Investigating Gender Differences in Students' Motivational Beliefs and inclusiveness of the Learning Environment in Introductory Physics Courses for Bioscience Majors

by

Sonja Cwik

B.A., Wellesley College, 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

Sonja Cwik

It was defended on

July 25, 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. Timothy Nokes-Malach, Professor, Department of Psychology

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

Investigating Gender Differences in Students' Motivational Beliefs and Inclusiveness of the Learning Environment in Introductory Physics Courses for Bioscience Majors

Sonja Cwik, PhD

University of Pittsburgh, 2022

Student grades and motivational outcomes in introductory physics courses can influence their retention in science, technology, engineering, and math (STEM) disciplines and future career aspirations. In recent years, many research studies have focused on inequities in calculus-based introductory physics courses, e.g., by investigating gender differences in students' motivational beliefs such as self-efficacy, interest, and identity and how they change from the beginning to the end of a physics course or course sequence. However, these issues have not been investigated in introductory physics courses for bioscience majors, in which women outnumber men. Although women outnumber men in these courses, societal stereotypes and biases about who can do and excel in physics may impact women in these courses unless there is an intentional effort to create equitable and inclusive learning environments.

In this dissertation I address the question of equity in introductory physics courses for students on the bioscience track by investigating the relationship between gender, motivational beliefs, and physics performance. Through my quantitative studies, I first analyzed gender differences in students' self-efficacy, sense of belonging, and recognition by others and how they predict course grade. Then, I investigated whether the relation between gender and physics identity was mediated by students' self-efficacy, interest, and recognition by others. Lastly, I investigated how students' perception of the inclusiveness of the learning environment including students' sense of belonging, interactions with their peers, and recognition by instructors and TAs predicts their physics identity and grades in the introductory physics courses. These findings can be invaluable for instructors striving to make these courses more equitable and inclusive. Throughout the thesis, there is discussion of how these findings can provide guidelines to improve women's experiences and achievement in these introductory physics courses.

Table of Contents

List of Tables

List of Figures

Preface

I would like to thank everyone who has supported me through my journey to achieve this milestone in my career.

I would like to thank my research advisor, Dr. Chandralekha Singh, for her help and encouragement throughout my four years at Pitt. Her guidance kept my motivation up and propelled me to become a better researcher. Knowing that I would meet with her kept me on my toes and working hard. Dr. Singh was very personable and interested in my life, which is something I admire about her. Wherever we would go, she would know people, their families and the work they were doing. She's the type of professor I'd like to be.

I would like thank all my education and physics professors at Wellesley College who helped me realize my potential and passion to pursue this research. Especially, I would like to thank Dr. Jerome Fung, the first to introduce me to physics education research and encouraged me to pursue my Ph.D.

I would like to thank all the members of my research group. Particularly, I would like to thank Yasemin Kalender, Kyle Whitcomb, and Danny Doucette for their mentorship and support in teaching me research methods and troubleshooting technical problems. I especially would like to thank my cohort within my research group, Doris Li and Ly Malespina, for their incredible friendship and support over the years. Their commiseration on difficult issues, and brainstorming research angles made the rigorous research more fun and rewarding.

My friends and family supported me in ways large and small. I would like to thank the friends I met at Wellesley, in Pittsburgh, as well as those from my childhood. Sara Vannah, who convinced me to take my first physics class in college. Biprateep Dey who helped fill out endless forms for presentations and with the technical aspects of Zoom. Bob and Lauren Caddy, my workout buddies who helped me to set reasonable boundaries and achieve balance while pursuing my research. Kristen George, who read every single Power Point presentation I ever made and helped to make them accessible. Of course my family, who was always there for me, and helped with recording presentations, editing my work, giving advice on job applications and professional milestones, and generally supporting me with love and concern.

While there isn’t room to mention everyone by name, I would like to thank all my friends, family, and colleagues who have helped me along the way!

1.0 Introduction

Prior research suggests that in many calculus-based introductory physics courses, women have lower physics motivational beliefs, including self-efficacy and identity, than men [1-3]. However, most of the prior studies in the college context concerning physics motivational beliefs have been conducted in classes in which women are underrepresented [4, 5]. Nevertheless, pervasive societal stereotypes and biases about who can excel in physics that women are bombarded with from a young age could impact their physics motivational beliefs, including their identity, even in physics courses in which they outnumber men, e.g., introductory physics courses for students on the bioscience track in which women are not underrepresented. The students' physics motivational beliefs in these courses could influence not only their participation and performance in the physics courses, which is very important in itself, but also in other science, technology, engineering, and math (STEM) disciplines in which they are intending to major or may be considering a career. For example, past studies have shown that the factors that influence students' physics identity influence their engineering career choices [6] and predict students' engineering identity [7] when the majority of students were first year students with intentions to major in a variety of disciplines (including STEM, humanities and social sciences). Similarly, these factors related to physics identity could influence students in the physics course sequence for bioscience and pre-health majors. In this dissertation, we focus on students who are on the bioscience track primarily interested in health-related professions, for whom two physics courses are mandatory. These physics courses are mandatory for them not only because the foundational knowledge of physics is central to developing a deep understanding of bioscience, but also because the kind of reasoning skills students develop in physics can help them in their majors and future careers. A majority of these bioscience majors are on the pre-health track and want to become health professionals. Higher physics motivational beliefs can help them engage and perform better, e.g., on the Medical College Admission Test (MCAT) in which 5-8% of the material consists of physics concepts, to realize their career goals. Several prior studies have shown performance gaps between men's and women's grades and scores on conceptual tests in calculus-based and algebra-based physics courses [8-13] that some hypothesize may be due to gender gaps in prior preparation or motivational beliefs [12-22].

Explicit and implicit societal stereotypes and biases that women internalize about physics could influence their motivational beliefs, including their physics identity. One common societal stereotype is that genius and brilliance are important factors to succeed in physics [23]. However, studies have found that by the age of six, girls are less likely than boys to believe they are "really really smart" and less likely to choose activities that are intended for "brilliant people" [24]. These types of stereotypes may continue to negatively impact women from childhood through college physics courses. In formal and informal situations, women are often made to feel that they cannot excel in physics-related fields. Additionally, negative societal stereotypes and biases may lead to a "chilly climate" for women in science classes [25]. These issues emanating from structural societal inequities related to physics can lower the physics motivational beliefs of women compared to men.

Moreover, in prior investigations, researchers have often hinted at the fact that the underrepresentation of women in those courses may be related to their lower physics motivational beliefs than men and the gender gap increasing from the beginning to the end of those courses [26]. In this dissertation, we investigated whether in physics courses in which women are not numerically underrepresented, a similar gender gap in physics motivational beliefs exists and whether these gender gaps get worse from the beginning to the end of the two course sequence. In particular, it is important to investigate whether numerical overrepresentation of women, e.g., in the physics courses for bioscience majors we focus on, would eliminate or significantly reduce the gender gap in physics motivational beliefs. If this is not the case, it may signify that classroom representation alone will not change the pernicious effects of systemic gender inequities in physics perpetuated by society and bolstered further by the physics learning environments. It may point to the need to view these gender gaps as more deep-rooted than simply due to women's underrepresentation in a physics course, as well as the need to address inequities in physics learning environments that exacerbate these gender gaps. Since physics motivational beliefs can impact student engagement and performance in physics courses and have the potential to impact their career choices, the persistence of a gender gap in these courses would be a marker of inequity that must be addressed.

1.1 Motivational Beliefs

Since students' motivational beliefs are important for students' success in introductory physics courses, we start with some background on the motivational beliefs that are studied in this dissertation. Self-efficacy in a discipline refers to students' belief in their ability to accomplish tasks or solve problems [27, 28]. It has been shown to influence students' engagement, learning, and persistence in science courses, in addition to contributing to students' science identity [14, 15, 29-42]. For example, when tackling difficult problems, students with high self-efficacy tend to view the problems as challenges that can be overcome, whereas those with low self-efficacy tend to view them as personal threats to be avoided [27]. However, in introductory physics courses in which women are underrepresented, studies have found a gender gap in self-efficacy disadvantaging women that widens by the end of the course, even in interactive engagement courses [29, 43]. Additionally, self-efficacy has been shown to predict students' engagement, learning, and persistence in science courses [14, 15, 27, 29-42].

Interest in a particular discipline may affect students' perseverance, persistence, and achievement [16, 38, 41, 44-46]. One study showed that changing the curriculum to stimulate the interest of female students helped improve all of the students' understanding at the end of the year [47]. Within expectancy-value theory, interest and self-efficacy are constructs that predict students' academic outcomes and career expectations [48]. We focus on intrinsic interest or an individual's interest and enjoyment in engaging with physics concepts in this dissertation.

Another belief, perceived recognition by others, has been shown to play an important role in a student's identity [49] and motivation to excel [50]. In a study on students' perception of support, teacher support was more strongly linked to the motivation and engagement of girls than boys [50]. However, studies have shown that female students are not recognized appropriately in many STEM disciplines even before they enter college [1, 4, 24]. One study found that science faculty members in biological and physical sciences exhibit biases against female students by rating comparable male students significantly more competent [51]. Moreover, our prior research suggests that students' perceived recognition by instructors and teaching assistants can impact their self-efficacy and interest in physics and students' physics self-efficacy can impact their interest in physics (e.g., see [52, 53]).

Students' identity in STEM disciplines is important to study since it plays a key role in students' participation in classes as well as career decisions [5, 49, 54-57]. Physics identity has been studied in physics classes and has been shown to be connected with a student's self-efficacy, interest, and perceived recognition [5, 58, 59]. The science identity framework draws on the work of Gee who defined an individual's identity as being recognized as a certain kind of person in a given context and emphasized that identity can change over time [56]. It has been adapted in the science context to address the identities of both students and scientists [55]. Carlone and Johnson's science identity framework includes three interrelated dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others") [55]. Hazari et al. modified the framework by adapting it to physics specifically rather than science more generally [5]. "Competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a good physics student. Lastly, a fourth dimension, interest, was added to the framework since students have highly varying levels of interest in physics [60, 61]. In subsequent studies by Hazari for introductory students, performance and competence are combined into one variable [58]. In a slightly reframed version of Hazari's physics identity framework by Kalender et al. [59], performance/competence was framed as self-efficacy (closely related to competency belief). Additionally, recognition was framed explicitly as "perceived recognition" by students for clarity.

In addition to motivational factors that predict physics identity, other motivational factors that contribute to the student perception of the inclusiveness of the learning environment can also influence how women perform in their STEM classes and beyond [62]. For example, students' sense of belonging in physics has been shown to correlate with retention and self-efficacy [50, 63]. In addition, it is shown to be a predictor of students' physics identity for senior physics majors [58]. In addition, students' interaction with peers has been shown to enhance understanding and engagement in courses.

1.2 Overview

In chapter 2, we begin with a theoretical framework that emphasizes the importance of creating an equitable and inclusive learning environment so that students from all demographic groups benefit from evidence-based active engagement curricula and pedagogies. We then discuss both quantitative and qualitative assessment findings (using survey data and ethnographic and interview data) that suggest that without explicit thoughts and measures, undergraduate physics learning environments are not equitable and inclusive. Our focus here is on two marginalized groups in physics: women and ethnic/racial minority students, who are severely underrepresented in physics. We discuss research that shows highly troubling trends, e.g., ethnic and racial minority students drop out of the physics major at twice the rate compared to white students and women drop out with significantly higher GPAs than men. In addition to course level performance gaps, we discuss findings about motivational beliefs that show that women and ethnic/racial minority students often have lower physics sense of belonging, self-efficacy, perceived recognition by others including instructors, and identity compared to students from the dominant group. We also discuss how stereotype threat can result in the deteriorated performance of marginalized students. Finally, we discuss how social psychological classroom interventions and an explicit focus on creating student-centered inclusive classrooms can lead to more equitable outcomes.

In chapter 3, we examine the self-efficacy of men and women with similar performance in introductory physics courses for students on the bioscience track in which women outnumber men. We find a gender gap in self-efficacy disadvantaging women at the beginning of the course. However, unlike courses in which women are underrepresented, in which the self-efficacy gender gap often increases from the beginning to the end of the courses, we find that the self-efficacy gender gap for students who received a certain grade either remained constant or decreased.

