete eigenvalues $d_i$ and eigenvalue equation $\hat{D}\phi_i(x) = d_i\phi_i(x)$, where $i$ =1,2,3,4…) is measured, the probability of obtaining $d_i$ is $\left|\int_{-\infty}^{\infty} \phi_i^*(x)\Psi(x)dx\right|^2$.
4. If a generic observable Q (with corresponding Hermitian operator $\hat{Q}$ which has eigenstates $\varphi_q(x)$ and continuous eigenvalues $q$ and eigenvalue equation $\hat{Q}\varphi_q(x) = q\varphi_q(x)$) is measured, the probability density for measuring $q$ is $\left|\int_{-\infty}^{\infty} \varphi_q^*(x)\Psi(x)dx\right|^2$.

**Note:**

- In $\hat{D}|\phi_i\rangle = d_i|\phi_i\rangle$, operator $\hat{D}$ is an operator corresponding to observable D, and $|\phi_i\rangle$ stands for an eigenstate of $\hat{D}$ with eigenvalue $d_i$
- $\hat{D}\phi_i(x) = d_i\phi_i(x)$ is the same eigenvalue equation written in position representation, where $\hat{D}$ is $\hat{D}\left(x, -i\hbar\frac{\partial}{\partial x}\right)$, and $\phi_i(x)$ is an eigenfunction of $\hat{D}$ with eigenvalue $d_i$ in the position representation.

**Class discussion for MQS 2.1-2.3**

Consider a generic wavefunction $\Psi(x)$ corresponding to a generic state $|\Psi\rangle$.

- If you measure energy, the probability of obtaining $E_m$ is $P(E_m) = |\langle\Psi_m|\Psi\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle\Psi_m|x\rangle\langle x|\Psi\rangle dx\right|^2 = \left|\int_{-\infty}^{\infty} \Psi_m^*(x)\Psi(x)dx\right|^2$ (where $|\Psi_m\rangle$ is an energy eigenstate satisfying $\hat{H}|\Psi_m\rangle = E_m|\Psi_m\rangle$)
- If you measure position, the probability density for measuring $x_0$ is $\rho(x_0) = |\langle x_0|\Psi\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle x_0|x\rangle\langle x|\Psi\rangle dx\right|^2 = \left|\int_{-\infty}^{\infty} \delta(x_0 - x)\Psi(x)dx\right|^2 = |\Psi(x_0)|^2$ (where $|x_0\rangle$ is a position eigenstate satisfying $\hat{x}|x_0\rangle = x_0|x_0\rangle$)
- If you measure the observable D whose corresponding Hermitian operator $\hat{D}$ has a discrete eigenvalue spectrum, the probability of obtaining $d_i$ is $P(d_i) = |\langle\phi_i|\Psi\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle\phi_i|x\rangle\langle x|\Psi\rangle dx\right|^2 = \left|\int_{-\infty}^{\infty} \phi_i^*(x)\Psi(x)dx\right|^2$ (where $|\phi_i\rangle$ is an eigenstate of $\hat{D}$ satifying $\hat{D}|\phi_i\rangle = d_i|\phi_i\rangle$)
- If you measure the observable Q whose corresponding operator $\hat{Q}$ has a continuous eigenvalue spectrum, the probability density for measuring $q$ is $\rho(q) = |\langle q|\Psi\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle q|x\rangle\langle x|\Psi\rangle dx\right|^2 = \left|\int_{-\infty}^{\infty} \varphi_q^*(x)\Psi(x)dx\right|^2$ (where $|q\rangle$ is an eigenstate of $\hat{Q}$ satifying $\hat{Q}|q\rangle = q|q\rangle$)

**Note:**

1. Emphasize to students that when considering the probability of measuring an observable in a generic state, they should be thinking about the measurement basis, which is the basis consisting of the eigenstates of the operator corresponding to the observable being measured. Then projecting the generic state onto an eigenstate of the operator corresponding to the

observable (in the measurement basis), the absolute square of the projection will give the probability (discrete eigenvalues) or probability density (continuous eigenvalues) of obtaining the corresponding eigenvalue.

2. The projections are given by the inner product of the generic state $\Psi(x)$ and the eigenstates of the operator corresponding to the observable. For example, for energy measurement, the projection of state $\Psi(x)$ onto energy eigenstate $\Psi_m(x)$ is $\int_{-\infty}^{\infty} \Psi_m^*(x)\Psi(x)dx$. If you have discussed Dirac notation, this projection can be simply represented by $\langle\Psi_m|\Psi\rangle$.
3. We can calculate the probability/probability density by projecting the generic state onto the eigenstates as discussed above. However, if we can expand the generic state in terms of a complete set of eigenstates of the operator corresponding to the observable we measure, the probability/probability density of measuring a particular value of the observable is the absolute square of the corresponding expansion coefficient.

   - For example, we can expand a generic state $\Psi(x)$ in terms of energy eigenstates, $\Psi(x) = \sum_n C_n \Psi_n(x)$, then $P(E_m) = |\langle\Psi_m|\Psi\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle\Psi_m|x\rangle\langle x|\Psi\rangle dx\right|^2 = \left|\int_{-\infty}^{\infty} \Psi_m^*(x)\Psi(x)dx\right|^2 = \left|\sum_n C_n \int_{-\infty}^{\infty} \Psi_m^*(x)\Psi_n(x)dx\right|^2 = |\sum_n C_n\, \delta_{mn}|^2 = |C_m|^2$
   - If you have done Dirac notation with students, you can also calculate the probability in the following way:
     $P(E_m) = |\langle\Psi_m|\Psi\rangle|^2 = |\sum_n C_n\langle\Psi_m|\Psi_n\rangle|^2 = |\sum_n C_n\, \delta_{mn}|^2 = |C_m|^2$
4. The probability density for measuring position $x_0$ is $\rho(x_0) = |\langle x_0|\Psi\rangle|^2 = |\Psi(x_0)|^2$

**Checkpoints**

- When a measurement is made, what measured values can you obtain if the state of the system is not an eigenstate of the operator corresponding to an observable (or when it IS an eigenstate of the operator corresponding to the observable)?
- When a measurement is made, what happens to the wave function instantaneously after the measurement if the state is not an eigenstate of the operator corresponding to the observable measured?
- Does a Hermitian operator corresponding to an observable acting on the quantum state correspond to a measurement of the observable? No!
- How does one calculate probability (discrete eigenvalue spectrum) or probability density (continuous eigenvalue spectrum) of measuring an observable?

**(MQS 3.1)**

Choose all of the following statements that are correct.

1. The stationary states refer to the eigenstates of any operator corresponding to a physical observable.
2. Any wavefunction for a system can be expressed as a linear superposition of the energy eigenstates.
3. If a system is in an eigenstate of any operator that corresponds to a physical observable, it stays in that state unless an external perturbation is applied.

A. 1 only ***B. 2 only*** C. 3 only
D. 2 and 3 only E. None of the above

**(MQS 3.2)**

Suppose at time $t = 0$, a particle in a 1D infinite square well has the initial wavefunction $\Psi(x,0) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$, where $\Psi_1(x)$ and $\Psi_2(x)$ are the ground state and first excited state wavefunctions. Choose all of the following expressions that can correctly represent the state $\Psi(x,t)$ of the particle after time *t*

1. $\Psi(x,t) = \frac{1}{\sqrt{2}} e^{\frac{-i(E_1+E_2)t}{2\hbar}} (\Psi_1(x) + \Psi_2(x))$
2. $\Psi(x,t) = \frac{1}{\sqrt{2}} (e^{\frac{-iE_1 t}{\hbar}} \Psi_1(x) + e^{\frac{-iE_2 t}{\hbar}} \Psi_2(x))$
3. $\Psi(x,t) = \frac{1}{\sqrt{2}} e^{\frac{-i\hat{H}t}{\hbar}} (\Psi_1(x) + \Psi_2(x))$

A. 1 only B. 2 only C. 1 and 3 only
***D. 2 and 3 only*** E. None of the above

**Class discussion for MQS 3.2**

- By solving the Time-dependent Schrödinger equation $i\hbar\frac{\partial}{\partial t}\Psi(\mathrm{x},t) = \hat{\mathrm{H}}\,\Psi(\mathrm{x},t)$, the time dependance of a generic state is given by $\Psi(\mathrm{x},t) = e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\Psi(\mathrm{x},0)$, where $e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}$ is the time-evolution operator.
- Any wavefunction for a system can be expressed as a linear superposition of the energy eigenstates:

$\Psi(\mathrm{x},0) = \sum_{\mathrm{n}} \mathrm{C_n}\,\psi_{\mathrm{n}}(\mathrm{x})$, where $\psi_{\mathrm{n}}(x)$ is an energy eigenstate with eigenvalue $\mathrm{E_n}$. Thus,

$$\Psi(\mathrm{x},t) = e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\Psi(\mathrm{x},0) = \sum_{\mathrm{n}} \mathrm{C_n} e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x})$$

Let's look at one term in this expansion $-\ \mathrm{C_n} e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x})$. We can expand the exponential function $e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}$ as follow: $e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}} = 1 + \frac{-i\hat{H}t}{\hbar} + \frac{1}{2}\left(\frac{-i\hat{H}t}{\hbar}\right)^2 + \frac{1}{6}\left(\frac{-i\hat{H}t}{\hbar}\right)^3 + \cdots$

Thus, $e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\psi_{\mathrm{n}}(x) = \psi_{\mathrm{n}}(x) + \frac{-i\hat{H}t}{\hbar}\psi_{\mathrm{n}}(x) + \frac{1}{2}\left(\frac{-i\hat{H}t}{\hbar}\right)^2\psi_{\mathrm{n}}(x) + \frac{1}{6}\left(\frac{-i\hat{H}t}{\hbar}\right)^3\psi_{\mathrm{n}}(x) + \cdots$

Because $\hat{\mathrm{H}}\psi_{\mathrm{n}}(x)$=$\mathrm{E_n}\psi_{\mathrm{n}}(\mathrm{x})$, we have $\frac{-i\hat{\mathrm{H}}t}{\hbar}\psi_{\mathrm{n}}(x) = \frac{-i\mathrm{E_n}t}{\hbar}\psi_{\mathrm{n}}(x)$, $\left(\frac{-i\hat{\mathrm{H}}t}{\hbar}\right)^2\psi_{\mathrm{n}}(x) = \left(\frac{-i\mathrm{E_n}t}{\hbar}\right)^2\psi_{\mathrm{n}}(x)$, $\left(\frac{-i\hat{\mathrm{H}}t}{\hbar}\right)^3\psi_{\mathrm{n}}(x) = \left(\frac{-i\mathrm{E_n}t}{\hbar}\right)^3\psi_{\mathrm{n}}(x)$, ...

Therefore, $e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\psi_{\mathrm{n}}(x) = \psi_{\mathrm{n}}(x) + \frac{-i\mathrm{E_n}t}{\hbar}\psi_{\mathrm{n}}(x) + \frac{1}{2}\left(\frac{-i\mathrm{E_n}t}{\hbar}\right)^2\psi_{\mathrm{n}}(x) + \frac{1}{6}\left(\frac{-i\mathrm{E_n}t}{\hbar}\right)^3\psi_{\mathrm{n}}(x) + \cdots = e^{\frac{-i\mathrm{E_n}t}{\hbar}}\psi_{\mathrm{n}}(x)$

Thus, $\Psi(\mathrm{x},t) = e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\Psi(\mathrm{x},0) = \sum_{\mathrm{n}} \mathrm{C_n} e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x}) = \sum_{\mathrm{n}} \mathrm{C_n} e^{\frac{-i\mathrm{E_n}t}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x})$

Therefore, the initial state $\Psi(\mathrm{x},0)$ evolves in time such that each term in the expansion of the state in terms of the energy eigenstates is multiplied by its corresponding time dependent factor $e^{\frac{-i\mathrm{E_n}t}{\hbar}}$. Note that in general, the time dependent factor is different for each term because the energy corresponding to each term is generally different.

- As a special case, if $\Psi(\mathrm{x},0)$ itself is an energy eigenstate with eigenvalue $E_{\mathrm{n}}$, then $\Psi(\mathrm{x},0) = \psi_{\mathrm{n}}(\mathrm{x})$ and $\Psi(\mathrm{x},t) = e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\Psi(\mathrm{x},0) = e^{\frac{-i\hat{\mathrm{H}}t}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x}) = e^{\frac{-i\mathrm{E_n}t}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x})$. In this case, the time-evolution of the state is trivial because the state is multiplied by an overall phase factor.