Moreover, except for the students who received an A grade, the average self-efficacy of most of the other student groups decreased from the beginning to the end of the semester. Additionally, we find that most of the self-efficacy gender gap is due to students' biased perceptions about their capability rather than the performance difference between women and men.

In the next few chapters, we examine how motivational beliefs predict students' grades in the introductory physics courses. In chapter 4, we investigated how students' self-efficacy predicts female and male students' grades at the end of the physics course using Structural Equation Modeling (SEM). We found that women had a lower self-efficacy and grades than men in this course and that the students' self-efficacy played a major role in predicting students' grades in the course.

In chapter 5, we investigated how the students' perceived recognition by instructors and teaching assistants (TAs) as a physics person predicts their grades at the end of their physics course. We found that overall, women had lower perceived recognition than men as a physics person and their perceived recognition played an important role in predicting course grades, controlling for high school GPA and math SAT scores. Since physics as a discipline presents a barrier to women due to deep-rooted societal stereotypes and biases about who can excel in it, the numerical representation of women alone in these courses does not imply that they will feel recognized by their instructors and TAs as a physics person without an intentional effort to make the learning environment equitable and inclusive. These findings suggest that physics instructors and TAs should focus on changing the culture in their physics classes and create an equitable and inclusive learning environment in which students from traditionally marginalized demographic groups, e.g., women, feel recognized and can excel.

In chapter 6, we investigated how students' sense of belonging predicts female and male students' grades at the end of the physics course. We found that women had a lower sense of belonging and average grade than men in this course and that the students' sense of belonging played a major role in predicting students' grades in the course. In addition, while men's sense of belonging significantly increased from the beginning to the end of the physics course, women's sense of belonging did not significantly change by the end of the course.

In chapter 7, we used the physics identity framework to investigate whether the relation between gender and physics identity was mediated by other motivational beliefs including perceived recognition by others, self-efficacy, and interest in physics 1. The model shows that perceived recognition by others, self-efficacy, and interest mediated students' physics identity and there was no direct path from gender to identity. In addition, we find that although female students outnumbered male students in these courses, they had lower physics motivational beliefs including self-efficacy and identity at the beginning of the physics course, and this gender gap increased by the end of the course. These findings related to the gender gap in physics motivational beliefs are valuable because they may signify that classroom representation alone will not change the pernicious effects of systemic gender inequities in physics perpetuated by society and bolstered further by the inequitable and non-inclusive physics learning environments.

In chapter 8, we analyzed the motivational beliefs of students longitudinally in introductory physics 1 and physics 2 and used the physics identity framework to investigate whether the relation between gender and physics identity was mediated by self-efficacy, interest, and perceived recognition by others. Our findings show a gender gap in beliefs disadvantaging women throughout the physics course sequence. Additionally, the SEM model of physics identity shows that physics perceived recognition by others plays a central role in predicting students' physics identity throughout the two-semester course sequence.

Since students in these physics course were mainly bioscience majors or chemistry majors on the bioscience track, their science identity is also important for their major. Therefore, in chapter 9 we investigated how the physics self-efficacy and perceived recognition by others for bioscience majors enrolled in the second semester of a physics sequence predicted their overall science identity aligned with their disciplinary major. We find that bioscience majors' physics self-efficacy and perceived recognition not only predict their physics identity but also their overall science identity. These relations between physics self-efficacy and perceived recognition and the overall science identity of bioscience majors suggest interdisciplinary connections that may provide additional pathways for boosting students' science identity, e.g., by enhancing their self-efficacy and perceived recognition in their other mandatory courses such as physics. We also find that on average, women majoring in bioscience had lower physics self-efficacy, perceived recognition, physics identity, and overall science identity than men even though women were not underrepresented in the physics course.

In chapters 10 and 11, we investigated how the perception of the inclusiveness of the learning environment (which includes perceived recognition, peer interaction, and sense of belonging) predicts male and female students' physics outcomes, including their physics self-efficacy, interest, and identity in physics 1 and physics 2, respectively. We found that in general, women had lower physics beliefs than men and the learning environment plays a major role in explaining student outcomes. We also found that perceived recognition played an important role in predicting students' physics identity and students' sense of belonging played an important role in predicting students' physics self-efficacy. These findings can be useful to contemplate strategies to create an equitable and inclusive learning environment to help all students to excel in these physics courses.

In chapters 12 and 13, we added students' academic outcomes, such as their course grade, into the models for physics 1 and physics 2, respectively. In particular, we investigated female and male students' perceptions of the inclusiveness of the learning environment (including their sense of belonging, perceived recognition by others such as instructors, and perceived effectiveness of peer interaction) and how it predicted their physics course grades, self-efficacy, interest, and identity at the end of the course. We find gender differences in perceptions of the inclusiveness of the learning environment disadvantaging female students. Moreover, using SEM, we find that all factors in students' perception of the inclusiveness of the learning environment are essential to predicting their self-efficacy, interest, grade, and identity at the end of the physics courses. The findings about gender differences call into question equity in learning in these physics courses. How the perception of the inclusiveness of the learning environment predicts the outcomes can provide baseline data and be invaluable for strategies for creating an equitable and inclusive learning environment to help all students excel in these algebra-based physics courses.

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

2.0 Framework for and Review of Research on Assessing and Improving Equity and Inclusion in Undergraduate Physics Learning Environments

2.1 Introduction and Framework

Women and ethnic/racial minority (ERM) students are severely underrepresented in physics courses, majors, and careers [64, 65]. The societal stereotypes and biases about who belongs in physics and can excel in it, the lack of role models, and the chilly unsupportive, and competitive climate in physics make the playing field uneven for the traditionally underrepresented groups.

In physics education research, equity and inclusion have been studied with a focus on different traditionally underrepresented demographic student groups, e.g., women, ERM students, students with disability and LGBTQ+ etc. at different points in their physics learning, e.g., high school, different levels in college, graduate school, etc. In this chapter, we discuss a theoretical framework and then review research on assessing and improving equity in physics learning environments focusing only on women and ERM *undergraduate students* with data collected from students in physics classes. In other words, our focus is *only* on research involving assessment and strategies to make physics learning environments equitable and inclusive for women and ERM students in college level undergraduate physics courses for both physics majors and non-majors. Prior studies have focused on assessing differences in the experiences and outcomes of different demographic groups in physics courses and physics majors as a whole to measure inequities in order to devise strategies to improve outcomes and level the playing field.

In conducting and interpreting research on underrepresented student groups in physics, intersectionality is a useful framework because a single demographic characteristic, e.g., gender or ethnicity/race alone cannot fully explain the intricacies of the obstacles that students face [66-68]. In particular, a combination of different aspects of an individual's social identity (e.g., gender and ethnicity/race) leads to unique levels of disadvantages that cannot be explained by simply adding together the effects of the individual components of their identity [69]. For example, according to the framework of intersectionality, in many STEM disciplines where the societal norm expects that students are white men, the experience of a Black woman is not a simple sum of the experiences of being a woman and being Black [68, 70]. In particular, some researchers have argued for the use of critical race theory and feminist standpoint theory in physics education research [71]. Before proceeding further, we first explicate how we conceptualize equity in physics learning.

Our conceptualization of equity in physics learning includes three pillars: equitable access and opportunity to learn physics, equitable and inclusive learning environment, and equitable outcomes. Thus, by equity in physics 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 physics courses have comparable learning outcomes. This conceptualization of equitable outcome is consistent with Rodrigues et al.'s equity of parity model [72]. The physics learning outcomes include student performance in courses as well as evolution in their motivational beliefs such as physics self-efficacy, identity, etc. because regardless of performance, students' motivational beliefs can influence their short and long-term retention in physics courses, major, and careers. In other words, an equitable and inclusive learning environment should be student-centered so that all students are provided appropriate support and 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 enjoy learning physics and embrace challenges as learning opportunities instead of being threatened by them. Equitable learning outcomes for physics and other science, technology, engineering, and math (STEM) majors also include the ability of the physics courses to empower students from all demographic groups and make them passionate about pursuing further learning and careers in related areas. We note that equitable access and opportunity to learn physics, equitable and inclusive learning environment, and equitable outcomes are strongly entangled with each other. For example, if the physics learning environment is not equitable and inclusive, the learning outcomes are unlikely to be equitable.

Our conceptualization of equity in physics learning is mindful of the pervasive societal stereotypes and biases about physics as well as the lack of role models that can have a detrimental psychological impact on women and ERM students who are severely underrepresented. In general, when students struggle to solve challenging physics problems, they often respond in one of two ways. Some question whether they have what is needed to excel in physics. Others enjoy the struggle because it means that they are tackling new physics and learning. The negative reaction is a manifestation of a fixed mindset, i.e., believing that intelligence is immutable and struggling is a sign of a lack of intelligence, whereas the positive reaction is the sign of a growth mindset (the fact that your brain's capabilities can grow with deliberate effort and one can become an expert in a field by working hard). In an inequitable and non-inclusive learning environment, due to societal stereotypes and lack of role models, the underrepresented students are more likely than others to fall prey to the fixed mindset trap and view their struggle with challenging physics problems in a negative light. In fact, compared to most other STEM fields, the societal stereotypes are stronger in physics, a field whose history is often told through the stories of brilliant men. These stereotypes and lack of role models can also contribute to a lower sense of belonging for women and ERM students in physics learning environments unless explicit efforts are made to make them equitable and inclusive.

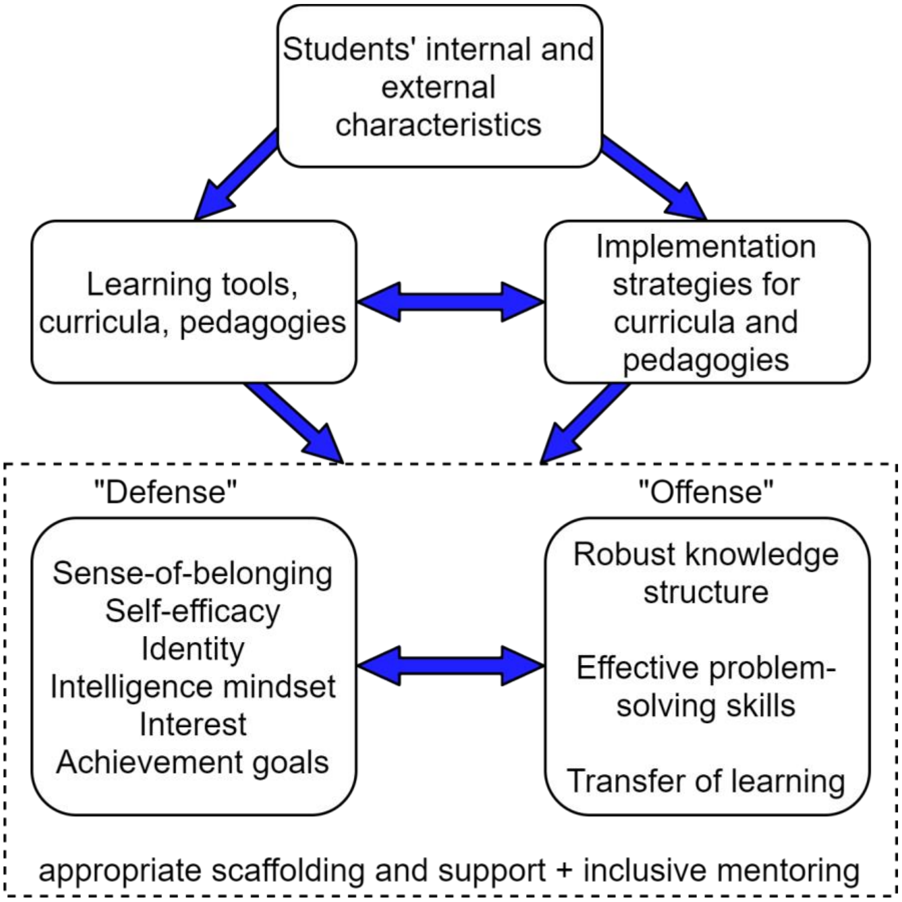


Figure 1 Holistic Ecosystem for Learning Physics in an Inclusive and Equitable Environment (HELPIEE) framework.