**(MQS 3.3)**

Suppose at time $t = 0$, the initial wavefunction of a particle in a 1D infinite square well is $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$, where $\Psi_1(x)$ and $\Psi_2(x)$ are the ground state and first excited state wavefunctions. Choose all of the following statements that are correct for measurements on the system in this state at $t = 0$.

1. If **energy** is measured, the wavefunction will become either $\Psi_1(x)$ or $\Psi_2(x)$ immediately after the energy measurement but go back to $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$ a long time after the measurement.
2. If **position** is measured, the wavefunction $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$ will become a delta function immediately after the position measurement, but go back to $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$ a long time after the measurement.
3. If **position** is measured, the wavefunction $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$ will become a delta function immediately after the position measurement, and the wavefunction will remain the delta function a long time after the measurement.

A. 2 only B. 3 only C. 1 and 3 only
D. 1 and 2 only ***E. None of the above***

**Class discussion for MQS 3.3**

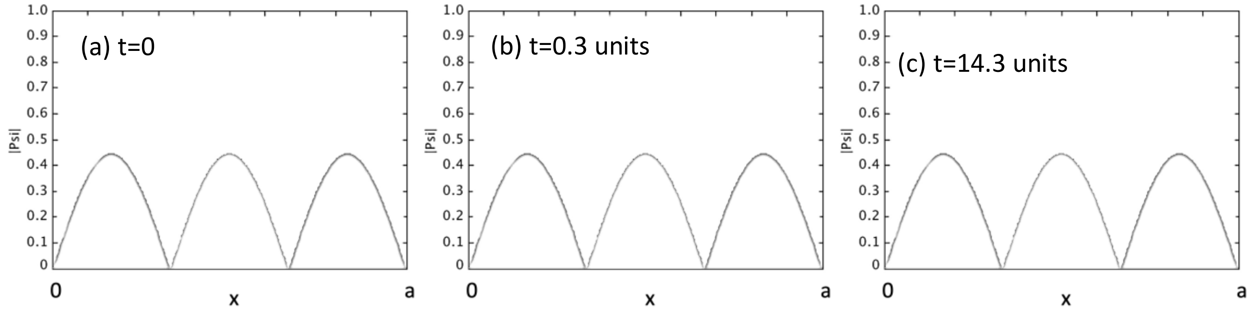
It should be emphasized to students that whenever talking about time evolution, we should expand the wave function in terms of energy eigenstates, and then multiply each term in the expansion by the corresponding time-dependent phase factor.

- If we measure energy at $t = 0$ and obtain $\mathrm{E_n}$, the wavefunction will instantaneously collapse to $\psi_n(x)$. After time $t_1$, this wave function becomes $e^{\frac{-iE_n t_1}{\hbar}}\psi_n(x)$. Even though there is a time dependent phase factor, $e^{\frac{-iE_n t_1}{\hbar}}\psi_n(x)$ is still an energy eigenstate with eigenvalue $E_n$. Thus, the spatial part of the wavefunction remains $\psi_n(x)$ after the energy measurement.

$$\begin{array}{ccc} t=0 & & t=t_1 \\ \Psi(x,0) = \psi_n(x) & -------\rightarrow & \Psi(x,t_1) = e^{\frac{-iE_n t_1}{\hbar}}\psi_n(x) \end{array}$$

**Note:**

- Multiplying a state with an overall phase factor doesn't change the state. Therefore, $e^{\frac{-iE_n t_1}{\hbar}}\psi_n(x)$ is the time dependence of the stationary state with energy $\mathrm{E_n}$.
- For example, if we measure energy at $t = 0$ and obtain $E_3$, the wave function will collapse to $\psi_3(x)$ right after an energy measurement. The three time-lapsed pictures from a simulation below show the time evolution of the absolute value of $\psi_3(x)$ in a 1D infinite square well with boundary (0<x<a). (a) shows the absolute value of the wavefunction right after the energy measurement. (b) and (c) show the absolute value of the wave function at $t = 0.3$ units and $t = 14.3$ units. As shown in the simulation, the absolute value of the wavefunction does not change with time. Thus, the wavefunction will remain $\psi_3(x)$ after the energy measurement.

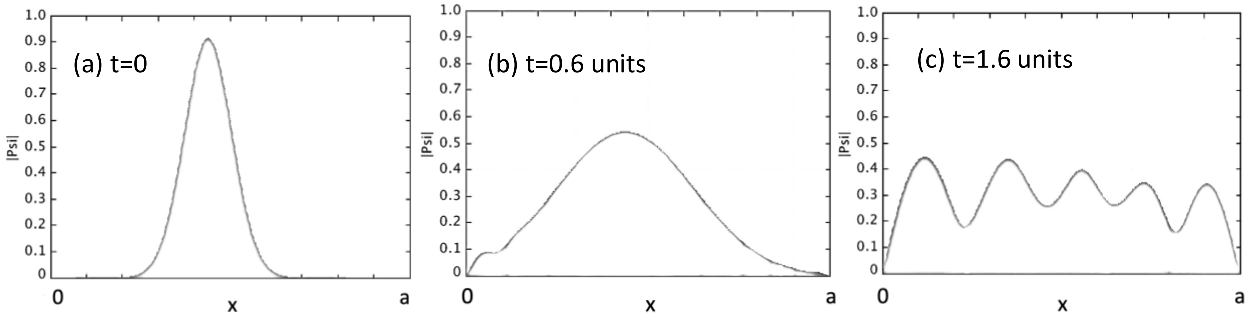



- If we measure position at $t = 0$ and obtain $x_0$, the wavefunction will instantaneously collapse to $\delta(x-x_0)$, which can be expanded as $\sum_n C_n\,\psi_n(x)$. After time $t_1$, the wave function becomes $\sum_n C_n\,e^{\frac{-iE_nt_1}{\hbar}}\psi_n(x)$. Thus, the wavefunction will neither remain $\delta(x-x_0)$ nor go back to $\frac{1}{\sqrt{2}}\left(\Psi_1(x)+\Psi_2(x)\right)$.

$$\Psi(x,0)=\delta(x-x_0)=\overset{t=0}{\sum_n C_n\,\psi_n(x)}\,-------\!\!\longrightarrow \Psi(x,t_1)=\overset{t=t_1}{\sum_n C_n\,e^{\frac{-iE_nt_1}{\hbar}}\psi_n(x)}$$

**Note:**

- The expansion coefficient in this case is $C_n=\int_{-\infty}^{\infty}\Psi(x,0)\psi_n^*(x)\mathrm{dx}=\int_{-\infty}^{\infty}\delta(x-x_0)\psi_n^*(x)\mathrm{dx}=\psi_n^*(x_0)$.
- In general, $\sum_n C_n\,e^{\frac{-iE_nt_1}{\hbar}}\psi_n(x)$ is not equal to $\delta(x-x_0)=\sum_n C_n\,\psi_n(x)$ , because $e^{\frac{-iE_nt_1}{\hbar}}$ are generally different for each $E_n$.
- The three time-lapsed pictures from a simulation below show the time evolution of $\delta(x-x_0)$ in a 1D infinite square well with boundary (0<x<a). (a) shows the absolute value of the wave function which is very localized (the actual wavefunction right after a position measurement is closer to a delta function which will be highly localized but the simulation is unable to show that level of a localized function so students would have to imagine a very peaked function). (b), (c) show the absolute value of the wave function at $t$ = 0.6 units and $t$ = 1.6 units. As shown in the simulation, the wave function does not remain localized (similarly, a delta function $\delta(x-x_0)$ after the position measurement will not remain localized at future times).
- Since delta function $\delta(x-x_0)$ contains nonzero coefficients $C_n$ for higher energy eigenstate wave function $\psi_n(x)(n>2)$, the probability of measuring these higher energies $\left|C_ne^{\frac{-iE_nt_1}{\hbar}}\right|^2$ would not be zero at future times. Therefore, the system will not return to the initial state $\frac{1}{\sqrt{2}}\left(\Psi_1(x)+\Psi_2(x)\right)$ after the position measurement, no matter how long we wait.

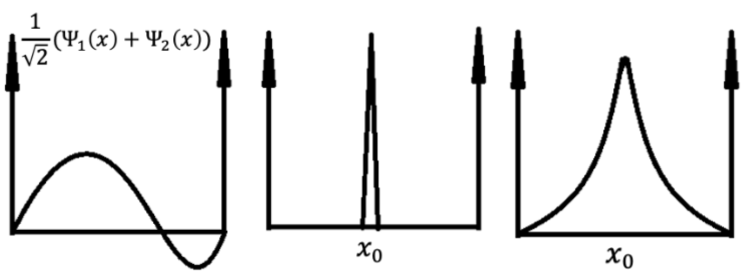



**(MQS 3.4)**

Suppose at time $t = 0$, a particle in a 1D infinite square well has the initial wavefunction $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$, where $\Psi_1(x)$ and $\Psi_2(x)$ are the ground state and first excited state wavefunctions (see the first graph below). You measure the position of the particle and obtain $x_0$. Choose all of the following statements that are correct.

1. The wavefunction will instantaneously collapse to a delta function at $x = x_0$ (the second graph is a schematic) when the position measurement is performed.
2. The wavefunction will remain a delta function at $x_0$ (one example is the second graph) a long time after the position measurement.
3. The wavefunction must evolve from a delta function at $x_0$ (one example is the third graph) so that it has different shapes a long time after the position measurement.

A. 1 only B. 1 and 2 only ***C. 1 and 3 only***
D. 2 and 3 only E. all of the above.



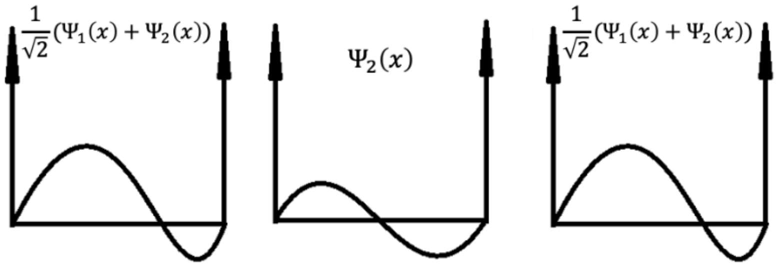

**(MQS 3.5)**

Suppose at time $t = 0$, a particle in a 1D infinite square well (0 < x < a) has the initial wavefunction $\Psi(x) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$, where $\Psi_1(x)$ and $\Psi_2(x)$ are the ground state and first excited state wavefunctions. You measure the energy of the particle and obtain $E_2$. Choose all of the following statements that are correct.

1. The wavefunction will instantaneously collapse to a stationary state when the energy measurement yielding $E_2$ is performed (the second graph).
2. The wavefunction will remain the first excited state a long time after the energy measurement.
3. The wavefunction must return to its initial state before the measurement a long time after the energy measurement (the third graph).

A. 1 only ***B. 1 and 2 only*** C. 3 only

D 1 and 3 only E. None of the above.




**(MQS 3.6)**

At time $t = 0$, the initial wavefunction of a particle is $\Psi(\mathrm{x},\ 0) = \frac{1}{\sqrt{2}}(\Psi_1(\mathrm{x}) + \Psi_2(\mathrm{x}))$, where $\Psi_1(\mathrm{x})$ and $\Psi_2(\mathrm{x})$ are the ground and first-excited energy eigenstates. Choose all of the following statements that are correct about a measurement performed on this system at $t = t_1$.

1. If we measure the position of the particle, the probability density for measuring the position of the particle will depend on the time $t_1$.
2. If we measure the energy of the particle, the probability of obtaining the energy $E_1$ or $E_2$ will depend on the time $t_1$.
3. If we measure an observable Q (with corresponding Hermitian operator $\widehat{\mathrm{Q}}$ which has eigenstates $\varphi_{\mathrm{q}}(\mathrm{x})$ and continuous eigenvalues q and eigenvalue equation $\widehat{\mathrm{Q}}\varphi_{\mathrm{q}}(\mathrm{x}) = \mathrm{q}\varphi_{\mathrm{q}}(\mathrm{x})$), the probability density for measuring $q$ will depend on the time $t_1$ (Note: $\widehat{\mathrm{Q}}$ does NOT commute with $\widehat{\mathrm{H}}$).

A. 1 only B. 2 only C. 3 only

***D. 1 and 3 only*** E. all of the above

**Class discussion for MQS 3.6**

Emphasize to students that when talking about time evolution, since the Hamiltonian of the system governs the time-evolution, we should expand the wave function in terms of energy eigenstates, and multiply each term by the corresponding time-dependent phase factor.

Initial state After waiting for $t$

$$\Psi(\mathrm{x},\ 0) = \sum_{\mathrm{n}} \mathrm{C_n}\ \psi_{\mathrm{n}}(\mathrm{x}) --------\rightarrow \Psi(\mathrm{x,t}) = \sum_{\mathrm{n}} \mathrm{C_n}\ \mathrm{e}^{\frac{-\mathrm{iE_n t}}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x})$$

- **If we measure energy at *t* = $t_1$, the probability of obtaining $E_{\mathrm{m}}$ is**
$\mathrm{P(E_m)} = |\langle\Psi_{\mathrm{m}}|\Psi(\mathrm{t_1})\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle\Psi_{\mathrm{m}}|x\rangle\langle x|\Psi(\mathrm{t_1})\rangle\mathrm{dx}\right|^2 = |\int_{-\infty}^{\infty}\Psi_{\mathrm{m}}^*(\mathrm{x})\Psi(\mathrm{x,t_1})\mathrm{dx}|^2 = \left|\sum_{\mathrm{n}}\mathrm{C_n e}^{\frac{-\mathrm{iE_n t_1}}{\hbar}}\int_{-\infty}^{\infty}\Psi_{\mathrm{m}}^*(\mathrm{x})\Psi_{\mathrm{n}}(\mathrm{x})\mathrm{dx}\right|^2 = \left|\sum_{\mathrm{n}}\mathrm{C_n e}^{\frac{-\mathrm{iE_n t_1}}{\hbar}}\delta_{mn}\right|^2 = \left|\mathrm{C_m e}^{\frac{-\mathrm{iE_m t_1}}{\hbar}}\right|^2 = |\mathrm{C_m}|^2$, which is independent of $t_1$.
- **If we measure position at *t* = $t_1$, the probability density for measuring $\mathrm{x_0}$ is**
$\rho(\mathrm{x_0}) = |\langle\mathrm{x_0}|\Psi(\mathrm{t_1})\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle\mathrm{x_0}|x\rangle\langle x|\Psi(\mathrm{t_1})\rangle\mathrm{dx}\right|^2 = \left|\int_{-\infty}^{\infty}\delta(x_0 - x)\Psi(\mathrm{x,t_1})\mathrm{dx}\right|^2 =$
$|\Psi(\mathrm{x_0,t_1})|^2 = \left|\sum_{\mathrm{n}}\mathrm{C_n}\ \mathrm{e}^{\frac{-\mathrm{iE_n t_1}}{\hbar}}\psi_{\mathrm{n}}(\mathrm{x_0})\right|^2$, which depends on $t_1$.
- **If we measure Q at *t* = $t_1$, the probability density for measuring *q* is**

$\rho(q) = |\langle q|\Psi(t_1)\rangle|^2 = \left|\int_{-\infty}^{\infty}\langle q|x\rangle\langle x|\Psi(t_1)\rangle dx\right|^2 = \left|\int_{-\infty}^{\infty} \varphi_q^*(x)\Psi(x,t_1)dx\right|^2 =$
$\left|\sum_n C_n\, e^{\frac{-iE_nt_1}{\hbar}} \int_{-\infty}^{\infty} \varphi_q^*(x)\, \psi_n(x)dx\right|^2$

Because $\hat{Q}$ does NOT commute with $\hat{H}$, $\hat{Q}$ and $\hat{H}$ don't have a complete set of simultaneous eigenstates. Therefore, the time factor in $\left|\sum_n C_n\, e^{\frac{-iE_nt_1}{\hbar}} \int_{-\infty}^{\infty} \varphi_q^*(x)\, \psi_n(x)dx\right|^2$ will not cancel out. Thus, $\rho(q)$ depends on $t_1$.

**Note:**

1. For the position measurement, the probability density for measuring position $x_0$ is $\rho(x_0) = |\langle x_0|\Psi(t_1)\rangle|^2 = |\Psi(x_0,t_1)|^2 = \left|\sum_n C_n\, e^{\frac{-iE_nt_1}{\hbar}}\psi_n(x_0)\right|^2$, which depends on time.
2. If you have discussed Dirac notation with students, you can also discuss calculation of the probability of obtaining $E_m$ in the following way: $P(E_m) = |\langle\Psi_m|\Psi(t_1)\rangle|^2 = \left|\sum_n C_n e^{\frac{-iE_nt_1}{\hbar}}\langle\Psi_m|\Psi_n\rangle\right|^2 = \left|\sum_n C_n e^{\frac{-iE_nt_1}{\hbar}}\,\delta_{mn}\right|^2 = \left|C_m e^{\frac{-iE_mt_1}{\hbar}}\right|^2 = |C_m|^2$, which is independent of $t_1$.

**Checkpoints:**

Consider a generic state $\Psi(x,\ 0)$ at $t = 0$.

- If energy is measured, what is the wave function instantaneously after and long time after the energy measurement?
- If position is measured, what is the wave function instantaneously after and long time after the position measurement?

**(MQS 4.1)**

The energy eigenvalues for a one-dimensional simple harmonic oscillator (SHO) are $E_n = \left(n+\frac{1}{2}\right)\hbar\omega\ \ (n = 0,1,2,3,\ldots)$. The initial wavefunction of a particle in a SHO potential energy well is $\Psi(x,\ 0) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$. You first measure the energy of the particle at $t = 0$ and obtain $\frac{3}{2}\hbar\omega$. Then immediately following the energy measurement, you measure the position of the particle. What is the probability of finding the particle in the region between $x_0$ and $x_0 + dx$?

1. $\frac{1}{2}(|\Psi_1(x_0)|^2+|\Psi_2(x_0)|^2)dx$
2. $|\Psi_1(x_0)|^2 dx$
3. $|\hat{x}\Psi_1(x)|^2 dx$

A. 1 only ***B. 2 only*** C. 3 only
D. 2 and 3 only E. None of the above

**Note:**

Emphasize to students that even though MQS 4.1 is about a quantum harmonic oscillator, the expression of probability density for measuring $x_0$ is generalizable to other quantum systems with different Hamiltonians such as the 1D infinite square well.

**(MQS 4.2)**

At time $t = 0$, the initial state of a particle in a 1D infinite square well is $\Psi(x,\ 0) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$, where $\Psi_1(x)$ and $\Psi_2(x)$ are the ground and first-excited energy eigenstates. You first measure the energy of the particle at $t = 0$ and obtain $E_1$. Then you measure the position of the particle at a later time $t = t_1$ (not immediately after the measurement of energy). What is the probability of finding the particle in the region between $x_0$ and $x_0 + dx$?

1. $\left|e^{\frac{-iE_1t_1}{\hbar}}\Psi_1(x_0)\right|^2 dx$
2. $\left|e^{\frac{-iE_1t_1}{\hbar}}\Psi_1(x_0) + e^{\frac{-iE_2t_1}{\hbar}}\Psi_2(x_0)\right|^2 dx$
3. $\left|\sum_n c_n e^{\frac{-iE_nt_1}{\hbar}}\Psi_n(x_0)\right|^2 dx$, where $c_n = \langle\Psi_n|\Psi\rangle$, n = 1, 2, 3…. and ($c_n \neq 0$ for 1, 2, 3….)

***A. 1 only*** B. 2 only C. 3 only

D. 1 or 2 depending on how much time has elapsed between the measurements

E. None of the above

**Note:**

- If you haven't discussed Dirac notation with students, please feel free to replace choice 3 with:
  $\left|\sum_n c_n e^{\frac{-iE_nt_1}{\hbar}}\Psi_n(x_0)\right|^2 dx$, where $c_n = \int_{-\infty}^{\infty}\Psi_n^*(x)\Psi(x)dx$, n = 1, 2, 3… and $c_n \neq 0$ for 1, 2, 3….
- It is worth noting that $\left|e^{\frac{-iE_1t_1}{\hbar}}\Psi_1(x_0)\right|^2 dx = |\Psi_1(x_0)|^2 dx$, which means that if the system is in a stationary state, the probability density for measuring position does not change with time.

**(MQS 4.3)**

The initial state of a particle in a 1D infinite square well (0 < x < a) is $\Psi(x,\ 0) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$. You first measure the position of the particle and obtain $x_0$. Then immediately following the position measurement, you measure the position of the particle again. Choose all of the following that are correct:

1. The second measurement must also yield $x_0$.
2. The second measurement could yield any of the infinitely many position eigenvalues possible for the system (0 < x < a).
3. The wavefunction immediately after the second measurement is still the position eigenstate corresponding to eigenvalue $x_0$.

A. 1 only B. 2 only C. 3 only

***D. 1 and 3 only*** E. 2 and 3 only

**(MQS 4.4)**

The initial state of a particle in a 1D infinite square well (0 < x < a) is $\Psi(x,\ 0) = \frac{1}{\sqrt{2}}(\Psi_1(x) + \Psi_2(x))$ when you measure the position of the particle and obtain $x_0$. Then some time *t* later following the position measurement, you measure the position of the particle again. Choose all of the following that are correct:

1. The second measurement must also yield $x_0$.
2. The second measurement could yield any of the infinitely many position eigenvalues possible for the system 0 < x < a.
3. The probability density for the second position measurement will depend on how much time elapses between the two measurements.

A. 1 only B. 2 only C. 3 only
D. 1 and 3 only ***E. 2 and 3 only***

**(MQS 4.5)**

The initial state of a particle in a 1D infinite square well (0 < x < a) is $\Psi(\mathrm{x},\ 0) = \frac{1}{\sqrt{2}}(\Psi_1(\mathrm{x}) + \Psi_2(\mathrm{x}))$ when you measure the position of the particle and obtain $\mathrm{x}_0$. Then some time *t* later following the position measurement, you measure the energy of the particle. Choose all of the following that are correct:

1. The energy measurement can yield any energy $\mathrm{E_n}$, where $\mathrm{n} = 1,2,3 \ldots \infty$.
2. The wavefunction will become either $\Psi_1(\mathrm{x})$ or $\Psi_2(\mathrm{x})$ immediately after the energy measurement and the system will remain in that state at future times.
3. The probability of obtaining the energy $\mathrm{E_1}$ will depend on how much time elapses between the two measurements.

***A. 1 only*** B. 2 only C. 3 only
D. 1 and 3 only E. 2 and 3 only

**Class discussion for MQS 4.1-4.5**

Emphasize to students that when talking about time evolution, we should expand the wave function in terms of energy eigenstates, and multiply each term by the corresponding time-dependent phase factor.

- **If we measure energy at *t* = 0 and obtain $\boldsymbol{E_n}$**, the wavefunction will instantaneously collapse to $\psi_n(\mathrm{x})$. After time $t_1$, the wave function becomes $e^{\frac{-iE_nt_1}{\hbar}}\psi_n(\mathrm{x})$.

$$\begin{array}{ccc} t = 0 & & t = t_1 \\ \Psi(\mathrm{x}, 0) = \psi_n(\mathrm{x}) & --------\rightarrow & \Psi(\mathrm{x}, t_1) = e^{\frac{-iE_nt_1}{\hbar}}\psi_n(\mathrm{x}) \end{array}$$

  - **If we measure energy at *t* = $\boldsymbol{t_1}$**, the probability of obtaining $E_m$ is
  $$P(E_m) = \left|\int_{-\infty}^{\infty}\Psi_m^*(\mathrm{x})\Psi(\mathrm{x}, t_1)d\mathrm{x}\right|^2 = \left|e^{\frac{-iE_nt_1}{\hbar}}\int_{-\infty}^{\infty}\psi_m^*(\mathrm{x})\psi_n(\mathrm{x})d\mathrm{x}\right|^2 = \delta_{mn}\begin{cases}1, & \text{if } m = n\\ 0, & \text{if } m \neq n\end{cases}$$
  Thus, the second measurement of energy will still yield $E_n$, and the result is independent of $t_1$. This means that the measurement of energy immediately after or a long time after the first energy measurement will yield the same value.
  - **If we measure position at *t* = $\boldsymbol{t_1}$**, the probability density for measuring $\mathrm{x}_0$ is
  $\rho(\mathrm{x}_0) = |\Psi(\mathrm{x}_0, t_1)|^2 = \left|e^{\frac{-iE_nt_1}{\hbar}}\psi_n(\mathrm{x}_0)\right|^2 = |\psi_n(\mathrm{x}_0)|^2$, which is independent of $t_1$. This means that the measurement of position immediately after or a long time after the energy measurement will have the same probability density for measuring $\mathrm{x}_0$.