With our conceptualization of equity in learning, this chapter focuses on research on assessing and improving equity and inclusion using the Holistic Ecosystem for Learning Physics in an Inclusive and Equitable Environment (HELPIEE) framework adapted from the Strategies for Engaged Learning Framework (SELF) that emphasizes a holistic approach to helping all students learn physics [26]. The HELPIEE framework (see Figure 1) adapts the SELF framework to explicitly indicates that an important characteristic of an inclusive and equitable learning environment is that it will lead to equitable outcomes, which are comparable for all demographic groups with regard to students' physics knowledge structure and problem-solving skills as well as physics motivational beliefs (e.g., sense of belonging, self-efficacy, identity, etc.). This framework in Figure 1 expresses the fact that not only should physics courses have learning outcomes based upon physics related knowledge and skills we want students to learn but also those that focus on whether all students (and especially those from underrepresented groups) have a high sense of belonging, self-efficacy, growth mindset, and identity as people who can excel in physics.

Drawing analogy with sports, we note that to help players excel in any game, such as tennis, coaches must ensure both good defense and offense. Likewise, helping students learn physics well requires that instructors equip all students with both defensive and offensive strategies. Instructors can strengthen students' defenses by creating equitable and inclusive learning environments in which all students have a high sense of belonging, promoting and emphasizing a growth mindset, and ensuring that all students have a high physics self-efficacy and identity. Only if instructors help develop strong defenses in students pertaining to physics learning can they effectively engage with the offense by tackling challenging problems and developing physics problem solving, reasoning, and meta-cognitive skills. If instructors do not help students develop strong defenses, they are unlikely to risk struggling with challenging physics problems.

In the absence of physics instructors recognizing their role in centering students (which is particularly important for students from marginalized groups) and helping them build strong defenses, tackling challenging physics problems can make a student have the following type of worry: “I am struggling because I do not have what it takes to do well in physics. What is the point of even trying?” These kinds of negative thoughts can lead to a lack of engagement, even with effective approaches to learning physics, and can increase students’ anxiety. This, in turn, can start a detrimental feedback loop in which negative thoughts about struggling can lead to increased anxiety, procrastination, disengagement from effective learning approaches, and failure to take advantage of resources for learning. Moreover, anxiety can rob students of cognitive resources both while learning and taking tests. It can lead to deteriorated performance, which can then lead to further negative thoughts and anxiety. Having been bombarded by societal stereotypes and biases from a young age, women and ERM students are less likely than the students from the dominant group to have strong defenses when they enter physics classes. Therefore, if the instructor does not make a concerted effort to bolster student defenses and inoculate students against stereotype threats [73] (i.e., fear of confirming a negative stereotype about one’s group), the inequitable and non-inclusive learning environment is more likely to hurt women and ERM students. On the other hand, inculcating a growth mindset, i.e., intelligence is not immutable and one can excel in physics by working hard and working smart, can go a long way in helping all students engage effectively and benefit from research-validated tools and approaches [74, 75].

Consistent with our framework in Figure 1, instructors can empower students and strengthen their defenses by creating an inclusive and equitable learning environment. With proper coaching, all students can have a high sense of belonging, be unafraid to struggle and fail, and recognize that failures are normal and should be embraced since they are stepping-stones to learning [74, 75]. Although physics instructors have traditionally not considered it to be their responsibility to serve as coaches for their students (who focus on strengthening both their defensive and offensive skills), these issues are central to equity and inclusion in physics. Moreover, short classroom activities that take less than a class period at the beginning of the course can go a long way in improving students' sense of belonging and confidence, particularly for marginalized students who need the boost the most [74, 75]. To meet these objectives, it is critical to ensure that the learning environments in physics courses are equitable and inclusive. To ensure these equitable outcomes, physics instructors need to be trained in research-validated approaches to making the learning environments equitable and inclusive as well as in inclusive mentoring approaches. Physics instructors must also be given the opportunity to reflect upon and internalize the fact that students' physics interest and achievement goals are not fixed and can grow if the physics learning environments increase students' sense of belonging, self-efficacy, growth mindset, and physics identity in addition to helping them develop a good grasp of physics.

Finally, as shown in Figure 1, the HELPIEE framework emphasizes that equitable and inclusive learning entails that students' internal and external characteristics should be at the core of making decisions about what learning tools, curricula, and pedagogies would be effective, and equally importantly, how they should be implemented and what kinds of support students should be provided. Student internal characteristics include their prior preparation and physics beliefs (which are shaped by differential prior support from societal stakeholders and differential opportunities due to lack of resources in addition to biases and stereotypes, and lack of role models) and external characteristics include support of family members or other mentors/advisors and ability to balance the course work and outside work (e.g., many students must work to support themselves), among others.

While the majority of research studies reviewed in this chapter focus on the assessment of inequities based upon gaps in performance or motivational outcomes in undergraduate physics courses and the physics major as a whole using quantitative measures (e.g., how women and ERM students may be disadvantaged if the learning environment is not equitable or inclusive), some research studies discussed focus on qualitative investigations involving ethnographic class observations and individual interviews with students and instructors. We also review research on course-level interventions for making the physics learning environment equitable and inclusive and improving the experiences and achievements of the underrepresented groups [74, 76, 77].

The rest of the chapter is organized as follows. We first focus on research on equity issues in quantitative research at the level of physics major to get a big picture view of equity in undergraduate physics. Then we focus on quantitative research in physics courses, comparing the performance of the underrepresented groups and dominant group; followed by quantitative research on student motivational beliefs. In these contexts, we summarize findings from three types of courses: traditionally taught physics courses, research-based active-engagement courses with no special consideration for equity issues, and courses in which creating an equitable and inclusive learning environment is intentionally planned and incorporated into the course design. This is followed by research on stereotype threat since students from marginalized groups can be harmed by it. Then, we focus on qualitative research that helps us zoom in and obtain a deeper understanding of the experiences of women and ERM students and can be helpful in contemplating strategies to make physics learning environments equitable and inclusive. We close with a discussion of future directions and ongoing dialogues in the field.

2.2 Quantitative Research on Physics Major-Level Equity in Outcome

Research shows that women drop out of the physics major (and other STEM majors) with a significantly higher overall GPA than men and ERM students drop out of the physics major at a significantly higher rate than white students (65% of ERM students vs. 34% of white students) [78, 79]. Physics has one of the lowest enrollment as majors of all STEM disciplines, and student attrition after declaring a major is the highest from physics compared to other disciplines [78, 79]. Research also shows that physics is a STEM discipline that does not attract a second wave of majors in the third year and beyond, i.e., those who initially declare other STEM majors do not switch to physics in later years unlike most other STEM majors, which have a bi-directional flow of students in later years [79]. Moreover, it is important to keep in mind that this research on a documented unidirectional flow of physics majors out of physics [78, 79] does not include the attrition of students who were intending to major in physics in their first year (e.g., due to their better physics experiences in high school physics) but changed their mind after taking their college introductory physics courses even *before* declaring the physics major (since many students in US colleges declare the physics major at the end of their first year or in their second year).

In addition, students' experiences in introductory courses may affect their decisions in upper-level courses [80, 81]. In one study on introductory and advanced physics and math courses, gender differences in physics course performance (controlling for high school GPA) were only found in introductory physics courses [80]. Additionally, the introductory courses did not predict performance in advanced physics courses. Therefore, women could be making decisions about whether physics and related disciplines are the right fields to major for them based on the gender performance gap in the introductory physics courses. In another study, researchers investigated the student performance and persistence in upper level physics courses for students who experienced active learning (specifically, Modeling Instruction and the Investigative Science Learning Environment) in introductory physics courses [81]. In this study, women were more likely than men to graduate with a physics degree and they were just as likely as men to pass the upper-level courses. The highest risk of failure for women and men was in the first semester of upper division courses.

In another study investigating grades earned by students categorized by gender, ERM student status, low-income status, and first-generation college student status, ERM students experienced the largest penalty in their STEM GPA and overall GPA compared to non-ERM students [82]. While women had a higher overall GPA than men, that gender gap was either reduced or disappeared for students' STEM GPA depending on the demographic group. In addition, when students' STEM GPA was plotted by year for physical science (chemistry, computer science, engineering, mathematics, and physics) majors, ERM students experienced the largest penalty in their STEM GPA and in general students who were from intersecting marginalized demographic groups (for example low income and ERM) had larger penalties than students with the most advantages or students with another single demographic group disadvantage.

2.3 Quantitative Research on Course-Level Equity in Outcome: Performance

Early research on equity and inclusion in physics focused on documenting and/or reducing the gender gap in performance in physics courses. Some of the earlier work focused on the gender gaps on concept inventories like the Force Concept Inventory (FCI) and the Conceptual Survey of Electricity and Magnetism (CSEM) [9, 18, 19, 83-87]. Some of the research has shown that when high school backgrounds and pretest scores were matched, there was no difference in post-test physics performance [85]. Other research studies have found that particular questions on the FCI or CSEM may be biased against women [9, 84, 86, 88]. In addition, one study investigated gender differences in midterm and final exam scores and if those gender differences were correlated with final course grades for introductory physics courses [89]. Results show that performance on exams and final grades was weakly dependent on student gender [89]. Research suggests that gender differences in performance may be caused by societal stereotypes and biases about physics [85]. The gender differences in scores are most likely due to a combination of many factors rather than one factor that can be easily modified [18]. One study found that high school factors (including SAT scores, enrollment in calculus courses, and high school grades) predict student performance in introductory physics courses and that the pedagogy used in high school physics courses could differentially predict male and female students' performance [12].

A growing body of research has documented the gender gap when active learning pedagogies are implemented in the physics classroom. Lorenzo et. al and Coletta et al. found that interactive engagement pedagogy reduces the gender gap in performance in the physics course and on the FCI, respectively, and both women and men benefit from interactive engagement pedagogy [19, 83]. However, other studies have found that interactive engagement pedagogies do not reduce the gender gap and may worsen the gap [11, 13, 76, 90, 91]. That is, active engagement in an inequitable and non-inclusive learning environment can lead to inequitable outcomes, e.g., it can lead to an increased gender performance gap. For example, one study investigated the gender gap in an introductory physics 1 course on the FCI when Modeling Instruction (MI) was implemented as opposed to lecture-based courses [76]. Results show that while students in the MI course outperformed students in the lecture-based course, the gender gap in women's scores on the FCI increased by the end of the course. Another study investigated students' conceptual test performance in multiple introductory courses in which either partially or fully interactive classroom techniques were used (including student discussions on CoceptTests and Tutorials, among others) [90]. This study showed that there was a learning gap between male and female students in these courses and that in some instances, the gap increased by the end of the course. Additionally, in a study [92], the gender gap on the CSEM survey in calculus-based college introductory courses that primarily make use of lecture-based (LB) instruction (i.e., instructor lectured 90% or more of the class) was compared with courses that make significant use of evidence-based active-engagement (EBAE) strategies. The EBAE courses included flipped courses in which students watched lecture videos at home and answered a pre-lecture assignment before coming to class, and class time was used for clicker questions involving peer discussions and lecture demonstrations preceded by questions and collaborative problem solving in which students worked in groups of two to three on quantitative problems. Research shows that the gender gap on the CSEM remains relatively constant in LB courses (4% at the beginning and 6% at the end), and both male and female students exhibit similar normalized gains on the CSEM (18% and 22% for female and male students, respectively). For EBAE courses, the gender gap increases significantly from 4% to 10%, which is also reflected in the effect sizes comparing male and female student CSEM performance: 0.27 on the pretest and 0.54 on the posttest, i.e., the effect size doubles. Also, the data suggest that while both men and women benefit from EBAE instruction (larger normalized gains in EBAE courses compared to LB courses), male students benefit disproportionately more than female students: the normalized gain for male students in EBAE courses is 39% compared to only 28% for female students. Normalized gain is defined as (post%-pre%)/(100%-pre%).

One study suggests that the gender gap in interactive physics courses may be due to differences in prior physics and math knowledge as well as differences in attitudes and beliefs about physics [13]. However, in accordance with our HELPIEE framework, the instructors must take responsibility and strive to make the physics learning environments equitable and inclusive so that all students, regardless of their prior knowledge and beliefs (which are often the result of stereotypes, biases, and differential opportunities) can excel in physics courses.