➢ **If we measure position at $t$ = 0 and obtain $x_0$**, the wavefunction will instantaneously collapse to $\delta(x - x_0)$, which can be expanded to $\sum_n C_n \psi_n(x)$ (in which $C_n = \psi_n^*(x_0)$). After time $t_1$, the wave function becomes $\sum_n C_n e^{\frac{-iE_n t_1}{\hbar}} \psi_n(x)$.

$$\Psi(x,0) = \delta(x - x_0) = \overset{t\,=\,0}{\sum_n C_n \psi_n(x)} \dashrightarrow \Psi(x,t_1) = \overset{t\,=\,t_1}{\sum_n C_n e^{\frac{-iE_n t_1}{\hbar}} \psi_n(x)}$$

- **If we measure position at $t = t_1$,** the probability density for measuring $x_0$ is $\rho(x_0) = |\Psi(x_0,t_1)|^2 = \left|\sum_n C_n e^{\frac{-iE_n t_1}{\hbar}} \psi_n(x_0)\right|^2$, which depends on $t_1$. This means that the measurement of position immediately after or a long time after the first position measurement will generally have different probability density for measuring $x_0$.
- **If we measure energy at $t = t_1$**, the probability of obtaining $E_m$ is $P(E_m) = |\int_{-\infty}^{\infty} \psi_m^*(x)\Psi(x,t_1)dx|^2 = \left|\sum_n C_n e^{\frac{-iE_n t_1}{\hbar}} \int_{-\infty}^{\infty} \psi_m^*(x)\psi_n(x)dx\right|^2 = \left|\sum_n C_n e^{\frac{-iE_n t_1}{\hbar}} \delta_{mn}\right|^2 = |C_m|^2 = |\psi_m(x_0)|^2$, which is independent of $t_1$. This means that the measurement of energy immediately after or a long time after the position measurement will have the same probability of obtaining $E_m$.

**Note:**

It is important for instructors to discuss with students subtleties with regard to measurements of position. The eigenstates of the position operator are Dirac delta functions, which are not strictly normalizable or even physical in the sense that position is not infinitely resolved in any true measurement. A Dirac delta function can be expressed as an infinite series of energy eigenstates, each of which has an infinitesimally small amplitude. Students should understand the unrealistic aspects of working with delta functions, even though they have pedagogical value and are convenient approximations for spatially localized states.

**Overall Class Discussion/Summary**

If the particle is in an energy eigenstate $\psi_n(x)$ at $t$ = 0

- Measure energy at $t$ = 0:
  Probability of obtaining $E_n$ is 1 (and the probability would be the same if the measurement was made at $t = t_1$ instead of at $t$ = 0)
- Measure position at $t$ = 0:
  Probability density for measuring $x_0$ is $|\psi_n(x_0)|^2$ (and the probability density would be the same if the measurement was made at $t = t_1$ instead of at $t$ = 0)

If at $t$ = 0, the particle is in a state $\Psi(x)$ which is not an eigenstate of either $\hat{H}$ or any operator that commutes with $\hat{H}$.

- Measure energy at $t$ = 0:
  Probability of obtaining $E_n$ is $|\int_{-\infty}^{\infty} \psi_n^*(x)\Psi(x)dx|^2$ (and the probability would be the same if the measurement was made at $t = t_1$ instead of at $t$ = 0).
- Measure position at $t$ = 0:
  Probability density for measuring $x_0$ is $|\Psi(x_0)|^2$ (and the probability density would be different if the measurement was made at $t = t_1$ instead of at $t$ = 0). The position operator does not commute with the Hamiltonian, so position is not a constant of motion. The probability density for measuring position in a non-energy eigenstate depends on time.

## Appendix Q Questions in the Pre-Test and Post-Test

The pre- and post-test questions are reproduced below. The same information provided at the beginning of Appendix P applies to the questions in Appendix Q.

*All of the questions in this test refer to an isolated system in which a particle is in a 1-D infinite square well with Hamiltonian* $\hat{H} = \frac{\hat{p}^2}{2m} + V(x)$ $(V(x) = 0\ when\ 0 < x < a, V(x) = +\infty\ otherwise)$. *The energy eigenvalues are* $E_n = \frac{n^2\pi^2\hbar^2}{2ma^2}$ $(n = 1,2,3, \ldots)$, *and the energy eigenstate corresponding to* $E_n$ *is* $\Psi_n(x) = \sqrt{\frac{2}{a}} sin\,(\frac{n\pi x}{a})$ *when 0<x<a and* $\Psi_n(x) = 0$ *elsewhere.*

Q1. Choose all of the following statements that are correct about a generic state $\Psi(x)$ (which is not a stationary state).

I. $\hat{H}\Psi(x) = E_n\psi_n(x)$

II. $\hat{x}\Psi(x) = x\Psi(x)$

III. $\Psi(x)$ can be expressed as a linear superposition of the energy eigenstates.

Q2. The state of a particle at t=0 is given by $\Psi(x, 0) = \sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$.

(a) If you measure the energy of the particle at t = 0, what possible energies could you obtain and what is the probability of each? Explain.

(b) If you measure the position of the particle at t = 0, what possible values could you obtain, and what is the corresponding probability density? Explain.

Q3. The state of a particle at t=0 is given by $\Psi(x, 0) = \sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$

(a) If you measure energy at t = 0 and obtain a value of $E_1$, what is the normalized state of the system right after the measurement?

(b) Immediately after the measurement of energy in 3(a), you measure energy again. What is the probability of obtaining $E_1$?

(c) A long time after the measurement of energy in 3(a), you measure energy again. Will the probability of obtaining $E_1$ be the same or different as in 3(b)? Explain your reasoning.

(a) Immediately after the measurement of energy in 3(a), you measure position. What is the probability density of finding the particle at x = $x_0$? Explain.

(d) A long time after the measurement of energy in 3(a), you measure position. Will the probability density of finding the particle at x =$x_0$ be the same or different as in 3(d)? Explain your reasoning.

Q4. The state of a particle at t=0 is given by $\Psi(x, 0) = \sqrt{\frac{2}{7}}\Psi_1(x) + \sqrt{\frac{5}{7}}\Psi_2(x)$

(b) If you measure position and obtain a value of $x_0$, what is the wavefunction of the system right after the measurement?

(c) Immediately after the measurement of position in 4(a), you measure position again. Will the wavefunction of the system right after the measurement be the same or different as in 4(a)? Explain.

(d) A long time after the measurement of position in 4(a), you measure position again. Will the wavefunction of the system right after the measurement be the same or different as in 4(b)? Explain your reasoning.

(e) Immediately after the measurement of position in 4(a), you measure energy. What is the probability of obtaining $E_1$? Explain.

(f) A long time after the measurement of position in 4(a), you measure energy. Will the probability of obtaining $E_1$ be the same or different as in 4(d)? Explain your reasoning.

# Bibliography


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

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

[3] K. Rosa and F. M. Mensah, Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life, Phys. Rev. Phys. Educ. Res. **12**, 020113 (2016).

[4] 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).

[5] C. M. Steele and J. Aronson, Stereotype threat and the intellectual test performance of African Americans, J. Pers. Soc. Psychol. **69**, 797 (1995).

[6] 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).

[7] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, Phys. Rev. ST Phys. Educ. Res. **9**, 020121 (2013).

[8] J. R. Cimpian, T. H. Kim, and Z. T. McDermott, Understanding persistent gender gaps in STEM, Science **368**, 1317 (2020).

[9] J. Ellis, B. K. Fosdick, and C. Rasmussen, Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit, PLoS One **11**, e0157447 (2016).

[10] C. M. Ganley, C. E. George, J. R. Cimpian, and M. B. Makowski, Gender equity in college majors: Looking beyond the STEM/Non-STEM dichotomy for answers regarding female participation, Am. Educ. Res. J. **55**, 453 (2018).

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

[12] M. Pelch, Gendered differences in academic emotions and their implications for student success in STEM, International Journal of STEM Education **5**, 1 (2018).

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

[14] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, Phys. Rev. ST Phys. Educ. Res. **3**, 010107 (2007).

[15] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in modeling instruction in introductory university physics, Phys. Rev. ST Phys. Educ. Res. **6**, 010106 (2010).

[16] R. Ivie and A. Porter, Women in Physics and Astronomy, 2019, 2019.

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

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

[19] J. C. Blickenstaff, Women and science careers: Leaky pipeline or gender filter?, Gender and Education **17**, 369 (2005).

[20] 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).

[21] 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).

[22] 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).

[23] C. Lindstrøm and M. D. Sharma, Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter?, International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International) **19**, 1 (2011).

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

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

[26] R. Elliott, A. C. Strenta, R. Adair, M. Matier, and J. Scott, The role of ethnicity in choosing and leaving science in highly selective institutions, Res. High. Educ. **37**, 681 (1996).

[27] E. T. Pascarella and P. T. Terenzini, *How College Affects Students: Findings and Insights from Twenty Years of Research* (ERIC, 1991).

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

[29] S. Ahlqvist, B. London, and L. Rosenthal, Unstable identity compatibility: How gender rejection sensitivity undermines the success of women in science, technology, engineering, and mathematics fields, Psychol. Sci. **24**, 1644 (2013).

[30] S. G. Brainard and L. Carlin, A six-year longitudinal study of undergraduate women in engineering and science, J. Engin. Educ. **87**, 369 (1998).

[31] 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).

[32] S. Cheryan and V. C. Plaut, Explaining underrepresentation: A theory of precluded interest, Sex Roles **63**, 475 (2010).

[33] I. Rodriguez, E. Brewe, V. Sawtelle, and L. H. Kramer, Impact of equity models and statistical measures on interpretations of educational reform, Phys. Rev. ST Phys. Educ. Res. **8**, 020103 (2012).

[34] 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).

[35] K. G. Talley and A. M. Ortiz, Women's interest development and motivations to persist as college students in STEM: A mixed methods analysis of views and voices from a Hispanic-serving institution, International Journal of STEM Education **4**, 1 (2017).

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

[37] 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).

[38] 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).

[39] A. Bandura, Self-efficacy, in *Encyclopedia of Psychology*, 2nd ed., edited by R. J. Corsini (Wiley, New York, 1994), Vol. 3, p. 368.

[40] A. Bandura, Social cognitive theory of self-regulation, Organ. Behav. Hum. Decis. Process. **50**, 248 (1991).

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

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

[43] A. L. Zeldin, S. L. Britner, and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, J. Res. Sci. Teach. **45**, 1036 (2008).

[44] 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.

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

[46] 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).

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

[48] 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).

[49] K. A. Renninger and S. Hidi, Revisiting the conceptualization, measurement, and generation of interest, Educ. Psychol. **46**, 168 (2011).

[50] A. Wigfield, J. S. Eccles, U. Schiefele, R. W. Roeser, and P. Davis-Kean, Development of achievement motivation, in *Handbook of Child Psychology: Social, Emotional, and Personality Development, Vol. 3, 6th ed.*, edited by N. Eisenberg, W. Damon, and R. M. Lerner (John Wiley & Sons Inc, Hoboken, NJ, US, 2006), p. 933.

[51] H.-S. Lin, Z.-R. Hong, and Y.-C. Chen, Exploring the development of college students' situational interest in learning science, Int. J. Sci. Educ. **35**, 2152 (2013).

[52] 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).

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

[54] A. Wigfield and J. S. Eccles, Expectancy–value theory of achievement motivation, Contemp. Educ. Psychol. **25**, 68 (2000).

[55] 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).

[56] 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 the Physics Education Research Conference Sacramento, CA* (2016), p. 228.

[57] G. Potvin and Z. Hazari, The development and measurement of identity across the physical sciences, in *Proceedings of the 2013 Physics Education Research Conference Portland, OR* (2013), p. 281.

[58] 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).

[59] A. Y. Kim and G. M. Sinatra, Science identity development: an interactionist approach, International Journal of STEM Education **5**, 1 (2018).

[60] K. Atkins, B. M. Dougan, M. S. Dromgold-Sermen, H. Potter, V. Sathy, and A. Panter, "Looking at Myself in the Future": how mentoring shapes scientific identity for STEM students from underrepresented groups, International Journal of STEM Education **7**, 42 (2020).