While there have been many research studies focused on gender issues in physics, fewer studies have focused on race/ethnicity in physics. Similar to the gender gap in performance, ERM students tend to have lower performance in introductory physics courses overall and on conceptual tests than their white peers [76, 93-95]. This is the case even in classes that use active engagement. In one study, ethnic and racial minority students entered the introductory physics courses with lower conceptual understanding, and the gap was maintained to the end of the introductory physics course sequence when active learning (specifically Modeling Instruction) was implemented [76]. Other studies show that part of the gap may be explained by differences in prior preparation. For instance, one study showed that the gaps in ethnic and racial minority students' final exam scores disappeared when students' ACT/SAT math scores and pre test scores on a conceptual test were controlled for [95]. In addition, in another study, ERM students experienced a larger penalty in their STEM GPA than non-ERM students, and ERM students with additional disadvantages due to socioeconomic status or first-generation college status were further penalized in their average GPA [82]. Stewart et al. investigated physics performance differences in relation to gender, ERM status, and status as first-generation college students (FGCS) [96]. Results showed significant differences in gender, ERM status, and FGCS status on the pre and post conceptual test and significant differences were found in students' course grades for ERM and FGCS students [96]. In addition, path analysis was used to examine the relationship between demographic factors, academic factors (SAT math scores, and pre conceptual test scores), and course performance (post conceptual tests and course grades). For ERM and FGCS students, differences on their pre conceptual test scores and ACT math scores explained most of the difference in course achievement scores [96].

2.4 Quantitative Research on Course-Level Equity in Outcome: Motivational Beliefs

Starting around 2010, researchers began to investigate motivational beliefs of students including their physics identity and self-efficacy. Early work investigated gender gaps in these motivational constructs, especially in students' self-efficacy [30, 31]. Self-efficacy is one’s belief that they can succeed in a particular activity or course [27, 28]. Self-efficacy is an important motivational construct since it is one of the primary dimensions of physics identity and it influences students’ engagement, learning, and persistence in science courses [30, 97]. Sawtelle et al. investigated sources of students' physics self-efficacy by gender and found subtle distinctions in the predictive ability of the sources of self-efficacy in women and men [31]. According to Badura’s social cognitive theory, self-efficacy may be derived from four sources: mastery experiences, vicarious learning experiences, social persuasion experiences, and physiological state [27]. Mastery experiences are important because successful completion of a task should have a strong positive influence on one's confidence to complete a similar task. Vicarious learning experiences occur when observing someone else’s success on a task influences their own belief in their ability to perform a similar task. Social persuasion experiences are that verbal suggestions from others such as words of encouragement can result in an increase in one’s self-efficacy. Lastly, one's physiological state can act as a mediating source to amplify or undermine one's confidence in one's ability. Sawtelle et. al. found that the probability of passing an introductory physics course for women relies primarily on the vicarious learning experiences source while it relies on mastery experiences for men [31].

In general, self-efficacy of both men and women decrease throughout introductory physics courses and self-efficacy decreases more for women than men [43]. This could be detrimental to women since self-efficacy and test anxiety may be related to each other [98]. In one study, women had lower self-efficacy and higher test anxiety than men, and self-efficacy mediated the relationship between test anxiety and high-stakes assessment test scores [98]. The gender gap in self-efficacy is found in both traditional and most interactive engagement courses [29]. However, one study found that in a class that employed team and project based physics learning the gender gap in self-efficacy disappeared by the end of a physics class [99]. In general, this gender gap in self-efficacy persists for men and women even when controlling for students' performance in introductory calculus-based physics courses [43]. Specifically, women who received A's in the course had the same self-efficacy as men who received C's in the course [43]. For engineering students, the gender gap in physics self-efficacy does not close by their fourth year while it closes in other STEM subjects [100]. Self-efficacy may predict students' future careers in physics as well [101]. In particular, in a study of Finnish undergraduate students, women had lower self-efficacy and higher anxiety about physics than male students, and their self-efficacy predicted student goals of going to graduate school [101].

Around the same time physics self-efficacy was investigated, researchers also explored students' physics identity. Physics identity is defined as identifying with physics, i.e., whether students see themselves as physics people [2, 5] or those who can excel in physics. In 2010, Hazari et al. took advantage of Carlone and Johnson's framework for science identity [55] to formulate a framework of physics identity and began to investigate high school students' physics identity and college students' engineering identity [5]. In Hazari's framework, "competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a good physics student. Lastly, a fourth dimension, interest, was added to the framework since students can have highly varying levels of interest in physics [60, 61]. In more recent studies by Hazari et al. of introductory students, performance and competence were combined into one variable [58]. In a slightly reframed version of Hazari et al.'s physics identity framework by Kalender et al. [59], performance/competence was framed as self-efficacy, which is closely related to competency belief and recognition was renamed as "perceived recognition" for clarity to investigate introductory students' physics identity [59]. Hyater Adams et al. combined the frameworks of physics identity and racialized identity to create a Critical Physics Identity framework to understand the experiences of Black students in physics [102].

Our prior individual interviews suggest that students' perceived recognition by instructors and teaching assistants (TAs) impacts their self-efficacy and interest in physics (e.g., see [52, 53]). In addition, in order for students to feel validated and recognized by their instructors, instructors need to both explicitly recognize students (for example, by verbally acknowledging the progress and success of their students) and implicitly recognize students (for example, by setting high standards for all of their students and making it clear to students that they all have what it takes to excel if they work hard and work smart and take advantage of all of the resources) [103]. However, studies have shown that without intentional strategies to create an equitable and inclusive learning environment, female students do not feel recognized appropriately even before they enter college

[1, 4], which is at least partly due to the societal stereotypes and biases about who belongs in physics and who can excel in it that women are bombarded with over their lifetime.

Students' sense of belonging in physics courses has not been studied as extensively. However, it has been shown that students' sense of belonging in calculus-based introductory physics courses for physical science and engineering majors is so closely tied to their self-efficacy that it was difficult to separate them as distinct factors in factor analysis [4]. In addition, it is shown to be a predictor of students' physics identity for senior physics majors [58], and a predictor of students' grades in an introductory physics 1 course for students on the bioscience track [104]. Another study showed that female participants in a Physics Olympiad competition who endorsed negative stereotypes about female talent for physics felt a lower sense of belonging in physics [105]. One study investigated factors that affect women's sense of belonging in physics that include the lack of role models in physics and stereotypes about who can succeed in physics [106]. Some suggestions for instructors to improve women's sense of belonging in physics include sending messages that concerns about belonging are normal and fade with time, identifying and tempering cues that perpetuate the "geeky" scientist stereotype, and openly endorsing effort and hard work over brilliance [106].

Additionally, although general mindset research has been ongoing for many decades [107], mindset applied to college physics has started only recently to become more widely studied in the physics discipline [108, 109]. One study in an introductory calculus-based physics course laid out a framework in which there are four distinct mindset views (i.e., whether students' ability is fixed or malleable, whether their intelligence can grow from effort, and similarly, whether that view is about themselves or for the general population) and investigated how they predict physics course grade [108]. The researchers found that women were more likely to believe that innate talent was needed for them to excel in physics and they might not be gifted. In addition, mindset was a stronger predictor of physics grade for students’ malleable mindset views about themselves (“my ability”) than the other mindset groups. In another study in an introductory calculus-based physics course, there were only gender differences in students' “my ability” mindset group at the beginning of the course [109]. However, gender differences developed in each mindset view by the end of the course, and the gender difference in the “my ability” category increased over the course. In addition, “my ability” was the only mindset factor that predicted course grades.

While most of the research studies on motivational beliefs have been conducted in calculus-based introductory physics classes in which women are severely underrepresented, similar findings have been reported in algebra-based physics courses for students interested in health professions in which women are not underrepresented. Although women are not underrepresented in these physics courses, societal stereotypes and biases internalized by female students over their lifetime can still impact their motivational beliefs about physics. In studies of introductory physics courses for bioscience majors, women had lower motivational beliefs than men. Although women outnumber men in these introductory physics courses, women had lower self-efficacy than men controlling for the grade they received in the physics courses. In particular, women who received an A in the physics 1 and 2 courses (defined as A+, A or A-) had similar self-efficacy to men who received a B (defined as B+, B, or B-) [110]. This trend is similar to the trends in calculus-based introductory physics courses in which women with A grades have the same self-efficacy as men with C grades [43]. However, instructors may have the ability to help improve students' motivational beliefs in the introductory courses. In studies conducted in calculus-based introductory physics courses and introductory physics courses for bioscience majors, students' perception of the inclusiveness of the learning environment factors (consisting of recognition by others including physics course instructors and teaching assistants, interaction with their peers, and sense of belonging) predicted gender differences in students' motivational outcomes at the end of the physics course such as self-efficacy, interest, and physics identity [62, 111].

In addition, some research has focused on students' motivational beliefs in the lab context. One study investigated students' two different group work styles in introductory lab courses; Group A in which each student takes on a different task, but spends equal time on it, and Group B in which students divide the work equitably and each student participates in every aspect of the work [112]. The findings show that while students prefer Group A style work, students who participated in Group B style work were more likely to report that interacting with their peers increased their physics interest and women were more likely to report that peer interactions increased their self-efficacy [112]. In addition, in another study in the lab context, an online, hands-on laboratory option was implemented to investigate gender differences in students' epistemological beliefs, socialization, and help-seeking in the laboratory [113]. Results show that men had higher epistemological beliefs, women reported a greater willingness to seek assistance from instructors and peers, and there was no difference by gender in socialization. The students in the in-person labs placed a higher value on socialization in the lab while there was no difference in epistemological beliefs or help-seeking. However, when comparing male and female students in each lab type, there were no significant gender differences in the three factors.

Several factors have been proposed to explain the gender difference in students' motivational beliefs (including self-efficacy and identity) and on conceptual tests. Some studies suggest that differences in prior preparation for various reasons may account for the gender difference [13, 114]. In addition, societal stereotypes and biases about who belongs in physics can negatively impact women in physics courses [115]. Some of the elements of the environment in many science classrooms that can negatively impact women include a lack of female role models, pedagogy that favors male students' interests, and a "chilly climate" for women [25].

In addition, the TEAM-UP report lists five factors essential to improving African American students' persistence and success in physics including their belonging, physics identity, academic support, personal support, and leadership and structures [64]. In order to improve African American students' sense of belonging, the report includes suggestions such as establishing clear rules in common spaces to ensure everyone is welcome, assisting students in finding the support they need inside and outside the department, and consistently communicating norms and values of respect and inclusion. To improve physics identity, suggestions to physics departments include diversifying their faculty with respect to race/ethnicity/gender, emphasizing the ways a physics degree empowers graduates to improve society, and discussing a broad range of career options with undergraduate students. In addition, to increase the other three factors, the report suggests that departments develop evidence-based actionable plans to increase the persistence of all students to physics degrees, help students take advantage of campus resources for funding such as conference travel, and set norms of inclusion and belonging in the department.

2.5 Stereotype Threat

Stereotype threat is the anxiety associated with confirming a negative stereotype resulting in reduced performance for the stereotyped group [116]. Prior studies in the context of mathematics have found that activation of a negative stereotype about a group or stereotype threat, e.g., asking test takers to indicate their ethnicity before taking a test, can lead to a deteriorated performance of the stereotyped group. For example, in the high school physics context, Marchand and

Taasoobshirazi [117] conducted research that suggests that a stereotype threat is automatically triggered in a physics test-taking situation due to prevalent societal stereotypes. They used three different manipulations immediately before students took a four question quantitative physics test: (i) an explicit, (ii) an implicit, and (iii) a nullified stereotype threat condition in which students were either told that (i) female students had performed worse than male students on this test, (ii) not told anything, or (iii) told that the test had been found to be gender neutral. While male students performed similarly in all three conditions, female students in the explicit and implicit stereotype threat conditions had comparable performances but performed statistically significantly worse than female students in the nullified condition. The researchers interpreted this result to suggest that a stereotype threat is automatically triggered in a test-taking situation.

In the context of college physics courses, stereotype threat has been shown to impact student performance on exams in some studies, but not all. In one study [118], the pretest and post-test performance of female and male students on the Conceptual Survey of Electricity and Magnetism (CSEM) was analyzed in an introductory algebra-based course in these two conditions: students were or were not asked to provide gender information before taking the CSEM (gender salient or not salient condition, respectively. There were no statistically significant differences between the performance of male or female students under the two conditions (e.g., female students who wrote their gender before taking the CSEM did not perform worse than female students who wrote their gender after taking the CSEM) in the pretest or the post-test.

In the same study [118], investigators also focused on stereotype threat associated with gender stereotypes in physics and its impact on student performance on the CSEM in calculus-based college introductory courses taken by engineering, physical science, and mathematics majors in which female students are severely underrepresented,. In particular, the study investigated the extent to which agreeing with a gender stereotype (i.e., I expect men to generally perform better in physics than women) correlates with performance on the CSEM. The demographic information of students in these courses was: gender – 67% male and 33% female, race – 77% White, 11% Asian, 4% Latinx, 4% Multiracial, 3% Black, and 1% Other. It was found that women who agree with the gender stereotype performed statistically significantly worse than women who did not agree with the gender stereotype at the end of a year-long calculus-based physics course sequence even though there was no difference in their performance in the pretest given at the beginning of the course [118]. Cognitive science suggests that the anxiety associated with conforming to a stereotype is essentially a threat, and it can take up part of the working memory, thus robbing an individual of cognitive resources that could be used for problem solving and learning. Moreover, the anxiety can also lead to procrastination or less time spent on learning as well as reduced engagement and use of effective study strategies and asking for help. Prior research has suggested that a certain level of stereotype threat may be implicitly present for female students in an introductory physics course. These findings suggest that female students endorsing a gender stereotype may be undergoing additional stereotype threats over and above what might already be present for many women in physics courses. As shown by the research [118], over the course of the semester, this can have a significant negative impact, especially in a calculus-based course in which women are underrepresented.

2.6 Qualitative Methods Research

Qualitative research can be a powerful tool for unpacking the mechanisms underlying quantitative research discussed in the preceding sections and understanding the experiences of students (especially women and ERM students we focus on here). Therefore, there has also been research studies focused on qualitative methods, e.g., interviews with women and ERM students to understand their experiences in physics. One study focused on the masculine nature of doing physics in a variety of contexts [119] and how it negatively impacts women. Some of the research focuses on students' mindset and how mindset beliefs studied in the literature in other disciplines are not always consistent with challenges in college physics courses [120]. Another study investigated the interviews of five women physics students to understand their identity as physicists [121]. The study showed that some of the women relate to the masculine norms of the discipline. For example, one of the woman mentioned that she was likely to tinker with lab equipment and referring to herself as “laddish” which made it easier for her than other women to fit within the boundaries of physicists. Some of the women also dealt with norms and expectations about how women are supposed to behave in a physics discipline like being more likely to take on the secretarial role in physics labs. Other studies have focused on the lab setting, using interviews and ethnographic observations to understand gendered task division and how women often have to position themselves as secretaries or project managers in the labs and do gender and physics simultaneously [52, 122].

Other work on race and ethnicity utilizes a critical race theory lens to interview women of color to understand their experiences through the physics curriculum [123-125]. For example, in one study, six Black women were interviewed to address obstacles faced in their career paths, such as socialization in STEM [125]. Many of the women had similar experiences in that they were influenced to major in physics due to being exposed to a science environment at a young age from after-school or summer school programs and were able to do physics research over the summer in their undergraduate studies. However, most of the women experienced isolation in their graduate studies, particularly in study groups.   Strategies used to overcome the obstacles faced in choosing physics over other STEM fields and in their career included afterschool activities that focus on scientific practices in high school, college recruitment specifically targeting underrepresented groups, financial aid, and creating an inclusive and supporting environment [125]. In another study, ten women of color were interviewed to understand how their sense of belonging and competence are questioned due to the intersection in three realms: their field of study (physics), gender, and race/ethnicity [124]. It was concluded that in order to retain more women of color in the field, physics departments should focus on reforming hiring and recruiting policies to recruit more women and ERM students, structurally and financially supporting their membership in organizations (such as the formal groups for women and ERM students), and to create a more hospitable environment for all students by improving the pedagogical, social and cultural practices that attract potential science majors [124].

In another study, women of color and lesbian, gay bisexual, transgender, or queer women were interviewed about their identity as a physicist [123]. When asked to describe a physicist, the women mentioned common stereotypes about who can do physics (such as being a white, male, genius to succeed in physics) that have prevented them from identifying as a physicist and questioned the necessity of a formal degree to be a physicist. However, most of the women were able to eventually reject the common stereotype and self-identify as a physicist and thus came to accept that others saw them as a physicist as well. The study highlighted the importance of personally identifying as a physicist for empowerment and belonging and the importance of recognition by others as a component of identity [123]. In addition, some of the research has focused on graduate students. For example, one study investigated student responses on an application question to be a part of the APS Bridge program and interviewed 9 students who were accepted into the program to understand the barriers that ERM students face when applying to graduate school [126]. Findings showed that some of the barriers that ERM students face include the Graduate Record Exams (GRE), student research experience, student grades/GPA, deadlines for applying to physics graduate programs, and financial concerns [126].

In order to support equity and inclusion for intersectional identity in academic settings, such as in physics classrooms, the Intersectionally Conscious Collaboration (ICC) protocol was created [127]. The ICC protocol includes six elements that allow physics educators to locate and address biases in pedagogical practices and better design learning experiences that engage and motivate all students.

Another study investigated a physics department in which women and women of color feel successful, and that they belong to understand their physics identity in that setting [128]. The study consisted of interviews with students and faculty members and the researcher attending physics classes. Important components that promoted these women of color's physics identity included students working collaboratively together, physics faculty members taking responsibility for group work to go smoothly, protecting students from racist and sexist microaggressions, and believing that success in physics is a result of hard work instead of innate intelligence [128].

2.7 Interventions

In addition to research into assessing equity and inclusion in undergraduate physics, there has also been research into implementing changes in the classroom to make them more equitable and inclusive. Socio-psychological classroom interventions, e.g., those focusing on self-affirmation or sense of belonging and mindset interventions have been shown to improve the experiences of women in physics courses [74, 75]. One intervention, called values affirmation, done in an introductory physics course, involved students writing about their most important values (such as connections with friends or family) for 15 minutes twice at the beginning of the course (once during the first recitation of the semester and once for online homework shortly before the first midterm) [77]. Values affirmation can buffer people against psychological threats, such as the stereotype that men are better than women at math and science. At the end of the course, the gender gap in performance was reduced and benefited women who tended to endorse the stereotype that men do better than women in physics.

Some other interventions are implemented at the beginning of the semester in introductory physics courses to create an equitable and inclusive learning environment in which students from marginalized groups have a high sense of belonging and feel that it is safe to engage in collaboration and discussions with peers and instructors. With this type of short intervention, the performance gap between the underrepresented and dominant groups can sometimes be significantly reduced [74]. For example, a short ecological belonging/mindset intervention by Binning et al., which only requires half of a single recitation class period at the beginning of the semester, eliminated the gender gap in physics performance [74]. In addition, ERM students performed better in the intervention condition than in the control condition. However, the intervention did not statistically eliminate the gap between white and ERM students, potentially due to the low numbers of ERM students.

This type of intervention [74] was conducted in a required introductory calculus-based physics course, which is taken by physical science and engineering majors typically in their first year and their first semester in college. Two female physics graduate students were trained to facilitate the half-hour activity at the beginning of the semester in half of the recitations that were randomly selected. The facilitators introduced it as an activity that would help the physics department understand student concerns and how to foster better learning environments. At the beginning of the first recitation class in which the activity took place, students were handed a piece of paper and asked to write down their concerns about being in the physics course. Then they were shown some quotations from both male and female students from previous years who did very well in physics but also had similar concerns. The quotes emphasized the importance of working hard and working smart, learning from one’s mistakes, and taking advantage of all learning resources available to them because that is the way to perform well in physics. Then students were asked to get together in small groups to discuss what they wrote; during this session, they generally learned that their peers in the class had similar worries. A general class discussion followed in which the groups summarized their discussions, with explicit emphasis on the fact that adversity is common in college physics courses, but it is temporary. The facilitators re-emphasized that students should embrace challenging physics problems and use their failures as bridges to learning. Finally, using the principle of “saying is believing”, students were asked to write a short letter telling a future student about strategies for excelling in their physics classes. What is heartening is that this short intervention fully closed the gender performance gap and greatly reduced the gap between ethnic/racial minority and white students in performance compared to the group in which this short intervention did not take place. These types of interventions are productive because students’ sense of belonging and other motivational beliefs are strongly intertwined with feeling safe, increased cognitive engagement, reduced anxiety, and learning.

In order for these types of interventions to succeed, however, a variety of factors must be considered and carefully implemented [75]. Some of the elements that must be considered are that the intervention must deal with specific concerns students have, the message should be delivered without singling out any particular group, the intervention must use methods that psychologists have found to be long lasting, and they should not be framed explicitly as interventions but activities that are components of the course. In addition, there are multiple strategies for improving women’s participation in physics including creating a positive learning atmosphere, providing encouragement and support to women, and emphasizing the societal benefits of physics [129].

Besides interventions, another way instructors can improve the learning environment in their courses is by providing mentoring and support for underrepresented students [64]. Instructors can set an equity goal for their class to explicitly track whether the demographic differences in their courses are getting better. Additionally, the mindset of the instructor in a course also plays a pivotal role in predicting student achievement. In a study of 150 STEM instructors by Canning et al., courses taught by instructors with a fixed mindset had twice as large an achievement gap as courses taught by instructors with a growth mindset [130]. Only if the instructors themselves have a growth mindset about their students’ ability can they come across as trustworthy to their students. Then they can credibly and authentically emphasize a growth mindset to their students, that physics is mastered through hard work and deliberate practice and not through innate talent or genius [131].

2.8 Critiques, Ongoing Dialogues, and Future Directions

One critique of most of the current work on gender in physics is that gender is put into binary categories, when it is more complicated and on a spectrum. The recommendation is to have more qualitative research, based on a methodology that attends to the complexity of gender (among others) [132]. Most researchers conducting this type of research now acknowledge that gender is not a binary construct but since most of the data from students are provided by the university in a binary form, it is often used that way, particularly in quantitative research.

Another critique is that the gender gap in performance is framed as a comparison of women with men which could be interpreted as “women should be more like men”, which is a deficit model [132]. A similar critique has been put forward for the gap between white students and ERM students. The recommendation is to move beyond the “gap” framework (among others) [132].

Regarding comparing women with men or ERM students with white students in quantitative comparisons in physics courses that show gaps, these kinds of studies are important for revealing inequities in the physics learning environments. In particular, the framework of the research plays a pivotal role in whether the model is a student deficit model or a course deficit model (i.e., the learning environment is not equitable and inclusive and is disadvantaging the underrepresented groups). These inequities are typically caused by differential opportunities for students based upon their privilege and societal stereotypes and biases about who can succeed in physics that can accumulate over the students’ lifetime. For instance, even before stepping into undergraduate physics courses discussed here, throughout K-12 education, women and ERM students are often not treated the same way as the white male students in physical science courses by their teachers and high school counselors also give them differential advice. Even TV shows like The Big Bang Theory as well as the interactions of students in museums with adults perpetuate the stereotypes based on gender, ethnicity, and race [133]. Moreover, most famous physicists are white men and so women and ERM students do not have as many role models to show them that they can be successful in physics courses. It is not surprising then that in an inequitable and non-inclusive learning environment, even in introductory algebra-based physics courses for bioscience majors in which women are not underrepresented, women have lower motivational beliefs at the beginning of the course due to the societal stereotypes and biases [134]. It is important to center instruction in the physics courses in ways that focus on creating an equitable and inclusive learning environment and counter the impact of these pervasive societal stereotypes and biases.