[61] A. Singer, G. Montgomery, and S. Schmoll, How to foster the formation of STEM identity: Studying diversity in an authentic learning environment, International Journal of STEM Education **7**, 57 (2020).

[62] P. Vincent-Ruz and C. D. Schunn, The nature of science identity and its role as the driver of student choices, International Journal of STEM Education **5**, 48 (2018).

[63] 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).

[64] 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).

[65] Z. Y. Kalender, Doctoral Dissertation, University of Pittsburgh, 2019.

[66] 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).

[67] A. T. Danielsson, Exploring woman university physics students 'doing gender' and 'doing physics', Gender and Education **24**, 25 (2012).

[68] A. J. Gonsalves, Exploring how gender figures the identity trajectories of two doctoral students in observational astrophysics, Phys. Rev. Phys. Educ. Res. **14**, 010146 (2018).

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

[70] E. Seymour, N. M. Hewitt, and C. M. Friend, *Talking About Leaving: Why Undergraduates Leave the Sciences* (Westview Press, Boulder, CO, 1997), Vol. 12.

[71] 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).

[72] I. Rodriguez, G. Potvin, and L. H. Kramer, How gender and reformed introductory physics impacts student success in advanced physics courses and continuation in the physics major, Phys. Rev. Phys. Educ. Res. **12**, 020118 (2016).

[73] N. E. Betz and G. Hackett, Applications of self-efficacy theory to understanding career choice behavior, J. Soc. Clin. Psychol. **4**, 279 (1986).

[74] 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).

[75] 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).

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

[77] 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).

[78] Z. Hazari, E. Brewe, R. M. Goertzen, and T. Hodapp, The importance of high school physics teachers for female students' physics identity and persistence, The Phys. Teach. **55**, 96 (2017).

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

[80] T. J. Nokes-Malach, Z. Y. Kalender, E. Marshman, C. Schunn, and C. Singh, How is perception of being recognized by others as someone good at physics related to female and male students' physics identities, in *Physics Education Research Conference 2019 PER Conference* (2019).

[81] 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).

[82] 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).

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

[84] 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).

[85] 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).

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

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

[88] S. A. Crockett, A five-step guide to conducting SEM analysis in counseling research, Counseling Outcome Research and Evaluation **3**, 30 (2012).

[89] S. L. Hershberger and G. Marcoulides, The problem of equivalent structural models, in *Structural equation modeling: A second course*, edited by G. R. Hancock and R. O. Mueller (IAP, 2006), p. 13.

[90] 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).

[91] R. B. Kline, *Principles and Practice of Structural Equation Modeling* (Guilford publications, 2015).

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

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

[94] 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).

[95] 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, in *Proceedings of the 2017 Physics Education Research Conference Cincinnati, OH* (2017), p. 276.

[96] 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?,

in *Proceeding of the 2018 Physics Education Research Conference Washington, DC* (2018).

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

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

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

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

[101] 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. ST Phys. Educ. Res. **9**, 020115 (2013).

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

[103] W. S. Gosset, The probable error of a mean, Biometrika **6**, 1 (1908).

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

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

[106] R. P. Chalmers, mirt: A multidimensional item response theory package for the R environment, Journal of Statistical Software **48**, 1 (2012).

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

[108] A. G. Yong and S. Pearce, A beginner’s guide to factor analysis: Focusing on exploratory factor analysis, Tutorials in Quantitative Methods for Psychology **9**, 79 (2013).

[109] H. Akoglu, User's guide to correlation coefficients, Turkish Journal of Emergency Medicine **18**, 91 (2018).

[110] R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation, Vienna, 2013).

[111] See Supplemental Material at [URL will be inserted by publisher] for results of the other equivalent models.

[112] 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, The Phys. Teach. **58**, 484 (2020).

[113] C. S. Dweck, *Mindset: The New Psychology of Success* (Random House Digital, Inc., 2008).

[114] E. A. Canning, K. Muenks, D. J. Green, and M. C. Murphy, STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes, Sci. Adv. **5**, eaau4734 (2019).

[115] M. Mitchell, Situational interest: Its multifaceted structure in the secondary school mathematics classroom, J. Educ. Psychol. **85**, 424 (1993).

[116] Y. Lee, M. M. Capraro, and R. Viruru, The factors motivating students' STEM career aspirations: Personal and societal contexts, International Journal of Innovation in Science and Mathematics Education **26** (2018).

[117] S. Hicks, S. MacDonald, and E. Martin, Measuring Student Attitudes and Performance in order to Improve Future Performance and Enrolments in Senior Science Subjects, International Journal of Innovation in Science and Mathematics Education **25** (2017).

[118] R. Walker, Motivating science undergraduates: ideas and interventions, International Journal of Innovation in Science and Mathematics Education **18** (2010).

[119] M. Sarkar, T. Overton, C. Thompson, and G. Rayner, Undergraduate science students' perceptions of employability: Efficacy of an intervention, International Journal of Innovation in Science and Mathematics Education **25** (2017).

[120] K. Bartimote-Aufflick, A. Bridgeman, R. Walker, M. Sharma, and L. Smith, The study, evaluation, and improvement of university student self-efficacy, Studies in Higher Education **41**, 1918 (2016).

[121] J. D. Cribbs and K. Piatek-Jimenez, Exploring How Gender, Self-Identified Personality Attributes, Mathematics Identity, and Gender Identification Contribute to College Students' STEM Career Goals, International Journal of Innovation in Science and Mathematics Education **29** (2021).

[122] 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).

[123] M. Hill, M. D. Sharma, and Y. Xu, Pre-lecture online learning modules in university physics: Student participation, perceptions, and performance, International Journal of Innovation in Science and Mathematics Education **25** (2017).

[124] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, Understanding engineering identity through structural equation modeling, in *2013 IEEE Frontiers in Education Conference (FIE)*, IEEE (2013), p. 50.

[125] 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).

[126] S. Lee and S. Hershberger, A simple rule for generating equivalent models in covariance structure modeling, Multivariate Behavioral Research **25**, 313 (1990).

[127] 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 the 2018 Physics Education Research Conference Washington, DC* (2018), p. https://doi.org/10.1119/perc.2018.pr.Nokes.

[128] K. Whitcomb, Doctoral Dissertation, University of Pittsburgh, 2020.

[129] L. M. Santana and C. Singh, Negative impacts of an unwelcoming physics environment on undergraduate women, in *Physics Education Research Conference 2021 Virtual Conference* (2021), p. 377.

[130] M. T. Chi, Quantifying qualitative analyses of verbal data: A practical guide, The Journal of the Learning Sciences **6**, 271 (1997).

[131] H. E. Longino, Feminist standpoint theory and the problems of knowledge, Signs: Journal of Women in Culture and Society **19**, 201 (1993).

[132] K. Rolin, Standpoint theory as a methodology for the study of power relations, Hypatia **24**, 218 (2009).

[133] J. Fereday and E. Muir-Cochrane, Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development, International Journal of Qualitative Methods **5**, 80 (2006).

[134] M. M. Williams and C. George-Jackson, Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in STEM, Journal of Women and Minorities in Science and Engineering **20**, 99 (2014).

[135] K. M. Whitcomb, Z. Y. Kalender, T. J. Nokes-Malach, C. D. Schunn, and C. Singh, Comparison of self-efficacy and performance of engineering undergraduate women and men, International Journal of Engineering Education **36**, 1996 (2020).

[136] 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).

[137] A. Maries, K. Whitcomb, and C. Singh, Gender inequities throughout STEM, J. Coll. Sci. Teach. **51**, 27 (2022).

[138] 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, The Phys. Teach. **58**, 430 (2020).

[139] K. M. Cooper, V. R. Downing, and S. E. Brownell, The influence of active learning practices on student anxiety in large-enrollment college science classrooms, International Journal of STEM Education **5**, 1 (2018).

[140] B. J. England, J. R. Brigati, and E. E. Schussler, Student anxiety in introductory biology classrooms: Perceptions about active learning and persistence in the major, PLoS One **12**, e0182506 (2017).

[141] K. M. Whitcomb and C. Singh, Underrepresented minority students receive lower grades and have higher rates of attrition across STEM disciplines: A sign of inequity?, Int. J. Sci. Educ. **43**, 1054 (2021).

[142] E. Pollack, *The Only Woman in the Room: Why Science is Still a Boys' Club* (Beacon Press, 2015).

[143] A. Johnson, An intersectional physics identity framework for studying physics settings, in *Physics Education and Gender: Identity as an Analytic Lens for Research*, edited by A. J. Gonsalves and A. T. Danielsson (Springer International Publishing, Cham, 2020), p. 53.

[144] K. R. Binning, N. Kaufmann, E. M. McGreevy, O. Fotuhi, S. Chen, E. Marshman, Z. Y. Kalender, L. Limeri, L. Betancur, and C. Singh, Changing social contexts to foster equity in college science courses: An ecological-belonging intervention, Psychol. Sci. **31**, 1059 (2020).

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

[146] 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).

[147] J. D. Cribbs, Z. Hazari, G. Sonnert, and P. M. Sadler, Establishing an explanatory model for mathematics identity, Child Development **86**, 1048 (2015).

[148] 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).

[149] 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 *2018 Collaborative Network for Engineering and Computing Diversity Conference Crystal City, United States* (2018).

[150] S. Harter, *The construction of the self: A developmental perspective* (Guilford Press, New York, NY, US, 1999), The construction of the self: A developmental perspective.

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

[152] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice Hall, 1986).

[153] A. Bandura, W. Freeman, and R. Lightsey, Self-efficacy: The exercise of control, Journal of Cognitive Psychotherapy **13**, 158 (1999).

[154] 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).

[155] J. E. Côté and C. G. Levine, *Identity, formation, agency, and culture: A social psychological synthesis* (Psychology Press, 2014).

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

[157] 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.

[158] P. J. Silvia, Interest and interests: The psychology of constructive capriciousness, Review of General Psychology **5**, 270 (2001).

[159] P. J. Silvia, Self-efficacy and interest: Experimental studies of optimal incompetence, Journal of Vocational Behavior **62**, 237 (2003).

[160] R. W. Lent, S. D. Brown, and G. Hackett, Toward a unifying social cognitive theory of career and academic interest, choice, and performance, Journal of vocational behavior **45**, 79 (1994).

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

[162] D. Hammer, Epistemological beliefs in introductory physics, Cogn. Instr. **12**, 151 (1994).

[163] 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).

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

[165] 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, in *Proceedings of the 2017 Physics Education Research Conference Cincinnati, OH* (2017), p. 204.

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

[167] D. Kaplan, *Structural equation modeling: Foundations and extensions* (Sage Publications, 2008), Vol. 10.

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

[169] Z. Awang, *SEM made simple: A gentle approach to learning Structural Equation Modeling* (MPWS Rich Publication, 2015).

[170] C. K. Enders and D. L. Bandalos, The relative performance of full information maximum likelihood estimation for missing data in structural equation models, Structural Equation Modeling: A Multidisciplinary Journal **8**, 430 (2001).

[171] G. Cumming and S. Finch, Inference by eye: confidence intervals and how to read pictures of data, American Psychologist **60**, 170 (2005).

[172] T. E. Sutherland and C. C. Bonwell, *Using active learning in college classes: A range of options for faculty* (Jossey-Bass, 1996).

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

[174] L. C. McDermott, Research on conceptual understanding in mechanics, Physics today **37**, 24 (1984).

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

[176] 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).

[177] 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).

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

[179] 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).

[180] 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).

[181] N. I. Karim, A. Maries, and C. Singh, Do evidence-based active-engagement courses reduce the gender gap in introductory physics?, Eur. J. Phys. **39**, 025701 (2018).

[182] M.-C. Shanahan and M. Nieswandt, Creative Activities and Their Influence on Identification in Science: Three Case Studies, Journal of Elementary Science Education **21**, 63 (2009).

[183] 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).

[184] M. Good, A. Maries, and C. Singh, Impact of traditional or evidence-based active-engagement instruction on introductory female and male students' attitudes and approaches to physics problem solving, Phys. Rev. Phys. Educ. Res. **15**, 020129 (2019).

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

[186] C. Powell, C. Demetriou, and A. Fisher, Micro-affirmations in academic advising: Small acts, big impact, The Mentor: Innovative Scholarship on Academic Advising **15** (2013).

[187] D. S. Yeager, V. Purdie-Vaughns, J. Garcia, N. Apfel, P. Brzustoski, A. Master, W. T. Hessert, M. E. Williams, and G. L. Cohen, Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide, Journal of Experimental Psychology: General **143**, 804 (2014).