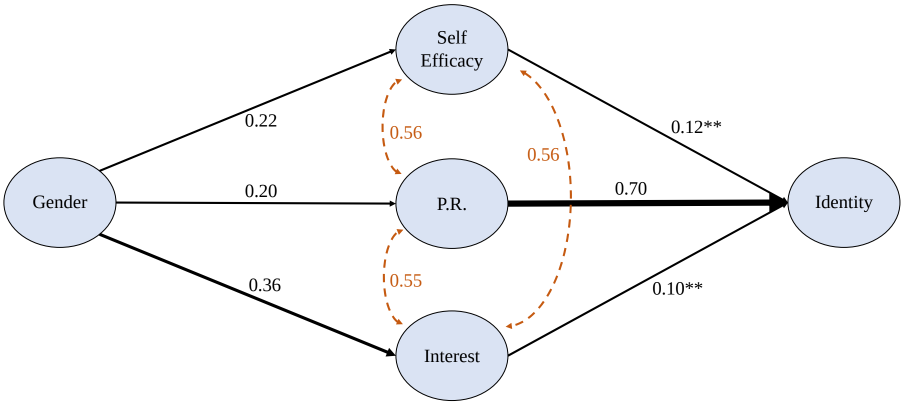
Another critique is on the narrow scope of the demographics in physics education research [135]. In particular, research disproportionately focuses on students in introductory calculus-based courses and at institutions that have a smaller population of ERM students than the overall college-bound population. In addition, there is less research on high school students, students at two-year colleges, and racially diverse colleges, e.g., minority serving institutions (MSI). Therefore, future work in physics education should include a wider variety of students and institutions.

Additionally, the connections between observational data and implied causal connections between factors in statistical approaches involving regression in PER have been criticized. In particular, to select good models from several statistically equivalent ones [136] researchers should consider different aspects of the models when selecting the best model. First, the potential instructional implications of each model should be considered, i.e., whether these instructional implications will have a positive influence on instructors and their pedagogical approaches. When a model is framed in this way, it can empower instructors so that they adopt effective practices and understand their role in recognizing and empowering students and affirming their work. Second, researchers should consider whether the instructionally beneficial models are also supported by additional evidence. This evidence may include but is not limited to researchers' own interview data or findings from prior studies. In other words, researchers should generate at least a few substantively meaningful different equivalent versions and deliberate based upon instructional implications and other evidence for why the proposed model is better than the others [137].

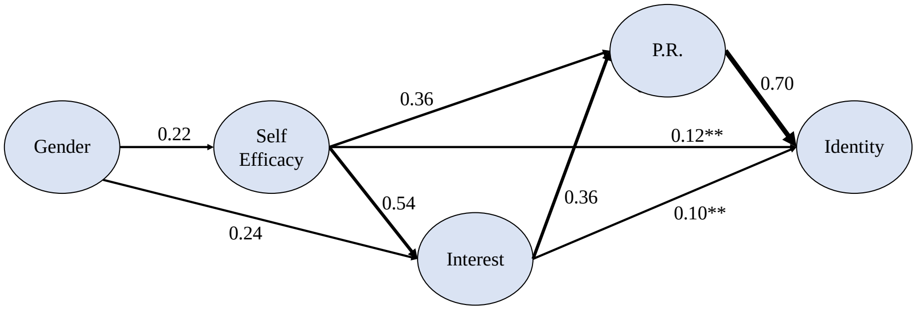
Here we illustrate this point with quantitative data from a motivational survey administered to students in the second introductory physics course for bioscience majors. We use the physics identity framework (as explained in section IV) as an example to show how the proposed approach is used to select a good model from several statistically equivalent ones. To quantify the significance and relative strength of our framework links, we used Structural Equation Modeling (SEM) [138]. The models predicted students' physics identity through self-efficacy, interest, and perceived recognition. While each model is statistically equivalent, the instructional implications of each model are different. We initially tested gender moderation between different constructs using multi-group SEM (between male and female students) to investigate whether the relationship between the different motivational constructs was different across gender. There were no group differences at the level of weak and strong measurement invariance and the level of regression coefficients. Therefore, we proceeded to gender mediation analysis to understand how self-efficacy, interest, and perceived recognition mediate the effect of gender on physics identity at the end of the second introductory physics course for bioscience majors. There are 27 statistically equivalent models with different predictive relations between the three mediating constructs (self-efficacy, interest, and perceived recognition) [136]. Here we discuss four of the models to show how our framework could guide us to select a good model. In all models, the model fit indices indicate a good fit to the data [139]. All path analysis results of the models are shown in Figure 2. More details about each of the models can be found in Appendix A.

First, we consider Model 1 in which there is no predictive relationship between self-efficacy, interest, and perceived recognition. Instead, there are covariances between each motivational factor. Figure 2 (a) shows the path analysis results of this SEM model. Next, we consider Model 2 where self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition (see Figure 2 (b)). In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition (see Figure 2 (c)). Finally, in Model 4 perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest (see Figure 2 (d)).

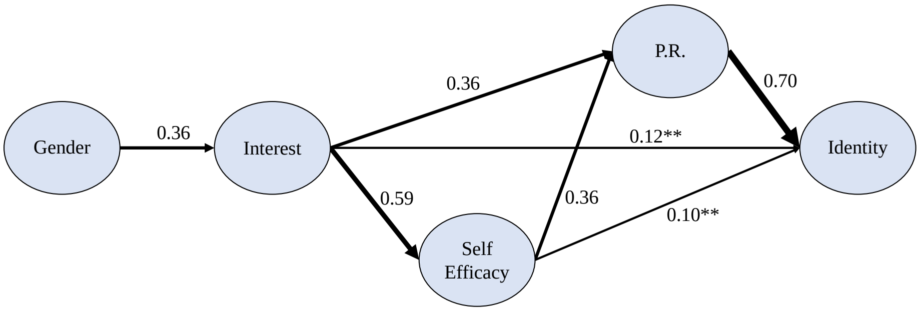
(a) Model 1



(b) Model 2



(c) Model 3



(d) Model 4

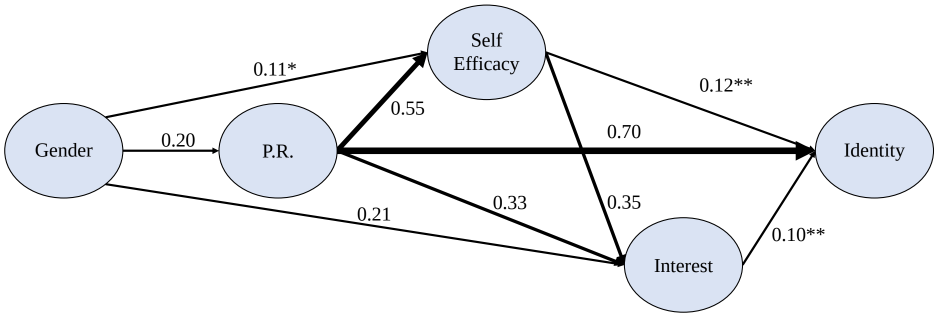


Figure 2 Results of the path analysis part of SEM Model 1-4 show how the relationship between gender and physics identity is mediated through self-efficacy, interest, and perceived recognition (P.R.). (a) In Model 1, there are only covariances between each pair of constructs: self-efficacy, P.R., and interest. (b) In Model 2, self-efficacy predicts interest and P.R., and interest predicts P.R. (c) In Model 3, interest predicts self-efficacy and P.R., and self-efficacy predicts P.R. (d) In Model 4, P.R. predicts interest and self-efficacy, and self-efficacy predicts 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 (β values), which represent the strength of the regression relations. Each regression line thickness qualitatively corresponds to the magnitude of the β value. Regression coefficients with $p < 0.050$ are indicated by superscript ***, $p < 0.010$ are indicated by superscript **, and $p \leq 0.001$ are indicated by no superscript. For clarity, we have removed the statistically insignificant regression path from gender to identity.

Since all 4 models in Figure 2 are statistically equivalent, according to the theoretical framework we now must consider the instructional implications of each model. In Model 1, self-efficacy and interest are only predicted by gender. However, self-efficacy and interest may be considered fixed by physics educators and instructional policy makers. Model 1 does not provide suggestions that instructors can use to improve those motivational factors. Since perceived recognition, interest, and self-efficacy covary, it can be difficult for instructors to interpret how the motivational factors relate to one another. In addition, in Models 2 and 3, gender only predicts self-efficacy or interest respectively which could be interpreted as a deficit model. In particular, these equivalent models can be interpreted to imply, that women are not feeling positively recognized by their instructors and teaching assistants (TAs) as much as men because they have lower interest and self-efficacy than men. While statistically equivalent to Models 1-3, Model 4 with perceived recognition predicting self-efficacy and interest is more likely to give them the message that students' interest and self-efficacy in physics can be influenced by the recognition they receive from instructors and teaching assistants. Thus, Model 4 is also more likely to inspire instructors to create a more inclusive and equitable learning environment in which all students including those from the underrepresented groups feel more positively recognized and affirmed.

Next, we must assess whether the instructionally beneficial models are also supported by additional evidence. Model 4 also reflects findings from prior interviews. The interviews show that recognition by others, especially from instructors or TAs, is critical in shaping students' self-efficacy and interest [52, 53, 140, 141]. Thus, we argue that Model 4 (in which physics self-efficacy, interest, and perceived recognition mediate the relation between gender and physics identity) is the best statistically equivalent model in Figure 2 and Model 4 is better than others based on our theoretical framework focusing on the model's instructional implications and supporting evidence from individual interviews with students.

In conclusion, to promote equity and inclusion, there is urgent need to dismantle inequitable structures and create an equitable and inclusive learning environment. Our conception of equitable outcomes discussed earlier emphasizes that all demographic groups should have comparable outcomes. Taking inspiration from prior studies, an effective approach to creating an equitable and inclusive learning environment humanizes learning and takes advantage of student assets using a culturally responsive pedagogy instead of using a deficit view of students [142]. In particular, institutions should recognize their duty to take action to create an equitable and inclusive learning environment and encourage the use of pedagogy in which all students have high motivational beliefs and can participate fully without the fear of being judged. It can also be beneficial for instructors to set an equity goal for their classes to explicitly track whether the demographic differences in outcomes in their courses are vanishing. In addition, more research is needed at the intersection of various demographic factors [143, 144], and diverse selections of schools such as 4-year institutions, community colleges, minority serving institutions, and all-Women's colleges.

3.0 Damage caused by societal stereotypes: Women have lower physics self-efficacy controlling for grade even in courses in which they outnumber men

3.1 Introduction

Recently, there has been significant focus on the underrepresentation of women in many science, technology, engineering, and math (STEM) fields in order to increase their representation and participation in these fields [115, 145-154]. Some of these research studies have focused on comparison of degree attainment by women and men or tracked processes of change in STEM undergraduate education [145-147]. Others focus on how women are still marginalized in these disciplines and strategies to improve the situation based upon a variety of research findings [115, 148-150], or ways to improve the retention of women via classroom interventions [151, 152]. Yet other research studies have focused on the root cause of women and other undergraduates leaving STEM majors [153, 154]. Motivational beliefs such as identity and self-efficacy in different STEM domains can influence students' continuation in related STEM courses, majors and careers [26, 32, 39, 41, 54, 155-159]. For example, students were more likely to take courses or pursue a career in science if they had higher competency belief or self-efficacy [32, 39, 41, 155, 156], display higher interest in science [157], or have a higher science identity [54, 158, 159]. Moreover, a gender gap favoring men in motivational factors [26, 97, 160] and conceptual tests [84] has been studied in STEM courses. Specifically, in the context of physics, several studies in the courses in which women are underrepresented have focused on understanding and addressing the low representation of women in these courses [18]. However, few studies on self-efficacy have focused on introductory algebra-based physics courses for biology and health science (bioscience) majors in which women tend to outnumber men. In a review paper that summarizes gender differences in physics courses by Madsen et. al, there is an explicit call to investigate the gender gap in various construct in physics courses where women are not underrepresented. [18]

Self-efficacy is a person's belief that they can succeed in a particular activity or course [27, 28]. A students' self-efficacy in academic courses may be influenced by the classroom environment [37, 161, 162] and different teaching strategies [29, 34, 38]. In addition, self-efficacy has been shown to impact students' engagement, learning, and achievement in science courses [14, 31, 36, 40-42, 163, 164]. For example, students' higher self-efficacy was found to correlate with higher course achievement [14, 31, 40, 163]. In addition, in other studies, students with higher self-efficacy were more likely to persist on more challenging tasks [36], less likely to second guess themselves and switch to an incorrect answer [164], and persist in STEM courses and careers [41, 42]. In addition, self-efficacy is one of the factors that contributes to students' science identity. Self-efficacy contributing to a students' identity is important because identity in a particular domain has been shown to play an important role not only in students' in-class participation and performance but also in their choices of future courses and careers [5, 49, 55-57, 165]. Additionally, identity frameworks show that identity is predicted by students' self-efficacy, interest, and recognition by others in physics [3, 5, 59].