[188] G. Sonnert and M. F. Fox, Women, men, and academic performance in science and engineering: The gender difference in undergraduate grade point averages, The Journal of Higher Education **83**, 73 (2012).

[189] F. McLure, M. Won, and D. F. Treagust, 'Even though it might take me a while, in the end, I understand it': a longitudinal case study of interactions between a conceptual change strategy and student motivation, interest and confidence, Disciplinary and Interdisciplinary Science Education Research **2**, 1 (2020).

[190] D. Fortus and I. Touitou, Changes to students' motivation to learn science, Disciplinary and Interdisciplinary Science Education Research **3**, 1 (2021).

[191] D. W. Jackson, Y. Cheng, Q. Meng, and Y. Xu, "Smart" greenhouses and pluridisciplinary spaces: supporting adolescents' engagement and self-efficacy in computation across disciplines, Disciplinary and Interdisciplinary Science Education Research **4**, 1 (2022).

[192] W. R. Penuel, B. J. Reiser, T. A. McGill, M. Novak, K. Van Horne, and A. Orwig, Connecting student interests and questions with science learning goals through project-based storylines, Disciplinary and Interdisciplinary Science Education Research **4**, 1 (2022).

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

[194] G. M. Walton and G. L. Cohen, A question of belonging: Race, social fit, and achievement, J. Pers. Soc. Psychol. **92**, 82 (2007).

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

[196] L. R. Hausmann, J. W. Schofield, and R. L. Woods, Sense of belonging as a predictor of intentions to persist among African American and White first-year college students, Res. High. Educ. **48**, 803 (2007).

[197] R. F. Kizilcec, A. Saltarelli, P. Bonfert-Taylor, M. Goudzwaard, E. Hamonic, and R. Sharrock, Welcome to the Course: Early Social Cues Influence Women's Persistence in Computer Science, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), p. 1.

[198] G. M. Walton, M. C. Murphy, and A. M. Ryan, Stereotype threat in organizations: Implications for equity and performance, Annual Review of Organizational Psychology and Organizational Behavior **2**, 523 (2015).

[199] B. London, L. Rosenthal, S. R. Levy, and M. Lobel, The influences of perceived identity compatibility and social support on women in nontraditional fields during the college transition, Basic and Applied Social Psychology **33**, 304 (2011).

[200] E. Seymour, A.-B. Hunter, H. Thiry, T. J. Weston, R. P. Harper, D. G. Holland, A. K. Koch, and B. M. Drake, *Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education* (Springer International Publishing AG, Cham, 2019).

[201] T. M. Freeman, L. H. Anderman, and J. M. Jensen, Sense of belonging in college freshmen at the classroom and campus levels, The Journal of Experimental Education **75**, 203 (2007).

[202] J. G. Stout, T. A. Ito, N. D. Finkelstein, and S. J. Pollock, How a gender gap in belonging contributes to the gender gap in physics participation, in *AIP Conf. Proc.*, American Institute of Physics (2013), p. 402.

[203] J. N. Hughes, M. H. Im, and P. J. Allee, Effect of school belonging trajectories in grades 6–8 on achievement: Gender and ethnic differences, Journal of school psychology **53**, 493 (2015).

[204] G. J. Kennedy and B. W. Tuckman, An exploration into the influence of academic and social values, procrastination, and perceived school belongingness on academic performance, Social psychology of education **16**, 435 (2013).

[205] S.-f. Lam, B. P. Wong, H. Yang, and Y. Liu, Understanding student engagement with a contextual model, in *Handbook of research on student engagement* (Springer, 2012), p. 403.

[206] K. E. Voelkl, School identification, in *Handbook of research on student engagement.* (Springer Science + Business Media, New York, NY, US, 2012), p. 193.

[207] L. D. Pittman and A. Richmond, Academic and psychological functioning in late adolescence: The importance of school belonging, The Journal of Experimental Education **75**, 270 (2007).

[208] M. J. O'Connor, State University of New York at Buffalo, 2018.

[209] C. R. Glass and C. M. Westmont, Comparative effects of belongingness on the academic success and cross-cultural interactions of domestic and international students, International journal of intercultural relations **38**, 106 (2014).

[210] L. Bao and K. Koenig, Physics education research for 21 st century learning, Disciplinary and Interdisciplinary Science Education Research **1**, 1 (2019).

[211] L. McCullough, Gender differences in student responses to physics conceptual questions based on question context, ASQ Advancing the STEM Agenda in Education (2011).

[212] R. R. Hake, Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization, in *Physics Education Research Conference* (2002), p. 1.

[213] J. Docktor and K. Heller, Gender differences in both force concept inventory and introductory physics performance, in *AIP Conf. Proc.*, American Institute of Physics (2008), p. 15.

[214] C. T. Richardson and B. W. O'Shea, Assessing gender differences in response system questions for an introductory physics course, Am. J. Phys. **81**, 231 (2013).

[215] S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch, and N. R. Walet, Gender differences in conceptual understanding of Newtonian mechanics: A UK cross-institution comparison, Eur. J. Phys. **34**, 421 (2013).

[216] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, and A. Traxler, Exploring the gender gap in the Conceptual Survey of Electricity and Magnetism, Phys. Rev. Phys. Educ. Res. **13**, 020114 (2017).

[217] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[218] L. E. Kost-Smith, S. J. Pollock, and N. D. Finkelstein, Gender disparities in second-semester college physics: The incremental effects of a ``smog of bias", Phys. Rev. ST Phys. Educ. Res. **6**, 020112 (2010).

[219] N. Gonsar, L. Patrick, and S. Cotner, Graduate-and undergraduate-student perceptions of and preferences for teaching practices in STEM classrooms, Disciplinary and Interdisciplinary Science Education Research **3**, 1 (2021).

[220] R. Hake, Six lessons from the physics education reform effort, Latin american journal of physics education **1**, 24 (2007).

[221] P. B. Kohl and H. V. Kuo, Introductory physics gender gaps: Pre‐and post‐studio transition, in *AIP Conf. Proc.*, American Institute of Physics (2009), p. 173.

[222] M. J. Cahill, K. M. Hynes, R. Trousil, L. A. Brooks, M. A. McDaniel, M. Repice, J. Zhao, and R. F. Frey, Multiyear, multi-instructor evaluation of a large-class interactive-engagement curriculum, Phys. Rev. ST Phys. Educ. Res. **10**, 020101 (2014).

[223] A. Maries, N. I. Karim, and C. Singh, Does stereotype threat affect female students' performance in introductory physics?, in *AIP Conf. Proc.*, AIP Publishing LLC (2019), p. 120001.

[224] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, The Phys. Teach. **30**, 141 (1992).

[225] S. M. Stoen, M. A. McDaniel, R. F. Frey, K. M. Hynes, and M. J. Cahill, Force Concept Inventory: More than just conceptual understanding, Phys. Rev. Phys. Educ. Res. **16**, 010105 (2020).

[226] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, Am. J. Phys. **53**, 1043 (1985).

[227] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, Am. J. Phys. **66**, 338 (1998).

[228] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, Comparing the force and motion conceptual evaluation and the force concept inventory, Phys. Rev. ST Phys. Educ. Res. **5**, 010105 (2009).

[229] R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64 (1998).

[230] N. Lasry, S. Rosenfield, H. Dedic, A. Dahan, and O. Reshef, The puzzling reliability of the Force Concept Inventory, Am. J. Phys. **79**, 909 (2011).

[231] J. Stewart, H. Griffin, and G. Stewart, Context sensitivity in the force concept inventory, Phys. Rev. ST Phys. Educ. Res. **3**, 010102 (2007).

[232] A. A. diSessa, Toward an epistemology of physics, Cogn. Instr. **10**, 105 (1993).

[233] A. A. diSessa, N. M. Gillespie, and J. B. Esterly, Coherence versus fragmentation in the development of the concept of force, Cognitive science **28**, 843 (2004).

[234] R. N. Steinberg and M. S. Sabella, Performance on multiple-choice diagnostics and complementary exam problems, The Phys. Teach. **35**, 150 (1997).

[235] D. Hestenes and I. Halloun, Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller, The Phys. Teach. **33**, 502 (1995).

[236] I. Halloun and D. Hestenes, The search for conceptual coherence in FCI data, preprint (1996).

[237] A. Savinainen and J. Viiri, The force concept inventory as a measure of students conceptual coherence, International Journal of Science and Mathematics Education **6**, 719 (2008).

[238] S. Vosniadou, Capturing and modeling the process of conceptual change, Learning and instruction **4**, 45 (1994).

[239] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis: International Version* (Pearson, New Jersey, 2010).

[240] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, Hillsdale, N.J., 1988).

[241] I. Devlieger and Y. Rosseel, Factor score path analysis: An alternative for SEM?, Methodology: European Journal of Research Methods for the Behavioral and Social Sciences **13**, 31 (2017).

[242] L. Hatcher and N. O'Rourke, *A step-by-step approach to using SAS for factor analysis and structural equation modeling* (Sas Institute, 2013).

[243] T. Hayes and S. Usami, Factor score regression in the presence of correlated unique factors, Educational and psychological measurement **80**, 5 (2020).

[244] J. B. Grace and K. A. Bollen, Interpreting the results from multiple regression and structural equation models, Bulletin of the Ecological Society of America **84**, 283 (2005).

[245] R. Fauber, R. Forehand, A. M. Thomas, and M. Wierson, A mediational model of the impact of marital conflict on adolescent adjustment in intact and divorced families: The role of disrupted parenting, Child development **61**, 1112 (1990).

[246] K. L. Nadal, R. King, D. G. Sissoko, N. Floyd, and D. Hines, The legacies of systemic and internalized oppression: Experiences of microaggressions, imposter phenomenon, and stereotype threat on historically marginalized groups, New Ideas in Psychology **63**, 100895 (2021).

[247] T. N. Brannon, E. R. Carter, L. A. Murdock‐Perriera, and G. D. Higginbotham, From backlash to inclusion for all: Instituting diversity efforts to maximize benefits across group lines, Social issues and policy review **12**, 57 (2018).

[248] C. Choenarom, R. A. Williams, and B. M. Hagerty, The role of sense of belonging and social support on stress and depression in individuals with depression, Archives of psychiatric nursing **19**, 18 (2005).

[249] A. Deiglmayr, E. Stern, and R. Schubert, Beliefs in “brilliance” and belonging uncertainty in male and female STEM students, Frontiers in Psychology **10**, 1114 (2019).

[250] L. Bian, S.-J. Leslie, M. C. Murphy, and A. Cimpian, Messages about brilliance undermine women's interest in educational and professional opportunities, Journal of Experimental Social Psychology **76**, 404 (2018).

[251] S. Cheryan, B. J. Drury, and M. Vichayapai, Enduring influence of stereotypical computer science role models on women’s academic aspirations, Psychol. Women Quarterly **37**, 72 (2013).

[252] J. G. Stout, N. Dasgupta, M. Hunsinger, and M. A. McManus, STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM), J. Pers. Soc. Psychol. **100**, 255 (2011).

[253] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).

[254] K. M. Whitcomb and C. Singh, For physics majors, gender differences in introductory physics do not inform future physics performance, Eur. J. Phys. **41**, 065701 (2020).

[255] R. M. Felder, G. N. Felder, M. Mauney, 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. Engin. Educ. **84**, 151 (1995).

[256] 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).

[257] A. W. e. al., Social emotionaly and personality development, in *Handbook of child psychology* (Wiley, New York, 2006), Vol. 3.

[258] F. Pajares and D. Schunk, The development of academic self-efficacy, Development of Achievement Motivation. United States **7** (2001).

[259] V. Tinto, Classrooms as communities: Exploring the educational character of student persistence, The Journal of Higher Education **68**, 599 (1997).

[260] S. Cwik, K. Whitcomb, and C. Singh, How the learning environment predicts male and female students' motivational beliefs in algebra-based introductory physics courses, in *Physics Education Research Conference 2020 Virtual Conference* (2020), p. 104.

[261] Y. Li, K. Whitcomb, and C. Singh, How learning environment predicts male and female students' physics motivational beliefs in introductory physics courses, in *Physics Education Research Conference 2020 Virtual Conference* (2020), p. 284.

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

[263] M. Meeuwisse, S. E. Severiens, and M. P. Born, Learning environment, interaction, sense of belonging and study success in ethnically diverse student groups, Res. High. Educ. **51**, 528 (2010).

[264] R. Masika and J. Jones, Building student belonging and engagement: Insights into higher education students' experiences of participating and learning together, Teach. High. Educ. **21**, 138 (2016).

[265] A. W. Astin, *What Matters in College: Four Critical Years Revisited* (Jossey-Bass, San Francisco, CA, 1993).

[266] D. Hooper and J. Coughlan, i Mullen, M.(2008). Structural equation modelling: Guidelines for determining model fit, Electronic Journal of Business Research Methods **6**, 5360.

[267] 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. ST Phys. Educ. Res. **2**, 010101 (2006).

[268] 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, in *the Proceedings of the 2017 Physics Education Research Conference* (2018).

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

[270] J. E. Stets and P. J. Burke, A sociological approach to self and identity, in *Handbook of Self and Identity* (The Guilford Press, New York, 2003), p. 128.

[271] 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).

[272] D. Cordova and M. Lepper, Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice, J. Educ. Psychol. **88**, 715 (1996).

[273] S. Hidi and W. Baird, Strategies for increasing text-based interest and students' recall of expository texts, Reading Research Quarterly **23**, 465 (1988).

[274] S. Hidi, J. Weiss, D. Berndorff, and J. Nolan, The role of gender, instruction and a cooperative learning technique in science education across formal and informal settings, in *Interest and learning: Proceedings of the Seeon conference on interest and gender*, IPN *Kiel, Germany* (1998), p. 215.

[275] M. R. Lepper and D. I. Cordova, A desire to be taught: Instructional consequences of intrinsic motivation, Motivation and Emotion **16**, 187 (1992).

[276] J. A. Sloboda and J. Davidson, The young performing musician, in *Musical Beginnings: Origins and Development of Musical Competence* (Oxford University Press, 1996), p. 171.

[277] G. J. Feist, The development of scientific talent in Westinghouse finalists and members of the National Academy of Sciences, Journal of adult development **13**, 23 (2006).

[278] D. Rozgonjuk, T. Kraav, K. Mikkor, K. Orav-Puurand, and K. Täht, Mathematics anxiety among STEM and social sciences students: the roles of mathematics self-efficacy, and deep and surface approach to learning, International Journal of STEM Education **7**, 1 (2020).

[279] K. A. Blotnicky, T. Franz-Odendaal, F. French, and P. Joy, A study of the correlation between STEM career knowledge, mathematics self-efficacy, career interests, and career activities on the likelihood of pursuing a STEM career among middle school students, International journal of STEM education **5**, 1 (2018).

[280] L. Blatt, C. D. Schunn, E. Votruba-Drzal, and B. M. Rottman, Variation in which key motivational and academic resources relate to academic performance disparities across introductory college courses, International Journal of STEM Education **7**, 1 (2020).

[281] S. Cotner, L. M. Jeno, J. Walker, C. Jørgensen, and V. Vandvik, Gender gaps in the performance of Norwegian biology students: the roles of test anxiety and science confidence, International Journal of STEM Education **7**, 1 (2020).

[282] Z. D. Kirbulut and E. Uzuntiryaki-Kondakci, Examining the mediating effect of science self-efficacy on the relationship between metavariables and science achievement, Int. J. Sci. Educ. **41**, 995 (2019).

[283] E. B. Witherspoon, C. D. Schunn, R. M. Higashi, and E. C. Baehr, Gender, interest, and prior experience shape opportunities to learn programming in robotics competitions, International Journal of STEM Education **3**, 1 (2016).

[284] C. Carmichael, R. Callingham, I. Hay, and J. Watson, Statistical literacy in the middle school: The relationship between interest, self-efficacy and prior mathematics achievement, Australian journal of educational & developmental psychology **10**, 83 (2010).

[285] M. M. Nauta, J. H. Kahn, J. W. Angell, and E. A. Cantarelli, Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both?, Journal of Counseling Psychology **49**, 290 (2002).

[286] T. J. Tracey, Development of interests and competency beliefs: A 1-year longitudinal study of fifth-to eighth-grade students using the ICA-R and structural equation modeling, Journal of counseling psychology **49**, 148 (2002).

[287] J. Wang, Q. Li, and Y. Luo, Physics identity of Chinese students before and after Gaokao: The effect of high-stake testing, Research in Science Education, 1 (2020).

[288] H. Cheng, G. Potvin, R. Khatri, L. H. Kramer, R. M. Lock, and Z. Hazari, Examining physics identity development through two high school interventions, in *Physics Education Research Conference 2018* (2018).

[289] A. Prybutok, A. Patrick, M. Borrego, C. C. Seepersad, and M. Kirisits, Cross-sectional survey study of undergraduate engineering identity, in *American Society for Engineering Education Annual Conference* (2016).

[290] J. Cribbs, Z. Hazari, P. M. Sadler, and G. Sonnert, Development of an explanatory framework for mathematics identity, in *Psychology of Mathematics Education-North American (PME-NA) Chapter Conference* (2012).

[291] P. W. Lei and Q. Wu, Introduction to structural equation modeling: Issues and practical considerations, Educational Measurement: issues and practice **26**, 33 (2007).

[292] R. P. Bagozzi and Y. Yi, On the evaluation of structural equation models, Journal of the academy of marketing science **16**, 74 (1988).

[293] O. J. Dunn, Multiple comparisons among means, Journal of the American statistical association **56**, 52 (1961).

[294] R. M. Felder, G. N. Felder, and E. J. Dietz, A longitudinal study of engineering student performance and retention. V. Comparisons with traditionally‐taught students, J. Engin. Educ. **87**, 469 (1998).

[295] R. Barnett, *The Ecological University: A Feasible Utopia* (Routledge, 2017).

[296] D. W. Johnson, R. T. Johnson, and K. A. Smith, Cooperative learning: Improving university instruction by basing practice on validated theory, Journal on Excellence in University Teaching **25**, 1 (2014).

[297] M. W. Ohland, S. D. Sheppard, G. Lichtenstein, O. Eris, D. Chachra, and R. A. Layton, Persistence, engagement, and migration in engineering programs, J. Engin. Educ. **97**, 259 (2008).

[298] R. Suresh, The relationship between barrier courses and persistence in engineering, Journal of College Student Retention: Research, Theory & Practice **8**, 215 (2006).

[299] J. B. Main, K. J. Mumford, and M. W. Ohland, Examining the influence of engineering students' course grades on major choice and major switching behavior, International Journal of Engineering Education **31**, 1468 (2015).

[300] J. D. Cribbs, C. Cass, Z. Hazari, P. M. Sadler, and G. Sonnert, Mathematics identity and student persistence in engineering, International Journal of Engineering Education **32**, 163 (2016).

[301] C. E. Brawner, S. M. Lord, R. A. Layton, M. W. Ohland, and R. A. Long, Factors affecting women’s persistence in chemical engineering, International Journal of Engineering Education **31**, 1431 (2015).

[302] S. M. Lord, M. M. Camacho, R. A. Layton, R. A. Long, M. W. Ohland, and M. H. Wasburn, Who's persisting in engineering? A comparative analysis of female and male Asian, Black, Hispanic, Native American, and White students, Journal of Women and Minorities in Science and Engineering **15** (2009).

[303] H. M. Matusovich, R. A. Streveler, and R. L. Miller, Why do students choose engineering? A qualitative, longitudinal investigation of students' motivational values, J. Engin. Educ. **99**, 289 (2010).

[304] B. L. Yoder, *Engineering by the Numbers: ASEE Retention and Time-to-Graduation Benchmarks for Undergraduate Engineering Schools, Departments and Programs* 2017).

[305] G. Lichtenstein, H. G. Loshbaugh, B. Claar, H. L. Chen, K. Jackson, and S. D. Sheppard, An engineering major does not (necessarily) an engineer make: Career decision making among undergraduate engineering majors, J. Engin. Educ. **98**, 227 (2009).

[306] D. E. Chubin, G. S. May, and E. L. Babco, Diversifying the engineering workforce, J. Engin. Educ. **94**, 73 (2005).

[307] A. Godwin and G. Potvin, Fostering female belongingness in engineering through the lens of critical engineering agency, International Journal of Engineering Education **31**, 938 (2015).

[308] Y. Min, G. Zhang, R. A. Long, T. J. Anderson, and M. W. Ohland, Nonparametric survival analysis of the loss rate of undergraduate engineering students, J. Engin. Educ. **100**, 349 (2011).

[309] K. P. Cross, On college teaching, J. Engin. Educ. **82**, 9 (1993).

[310] S. Tobias, *They're Not Dumb, They're Different: Stalking the Second Tier* (Research Corporation, Tucson, AZ, 1990).

[311] M. Besterfield‐Sacre, M. Moreno, L. J. Shuman, and C. J. Atman, Gender and ethnicity differences in freshmen engineering student attitudes: A cross‐institutional study, J. Engin. Educ. **90**, 477 (2001).

[312] C. Seron, S. S. Silbey, E. Cech, and B. Rubineau, Persistence is cultural: Professional socialization and the reproduction of sex segregation, Work and Occupations **43**, 178 (2016).

[313] B. M. Capobianco, B. F. French, and H. A. Diefes-Dux, Engineering identity development among pre‐adolescent learners, J. Engin. Educ. **101**, 698 (2012).

[314] K. L. Tonso, *On the Outskirts of Engineering: Learning Identity, Gender, and Power via Engineering Practice* (Sense, Rotterdam, Netherlands, 2007).

[315] K. L. Meyers, M. W. Ohland, A. L. Pawley, S. E. Silliman, and K. A. Smith, Factors relating to engineering identity, Global Journal of Engineering Education **14**, 119 (2012).

[316] J. R. Morelock, A systematic literature review of engineering identity: Definitions, factors, and interventions affecting development, and means of measurement, European Journal of Engineering Education **42**, 1240 (2017).

[317] A. Patrick and M. Borrego, A review of the literature relevant to engineering identity, in *American Society for Engineering Education Annual Conference New Orleans, LA* (2016).

[318] 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, Journal of Social Issues **67**, 469 (2011).

[319] B. M. Capobianco, H. A. Diefes-Dux, and M. M. Habashi, Generating measures of engineering identity development among young learners, in *2009 39th IEEE Frontiers in Education Conference,* IEEE (2009), p. 1.

[320] A. Patrick, C. Seepersad, and M. Borrego, A combined model for predicting engineering identity in undergraduate students, in *2018 ASEE Annual Conference & Exposition Salt Lake City, UT* (2018).

[321] A. Godwin, The development of a measure of engineering identity, in *ASEE Annual Conference & Exposition* (2016).

[322] J. P. Gee, Chapter 3: Identity as an analytic lens for research in education, Review of Research in Education **25**, 99 (2000).

[323] F. Dehing, W. Jochems, and L. Baartman, Development of an engineering identity in the engineering curriculum in Dutch higher education: An exploratory study from the teaching staff perspective, European Journal of Engineering Education **38**, 1 (2013).

[324] L. Mann, P. Howard, F. Nouwens, and F. Martin, Influences on the development of students' professional identity as an engineer, in *Proceedings of the Research in Engineering Education Symposium Palm Cove, QLD* (2009), p. 1.

[325] L. Katehi, G. Pearson, and M. Feder, *Engineering in K-12 Education: Understanding the Status and Improving the Prospectus* (National Academies Press, Washington, DC, 2009).

[326] K. Whitcomb, Z. Y. Kalender, T. J. Nokes-Malach, C. Schunn, and C. Singh, Engineering students' performance in foundational courses as a predictor of future academic success, International Journal of Engineering Education **36**, 1340 (2020).

[327] K. Whitcomb, A. Maries, and C. Singh, Examining gender differences in a mechanical engineering and materials science curriculum, International Journal of Engineering Education **37**, 1261 (2021).

[328] J. Miles and M. Shevlin, *Applying regression and correlation: A guide for students and researchers* (Sage, 2001).

[329] A. Godwin and W. C. Lee, A cross-sectional study of engineering identity during undergraduate education, (2017).

[330] P. Jolly, D. Zollman, S. Rebello, and A. Dimitrova, Visualizing potential energy diagrams, Am. J. Phys **66**, 57 (1998).

[331] C. Singh, Student understanding of quantum mechanics, Am. J. Phys. **69**, 885 (2001).

[332] R. Müller and H. Wiesner, Teaching quantum mechanics on an introductory level, Am. J. Phys. **70**, 200 (2002).