Students with higher self-efficacy are more likely to embrace challenging goals, take on difficult tasks, and use effective learning strategies in their classes [33]. Additionally, students' self-efficacy in different STEM courses can lead to long term effects by forcing them into a positive or negative feedback loop. When students complete short-term goals as they work towards completing their major (e.g., finishing introductory physics courses, etc.) they receive feedback on their performance. Students with high self-efficacy in a domain are less likely to have anxiety that can rob them of their cognitive resources while learning and test taking, since the working memory during problem solving has limited capacity [166]. They are also more likely to employ better academic strategies, such as goal-setting, self-monitoring, and self-evaluation and their self-efficacy can be a predictor of their performance in that domain [33]. Thus, the domain specific self-efficacy has the potential to create a feedback loop with performance, e.g., when students with low or high self-efficacy start receiving feedback on graded assignments and exams in their classes. In particular, a student with greater self-efficacy may perform better (than a student with similar knowledge but lower self-efficacy), which can strengthen their self-efficacy further. On the other hand, a student with lower self-efficacy may perform worse, which can further weaken their self-efficacy. Since self-efficacy can increase students' persistence in science majors and careers [167] and a gender gap in self-efficacy has been observed in physics courses in which women are underrepresented [26, 30, 43, 168], it is important to investigate whether there is a gender gap in physics self-efficacy in introductory physics courses in which women are the majority.

Even though women outnumber men in the introductory algebra-based physics courses for bioscience majors, pervasive societal stereotypes and biases about who belongs and can succeed in physics that women internalize from a young age can impact women's physics self-efficacy. One common stereotype is that one needs to be a genius or brilliant to succeed in physics [23]. Genius is often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius [24]. Additionally, by high school, girls are less likely to believe that physics is for them and that their teacher thinks they are good at physics [1]. Furthermore, the environment in many science classrooms, including a pedagogy that favors male students' interests, a "chilly climate" for women, and a lack of female role models can negatively impact women [25]. All these factors can influence female students' perception of their ability to do physics before they enter the classroom. Therefore, although women are not underrepresented in algebra-based physics courses, these societal stereotypes can still impact their motivational beliefs in those courses.

The study presented here investigated gender difference in self-efficacy in introductory algebra-based physics courses in which women make up a majority of students in the class. In particular, we investigated the gender gap in self-efficacy at matched performance levels. It is also important to examine whether physics self-efficacy is a reflection of actual performance difference or whether the self-efficacy gender gap is due to pervasive societal stereotypes and biases female students must contend with throughout their lifetime. Below, we first delineate the research questions, then describe the methodology, present and discuss results, and finally conclude with instructional implications and future directions. The following research questions were answered by analyzing data from a validated survey administered to students in introductory algebra-based physics courses mainly for bioscience majors in which women outnumber men at a large public research university in the US:

3.2 Research Questions

RQ1. Are there gender differences in self-efficacy throughout a two-semester physics course sequence controlling for student performance?

RQ2. How do women's and men's self-efficacy change from the beginning to the end of each physics course (physics 1 and physics 2)?

RQ3. If there is a self-efficacy gender gap at the beginning of the first physics course, does the gap get bridged by the end of a year-long physics course sequence in which women outnumber men?

RQ4. If there is a self-efficacy gender gap, is it a reflection of the performance differences by gender or is there a gender gap beyond what is predicted by course performance due to pervasive stereotypes and biases female students must contend with throughout their lifetime?

3.3 Methodology

3.3.1 Participants who were administered the survey

The data were collected using a validated motivational survey by the authors and administered at a large public research university in the U.S. to students in a two-semester introductory algebra-based physics sequence in which women outnumber men. The survey was administered on scantrons in the recitation at the beginning (pre – during the first recitation) and end (post-in the last two weeks of classes) of physics 1 and physics 2. These two courses are generally taken in consecutive semesters, and in a normal sequence. We analyzed data across two consecutive academic years from courses that are traditionally taught. The classes are typically taken as a requirement by students on the pre-medical or pre-health track (mainly bioscience majors) primarily in their junior or senior year of undergraduate studies. Both of these are mandatory physics courses for all bioscience majors.

In this study, we analyzed 474 matched students (students who took the survey at both time points in physics 1 and physics 2). The university provided demographic information such as age, gender, and ethnicity/race information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the university data, the matched participants were 34% male and 66% female students. Thus, female students significantly outnumbered male students in these introductory physics courses. We note that the data provided by the university only include binary options of male and female. We recognize that gender is not a binary construct. However, we are limited to the binary data here (less than 1% of the students did not provide this information and thus were not included in this study).

3.3.2 Instrument validity

This study focused on students' responses to the physics self-efficacy survey items (see Table 1) [170-172] in the motivational survey, which included other items on other motivational constructs such as physics perceived recognition, interest, identity, etc. The items in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173]. A lower score was indicative of a negative endorsement of self-efficacy while a higher score was related to a positive endorsement. The items for different motivational beliefs including self-efficacy were adapted from previously validated surveys [6, 26, 170, 172, 174, 175]. The questions on self-efficacy covered students' belief in their ability to succeed in the physics course [27, 28]. Self-efficacy question Q1 was taken from the Peer Instruction self-efficacy instrument [172]. The other questions are adapted from Godwin et. al.'s performance/competence questions with minor changes based upon individual interviews during the validation of the survey at our institution [6].

Specifically, Q2, Q3, and Q4 (see Table 1) are the same or slightly modified questions from the survey used by Godwin et. al. [6]. The items on the survey were used in our prior work and revalidated in our own context [26] using one-on-one interviews with students to ensure that students interpreted the survey items correctly, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) [26], and Cronbach alpha [176]. Interviews with students suggest that they interpreted the items correctly. The Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.78 for physics 1 and 0.82 for physics 2, which is considered reasonable [176]. To ensure that items measured self-efficacy coherently and separately from other motivational constructs in the survey at the end of physics 1 and physics 2, an EFA was conducted. Additionally, we conducted a CFA with five motivational constructs in the broader survey. Since this study focuses only on the self-efficacy construct, the survey items and factor loadings for each self-efficacy item from the CFA are given in Table 1.

Table 1 Each self-efficacy survey item along with factor loadings from the Confirmatory Factor Analysis (CFA) for all students (N = 474) in physics 1 and physics 2. The Cronbach alpha for physics 1 is 0.78 and the Cronbach alpha for physics 2 is 0.82. The factor loading of each self-efficacy item to the self-efficacy construct (factor) is statistically significant to p < 0.001.

Survey Question	physics 1	physics 2
Q1. I am able to help my classmates with physics in the laboratory or in recitation.	0.581	0.672
Q2. I understand concepts I have studied in physics.	0.769	0.760
Q3. If I study, I will do well on a physics test.	0.730	0.762
Q4. If I encounter a setback in a physics exam, I can overcome it	0.771	0.703

In addition, we preformed item response theory (IRT) in order to check the response option distances for the survey constructs[177, 178]. The parametric grades response model using STATA was used to test the measurement precision of our response scale. The parametric grades item response model calculates the location parameter for each response and calculates the difference between the locations. The response scale discrimination values were 1.54 and 2.08. The numerical values for the location differences need to be approximately similar, as they are for the response scale, so we can use the means for these data. In addition, we estimated IRT-based domain scores. The correlation coefficient between the Likert mean scale values for self-efficacy and the IRT-based scores were calculated to be 0.98. Since the IRT-based domain score is so highly correlated with the mean score, using mean values for the scores is acceptable when analyzing the data and we can use parametric tests on Likert-scale data [179, 180]. In particular, because the psychological distance between adjacent response items and across items was approximately similar and complex factor scores derived from IRT or CFA are so highly correlated with the mean scores, it is reasonable to use mean scores [177, 181].

3.3.3 Analysis

We obtained students' final grades in each course from the university and linked those to students' gender and self-efficacy via the honest broker process discussed earlier. For analysis, we grouped students into bins by grade ranges (i) C- and below (considered insufficient to move on to the next course for this mandatory course for bioscience majors); (ii) C, C+; (iii) B-, B, B+ (iv) A-, A. For convenience, we labeled these bins D (e.g., this refers to C- and below), C, B, and A, respectively. Table 2 shows the distribution of students' final grades and the percentages of female students in each grade bin.

We first compared female and male students' mean self-efficacy scores by grade bins for statistical significance using *t*-tests and for calculating the effect size using Cohen's *d* [182]. In this analysis, we focused on self-efficacy differences of students who passed the course (C or above) since there were few students in the lowest grade bin. Next, to determine whether there are gender differences in self-efficacy controlling for grade (see Table 2), we performed a linear regression in which the dependent variable was students' average post self-efficacy score in physics 1 or physics 2 and the independent variables were students' gender and grade at the end of the course [183].

Table 2 The distribution of students' final grade (which includes women and men) and percentage of female students in each grade bin in physics 1 and physics 2.

Course	Total	D	C	B	A
Physics 1	N = 474 66% female	N = 1 100% female	N = 52 71% female	N = 195 73% female	N = 226 58% female
Physics 2	N = 474 66% female	N = 16 63% female	N = 94 72% female	N = 168 70% female	N = 196 60% female

3.4 Results and Discussion

3.4.1 Gender differences in self-efficacy by grade

Regarding **RQ1 for physics 1**, we find women have lower average self-efficacy scores than men in most of the grade bins (A, B, C) at the beginning and end of the course (see Table 3). Table 3 shows that the differences are statistically significant for students in the A and B bins at the beginning and end of the course. However, in regard to RQ2 for physics 1, the average self-efficacy scores of both men and women decreased for those students who received B's or C's in the physics 1 course. The differences between men and women range from a large effect size (greater than 0.50) to a medium effect size (greater than 0.30). Furthermore, Table 3 shows that women receiving A's have about the same self-efficacy as men who receive between a C and a B at the beginning of the course in physics 1. However, by the end of the course, women with A's have roughly the same self-efficacy as men with B's (Table 3). Table 3 also shows that women and men's average self-efficacy scores were stable (did not significantly change over the course of the semester) for the students who received A's.

Table 3 Self-efficacy scores in physics 1 by grade bin for female and male students. N = number of students, Pre = mean pre self-efficacy score, Post = mean post self-efficacy score. Cohen's d for the gender differences in pre and post scores for each grade bin is shown (all statistically significant gender differences have been bolded).

Physics 1 Self-Efficacy by Grade

	C			B			A		
	Pre	Post	N	Pre	Post	N	Pre	Post	N
Men	2.78	2.65	15	3.07	2.92	52	3.09	3.10	95
Women	2.79	2.43	37	2.76	2.58	143	2.90	2.96	131
Cohen's *d*	-0.01	0.36		**0.65**	**0.60**		**0.40**	**0.30**	
p-value	0.966	0.252		<0.001	<0.001		0.003	0.027	

Regarding **RQ1 for physics 2**, the gender gap in self-efficacy controlling for grade continued. Table 4 shows that women had lower self-efficacy scores than men in all grade bins at the beginning and at the end of physics 2, but the self-efficacy gender difference was not statistically significant in the group of students who received a C grade. However, by the end of the course in physics 2, the only statistically significant difference occurred for women and men in the A range with a medium to large effect size (Table 4). This finding differs from a similar study in the calculus-based introductory physics courses (in which women are severely underrepresented) in which the physics self-efficacy gender gap increased from the beginning to the end of the course [43].