[333] M. C. Wittmann, R. N. Steinberg, and E. F. Redish, Investigating student understanding of quantum physics: Spontaneous models of conductivity, Am. J. Phys. **70**, 218 (2002).

[334] D. Domert, C. Linder, and Å. Ingerman, Probability as a conceptual hurdle to understanding one-dimensional quantum scattering and tunnelling, Eur. J. Phys. **26**, 47 (2004).

[335] A. Kohnle, M. Douglass, T. J. Edwards, A. D. Gillies, C. A. Hooley, and B. D. Sinclair, Developing and evaluating animations for teaching quantum mechanics concepts, Eur. J. Phys. **31**, 1441 (2010).

[336] C. Manogue, E. Gire, D. McIntyre, and J. Tate, Representations for a spins-first approach to quantum mechanics, in *AIP Conf. Proc.*, American Institute of Physics (2012), p. 55.

[337] C. Singh, Student understanding of quantum mechanics at the beginning of graduate instruction, Am. J. Phys. **76**, 277 (2008).

[338] A. Kohnle, I. Bozhinova, D. Browne, M. Everitt, A. Fomins, P. Kok, G. Kulaitis, M. Prokopas, D. Raine, and E. Swinbank, A new introductory quantum mechanics curriculum, Eur. J. Phys. **35**, 015001 (2013).

[339] E. Gire and E. Price, Structural features of algebraic quantum notations, Phys. Rev. ST Phys. Educ. Res. **11**, 020109 (2015).

[340] G. Passante, P. J. Emigh, and P. S. Shaffer, Examining student ideas about energy measurements on quantum states across undergraduate and graduate levels, Phys. Rev. ST Phys. Educ. Res. **11**, 020111 (2015).

[341] P. J. Emigh, G. Passante, and P. S. Shaffer, Student understanding of time dependence in quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020112 (2015).

[342] V. Dini and D. Hammer, Case study of a successful learner's epistemological framings of quantum mechanics, Phys. Rev. Phys. Educ. Res. **13**, 010124 (2017).

[343] K. Krijtenburg-Lewerissa, H. J. Pol, A. Brinkman, and W. Van Joolingen, Insights into teaching quantum mechanics in secondary and lower undergraduate education, Phys. Rev. Phys. Educ. Res. **13**, 010109 (2017).

[344] U. S. di Uccio, A. Colantonio, S. Galano, I. Marzoli, F. Trani, and I. Testa, Design and validation of a two-tier questionnaire on basic aspects in quantum mechanics, Phys. Rev. Phys. Educ. Res. **15**, 010137 (2019).

[345] T. Wan, P. J. Emigh, and P. S. Shaffer, Investigating how students relate inner products and quantum probabilities, Phys. Rev. Phys. Educ. Res. **15**, 010117 (2019).

[346] T. Tu, C.-F. Li, Z.-Q. Zhou, and G.-C. Guo, Students' difficulties with partial differential equations in quantum mechanics, Phys. Rev. Phys. Educ. Res. **16**, 020163 (2020).

[347] C. Porter and A. Heckler, Graduate student misunderstandings of wave functions in an asymmetric well, Phys. Rev. Phys. Educ. Res. **15**, 010139 (2019).

[348] M. Wawro, K. Watson, and W. Christensen, Students' metarepresentational competence with matrix notation and Dirac notation in quantum mechanics, Phys. Rev. Phys. Educ. Res. **16**, 020112 (2020).

[349] T. Tu, C.-F. Li, J.-S. Xu, and G.-C. Guo, Students' difficulties with solving bound and scattering state problems in quantum mechanics, Phys. Rev. Phys. Educ. Res. **17**, 020142 (2021).

[350] P. Bitzenbauer, Effect of an introductory quantum physics course using experiments with heralded photons on preuniversity students' conceptions about quantum physics, Phys. Rev. Phys. Educ. Res. **17**, 020103 (2021).

[351] C. Singh, Transfer of learning in quantum mechanics, in *AIP Conf. Proc.*, American Institute of Physics (2005), p. 23.

[352] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. I. Performance on nonintuitive problem pair from introductory physics, Phys. Rev. ST Phys. Educ. Res. **4**, 010104 (2008).

[353] C. Singh, M. Belloni, and W. Christian, Improving students' understanding of quantum mechanics, Physics Today **59**, 43 (2006).

[354] C. Singh, Student difficulties with quantum mechanics formalism, in *AIP Conf. Proc.*, American Institute of Physics (2007), p. 185.

[355] C. Singh, Interactive learning tutorials on quantum mechanics, Am. J. Phys. **76**, 400 (2008).

[356] C. Singh and G. Zhu, Cognitive issues in learning advanced physics: An example from quantum mechanics, in *AIP Conf. Proc.*, AIP (2009), p. 63.

[357] A. Mason and C. Singh, Do advanced physics students learn from their mistakes without explicit intervention?, Am. J. Phys. **78**, 760 (2010).

[358] S.-Y. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, Eur. J. Phys. **31**, 57 (2009).

[359] G. Zhu and C. Singh, Improving students' understanding of quantum mechanics via the Stern–Gerlach experiment, Am. J. Phys. **79**, 499 (2011).

[360] C. Singh and G. Zhu, Improving students' understanding of quantum mechanics by using peer instruction tools, AIP Conf. Proc. **1413**, 77 (2012).

[361] G. Zhu and C. Singh, Surveying students' understanding of quantum mechanics in one spatial dimension, Am. J. Phys. **80**, 252 (2012).

[362] G. Zhu and C. Singh, Improving students' understanding of quantum measurement. I. Investigation of difficulties, Phys. Rev. ST Phys. Educ. Res. **8**, 010117 (2012).

[363] G. Zhu and C. Singh, Improving students' understanding of quantum measurement. II. Development of research-based learning tools, Phys. Rev. ST Phys. Educ. Res. **8**, 010118 (2012).

[364] G. Zhu and C. Singh, Improving student understanding of addition of angular momentum in quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **9**, 010101 (2013).

[365] E. Marshman and C. Singh, Investigating and improving student understanding of quantum mechanical observables and their corresponding operators in Dirac notation, Eur. J. Phys. **39**, 015707 (2017).

[366] S. DeVore and C. Singh, Development of an interactive tutorial on quantum key distribution, in *Physics Education Research Conference 2014 Minneapolis, MN* (2014), p. 59.

[367] B. R. Brown and C. Singh, Development and evaluation of a quantum interactive learning tutorial on Larmor precession of spin, in *Physics Education Research Conference 2014 Minneapolis, MN* (2014).

[368] C. Singh and E. Marshman, Review of student difficulties in upper-level quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020117 (2015).

[369] E. Marshman and C. Singh, Framework for understanding the patterns of student difficulties in quantum mechanics, Phys. Rev. ST Phys. Educ. Res. **11**, 020119 (2015).

[370] B. R. Brown, A. Mason, and C. Singh, Improving performance in quantum mechanics with explicit incentives to correct mistakes, Phys. Rev. Phys. Educ. Res. **12**, 010121 (2016).

[371] E. Marshman and C. Singh, Interactive tutorial to improve student understanding of single photon experiments involving a Mach–Zehnder interferometer, Eur. J. Phys. **37**, 024001 (2016).

[372] E. Marshman and C. Singh, Investigating and improving student understanding of the probability distributions for measuring physical observables in quantum mechanics, Eur. J. Phys. **38**, 025705 (2017).

[373] E. Marshman and C. Singh, Investigating and improving student understanding of the expectation values of observables in quantum mechanics, Eur. J. Phys. **38**, 045701 (2017).

[374] R. Sayer, A. Maries, and C. Singh, Quantum interactive learning tutorial on the double-slit experiment to improve student understanding of quantum mechanics, Phys. Rev. Phys. Educ. Res. **13**, 010123 (2017).

[375] S. Siddiqui and C. Singh, How diverse are physics instructors' attitudes and approaches to teaching undergraduate level quantum mechanics?, Eur. J. Phys. **38**, 035703 (2017).

[376] E. Marshman and C. Singh, Investigating and improving student understanding of quantum mechanics in the context of single photon interference, Phys. Rev. Phys. Educ. Res. **13**, 010117 (2017).

[377] A. Maries, R. Sayer, and C. Singh, Effectiveness of interactive tutorials in promoting "which-path" information reasoning in advanced quantum mechanics, Phys. Rev. Phys. Educ. Res. **13**, 020115 (2017).

[378] C. Keebaugh, E. Marshman, and C. Singh, Investigating and addressing student difficulties with a good basis for finding perturbative corrections in the context of degenerate perturbation theory, Eur. J. Phys. **39**, 055701 (2018).

[379] C. Keebaugh, E. Marshman, and C. Singh, Investigating and addressing student difficulties with the corrections to the energies of the hydrogen atom for the strong and weak field Zeeman effect, Eur. J. Phys. **39**, 045701 (2018).

[380] C. Keebaugh, E. Marshman, and C. Singh, Improving student understanding of corrections to the energy spectrum of the hydrogen atom for the Zeeman effect, Phys. Rev. Phys. Educ. Res. **15**, 010113 (2019).

[381] C. Keebaugh, E. Marshman, and C. Singh, Improving student understanding of fine structure corrections to the energy spectrum of the hydrogen atom, Am. J. Phys. **87**, 594 (2019).

[382] C. Keebaugh, E. Marshman, and C. Singh, Improving student understanding of a system of identical particles with a fixed total energy, Am. J. Phys. **87**, 583 (2019).

[383] E. Marshman and C. Singh, Validation and administration of a conceptual survey on the formalism and postulates of quantum mechanics, Phys. Rev. Phys. Educ. Res. **15**, 020128 (2019).

[384] P. Justice, E. Marshman, and C. Singh, Improving student understanding of quantum mechanics underlying the Stern–Gerlach experiment using a research-validated multiple-choice question sequence, Eur. J. Phys. **40**, 055702 (2019).

[385] A. Maries, R. Sayer, and C. Singh, Can students apply the concept of “which-path” information learned in the context of Mach–Zehnder interferometer to the double-slit experiment?, Am. J. Phys. **88**, 542 (2020).

[386] P. Justice, E. Marshman, and C. Singh, Student understanding of Fermi energy, the Fermi–Dirac distribution and total electronic energy of a free electron gas, Eur. J. Phys. **41**, 015704 (2019).

[387] S. DeVore and C. Singh, Interactive learning tutorial on quantum key distribution, Phys. Rev. Phys. Educ. Res. **16**, 010126 (2020).

[388] C. Singh, A. Asfaw, and J. Levy, Preparing students to be leaders of the quantum information revolution, arXiv preprint arXiv:2111.06438 (2021).

[389] G. Zhu and C. Singh, Students' difficulties with quantum measurement, in *AIP Conf. Proc.*, AIP (2012), p. 387.

[390] D. J. Griffiths and D. F. Schroeter, *Introduction to quantum mechanics* (Cambridge university press, 2018).

[391] P. Hu, Y. Li, and C. Singh, Challenges in addressing student difficulties with time-development of two-state quantum systems using a multiple-choice question sequence in virtual and in-person classes, Eur. J. Phys. (2022).

[392] P. Justice, E. Marshman, and C. Singh, Development, validation and in-class evaluation of a sequence of clicker questions on Larmor precession of spin in quantum mechanics, arXiv preprint arXiv:2006.05341 (2020).

[393] P. J. Emigh, G. Passante, and P. S. Shaffer, Developing and assessing tutorials for quantum mechanics: Time dependence and measurements, Phys. Rev. Phys. Educ. Res. **14**, 020128 (2018).

[394] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, N.J., 1997).

[395] C. H. Crouch and E. Mazur, Peer Instruction: Ten years of experience and results, Am. J. Phys. **69**, 970 (2001).

[396] P. Justice, E. Marshman, and C. Singh, Development and validation of a sequence of clicker questions for helping students learn addition of angular momentum in quantum mechanics, Physics Education Research Conference 2018 (2020).

[397] R. Sayer, E. Marshman, and C. Singh, Case study evaluating Just-In-Time Teaching and peer instruction using clickers in a quantum mechanics course, Phys. Rev. Phys. Educ. Res. **12**, 020133 (2016).

[398] D. Zingaro and L. Porter, Peer instruction in computing: The value of instructor intervention, Computers & Education **71**, 87 (2014).

[399] C. H. Crouch, J. Watkins, A. P. Fagen, and E. Mazur, Peer instruction: Engaging students one-on-one, all at once, Research-based reform of university physics **1**, 40 (2007).