In regard to **RQ2 for physics 2,** at the beginning of the course, women who receive A's have the same self-efficacy as men who receive C's (Table 4). However, by the end of the course, women with A's have the same self-efficacy as men with grades between a B and an A (see Table 4). Again, women's and men's average self-efficacy scores did not significantly change over the course of the semester for the students who received A's in the physics 2 course. Also, the average self-efficacy scores of both men and women decreased for those students who received B's or C's in the course.

Regarding **RQ3**, Table 3 and Table 4 show that the self-efficacy gender gap did not get bridged by the end of a year-long course sequence. Data from both physics 1 and physics 2 show that there is some protective effect in students' self-efficacy for men and women who receive A's in both courses. However, the gender gap is maintained even among students who received an A grade and the self-efficacy scores decreased for students in the B and C bins. These findings suggest that there may be a need to make the learning environment equitable and inclusive. It is important to note that women had lower self-efficacy scores than men in many of the grade bins (as evidenced by both the pre and post tests scores) even though they outnumber men in the course. Prior interviews with students suggest that some women in physics classes would get test grades comparable to their peers in the class, but they still felt that they weren't good at physics [53]. We believe that this disparity may be due to the deep-rooted societal stereotypes and biases about who can excel in physics that women must contend with from an early age, starting long before they step into the college physics classroom [25].

Table 4 Self-efficacy scores in physics 2 by grade bin for female and male students. N = number of students, Pre = mean pre self-efficacy score, Post = mean post self-efficacy score. Cohen's d for the gender differences in pre and post scores for each grade bin is shown (all statistically significant gender differences have been bolded).

Physics 2 Self-Efficacy by Grade

	C			B			A		
	Pre	Post	N	Pre	Post	N	Pre	Post	N
Men	2.88	2.64	26	3.00	2.74	51	3.10	3.10	79
Women	2.69	2.50	68	2.73	2.68	117	2.89	2.88	117
Cohen's d	0.40	0.27		**0.56**	0.10		**0.51**	**0.50**	
p-value	0.085	0.244		<0.001	0.561		<0.001	<0.001	

Lastly, in regard to **RQ4**, we investigated the fraction of the self-efficacy gender gap that was related to students' course performance versus students' biased perceptions. First, we found the raw post self-efficacy differences between women and men to be 0.28 ($p < 0.001$) in physics 1 and 0.20 ($p < 0.001$) in physics 2 in favor of men. Then we controlled for students' grade in each course with multiple linear regression [183] with post self-efficacy as the dependent variable and gender and grades as the independent variables, including an interaction effect between grade and gender (see Equation 1). The interaction effect (β_3) between gender and grade was not a significant predictor for self-efficacy in either physics 1 ($p = 0.221$) or physics 2 ($p = 0.299$) and thus was omitted in future analyses.

Equation 1

$$post\ self-efficacy\ = \beta_1(gender) + \beta_2(grade) + \beta_3(grade \times gender) + const$$

In physics 1, a significant overall regression correlation was found (F(2,471) = 47.96, $p < 0.001$) for post self-efficacy scores with main effects from both gender ($p < 0.001$) and final grade ($p < 0.001$) [182]. The regression equation is shown in Equation 2. In physics 2, a significant overall regression correlation was found (F(2, 471) = 34.78, $p < 0.001$) for post self-efficacy scores with main effects from gender ($p = 0.003$) and final grade ($p < 0.001$). The regression equation for this case is shown in Equation 3. Gender was coded 0 for female students, or 1 for male students and grade was reported continuously from 1-4. Both gender and grade were significant predictors of students' average self-efficacy in both physics 1 and physics 2 even though women make up a numerical majority in the course.

Equation 2

$$post\ self-efficacy\ in\ physics\ 1 = 1.67 + 0.21(gender) + 0.32(grade)$$

Equation 3

$$post\ self-efficacy\ in\ physics\ 2 = 2.00 + 0.15(gender) + 0.23(grade)$$

Compared to the raw post self-efficacy differences, the mean values of the gender gap in self-efficacy (from the multiple linear regression) decreased slightly to 0.21 ($p < 0.001$) in physics 1 (see Equation 2) and 0.15 ($p = 0.003$) in physics 2 (see Equation 3). Therefore, while a small portion of the gap in self-efficacy can be attributed to performance in the course, the self-efficacy gender gap mainly comes from women's biased perceptions of how well they can do in physics: 75% in both physics 1 and physics 2 (see Figure 3). Thus, similar to earlier investigation in calculus-based courses in which women are severely underrepresented [43], even in these algebra-based courses where women make up the majority of the students, biased beliefs about who can excel in physics that are pervasive in our society still affects their self-efficacy.

Figure 3 Relative contribution of performance and biased perception to the post self-efficacy gender gap in physics 1 and physics 2 for bioscience majors in which women outnumber men.

3.5 Implications and Future Directions

The physics self-efficacy gender gap we found at equal grade levels may seem unexpected since women are not underrepresented in these introductory algebra-based physics courses mainly taken by the bioscience majors (for whom both courses are mandatory). However, we hypothesize that the societal stereotypes and biases about who can excel in physics that women from a young age are bombarded with are at the heart of the gender gap in physics self-efficacy in these courses in which women outnumber men. Not only are female students exposed to gendered views of physics in informal situations growing up, e.g., from shows such as The Big Bang Theory or from their family and friends, they are also treated differently and given differential advice by K-12 educators and guidance counselors [184].

Our finding is important and this issue needs to be addressed since self-efficacy has been shown to affect students' persistence, learning and long term outcomes [33]. Students with higher self-efficacy are more likely to engage with challenging problems without anxiety and self-efficacy plays an important role in predicting student performance in courses in that domain [33]. Low self-efficacy can create a negative feedback loop when students receive feedback via graded assignments in their classes. In particular, a student's lower self-efficacy can cause anxiety and deteriorate their performance, which can then weaken their self-efficacy further.

However, there appears to be some benefit to having a large cohort of women in these introductory algebra-based physics courses for bioscience majors. We find that the percentage of the gender gap attributed to biased perception does not change significantly from physics 1 to physics 2 (Figure 3). However, the percentage of the gender gap in introductory calculus-based physics courses (in which women are severely underrepresented) attributed to biased perception started out lower in physics 1 and gets worse in physics 2 [43]. Thus, we observed a somewhat protective effect in women's self-efficacy when they make up about two-thirds of the students in these courses. Nevertheless, since women still have lower self-efficacy than men in these courses controlling for grade, there is urgent need to create an equitable and inclusive learning environment to bridge the self-efficacy gender gap.

In particular, equitable and inclusive pedagogical strategies are necessary to bridge the gender gap in self-efficacy and raise the self-efficacy of all students. Active learning pedagogies could decrease women's self-efficacy if not implemented with equity and inclusion as central constructs. In one study in engineering, women working in mixed gender cooperative groups tended to be undermined by their male classmates and played a less active role in those groups [15]. Additionally, in a study in physics, students in courses that utilized active engagement strategies performed better than students in lecture-based courses, however, the gender gap in performance on conceptual surveys persisted and even became worse in some cases [11]. In relation to self-efficacy, some interactive engagement physics courses have been shown to decrease women's self-efficacy [29] whereas others have been shown to improve women's self-efficacy [99]. Therefore, it is important for the instructor/TA to ensure that all students feel they belong and have high self-efficacy and can contribute equally. One strategy physics instructors can use to encourage equal contribution in a group is to assign each student a role that rotates throughout the course. However, this approach alone is unlikely to be effective in bridging the self-efficacy gender gap if the learning environment is otherwise not equitable and inclusive.

A field-tested strategy to make the learning environment equitable and inclusive is through classroom interventions [152, 185, 186]. In particular, social-psychological classroom interventions, e.g., sense of belonging and mindset interventions, have been shown to eliminate gender performance gaps [151, 186] and to have lasting effects beyond the time they are implemented [185]. At the same time, these interventions can help students develop positive feelings of being recognized by their peers, TAs, and instructors. We have implemented a short ecological belonging intervention in the calculus-based physics classes that eliminated the gender gap in physics performance [186]. These types of interventions could be adapted and their effectiveness investigated for the algebra-based physics courses discussed here.

4.0 Students self-efficacy in an introductory physics course for students on a bioscience track predicts their grade

4.1 Introduction and Theoretical Framework

Recently, research has focused on students' persistence in science, technology, engineering, and math (STEM) related fields. Students are more likely to persist in science majors and science related careers if they have a higher science identity [54, 158, 159], a higher interest in science [157], or higher self-efficacy [32, 39, 41, 42, 155, 156]. Studies have shown that academic self-efficacy, or a person's belief that they can succeed in a particular activity or course [27, 28], can impact students' engagement, learning, and achievement in science courses [14, 31, 36, 40-42, 163, 164]. For example, self-efficacy has been shown to predict students' motivation to embrace and persist on challenging tasks. Students with higher self-efficacy in a domain are also more likely to adopt superior methods of learning [33]. However, for students from underrepresented groups, such as women, their self-efficacy and other motivational beliefs may be undermined by societal stereotypes and biases about who belongs in physics and can excel in it. For example, a gender gap favoring men in motivational factors [26, 97, 160] and conceptual tests [84] has been studied in STEM courses. Therefore, understanding the role of women's STEM self-efficacy in multiple contexts is important to address issues of equity and inclusion in STEM classrooms.

One reason the gender gap in self-efficacy is concerning is that a higher self-efficacy was found to correlate with higher course achievement in science courses [14, 31, 40, 163]. Other studies have shown that students with higher self-efficacy were more likely to persist on challenging tasks [36] and less likely to second guess themselves and switch to an incorrect answer

[164]. Students with high self-efficacy in a domain, like physics, are less likely to have anxiety that can rob them of their cognitive resources while learning and test taking [166]. They are also more likely to employ better academic strategies, such as goal-setting, self-monitoring, and self-evaluation [33]. In addition, instructors have the potential to increase self-efficacy since self-efficacy may be influenced by the classroom environment [37, 161, 162] and different teaching strategies [29, 34, 38]. In physics courses, in particular, self-efficacy is one of the factors that contributes to students' physics identity along with recognition by others and interest in physics [3, 5, 59]. A students' identity is important because identity in a particular domain has been shown to play an important role not only in students' in-class participation and performance but also in their choices of future courses and careers [5, 49, 55-57, 165].

Although self-efficacy can increase students' persistence in science majors and careers [167], in physics courses in which women are underrepresented, a gender gap in self-efficacy [26, 30, 43, 168] and conceptual tests have been found. In many cases, the gender gap in self-efficacy widens by the end of the course, even in interactive engagement courses [29, 43]. Several factors have been proposed to explain the gender difference in students' motivational beliefs (including self-efficacy) and on conceptual tests. Some studies suggest that differences in prior preparation for various reasons may account for the gender difference [13, 114]. In addition, societal stereotypes and biases about who belongs in physics can negatively impact women in the introductory physics courses. Some of the elements of the environment in many science classrooms can negatively impact women including a lack of female role models, pedagogy that favors male students' interests, and a "chilly climate" for women [25]. By high school, girls are less likely to believe that physics is for them and that their teacher thinks they are good at physics [1]. These societal stereotypes and biases can impact women even in courses where they are not underrepresented, like introductory physics courses for students on the bioscience track. However, few studies focus on the effect of students' self-efficacy on their grades in these courses.

This study examined the difference between male and female students with regard to the role of their self-efficacy in predicting performance outcomes at the end of a mandatory second semester algebra-based introductory physics course sequence for students on the bioscience track, controlling for their high school GPA, SAT math scores, grade, and self-efficacy at the end of physics 1. We controlled for students' past performance in high school in order to isolate the effect of self-efficacy on student outcomes controlling for prior high school performance. A visual representation of our final model is shown in Figure 4. All paths were considered from left to right in our model. However, only some of the paths are shown for clarity. Our research questions are as follows:

RQ1. Are there gender differences in students' high school GPA and SAT math, as well as grades and self-efficacy in physics 1 and physics 2?

RQ2. To what extent does physics self-efficacy predict physics course grade?

RQ3. What unique role does the academic factors of SAT math, high school GPA, and grade in physics 1 play in mediating students' grade in physics 2?

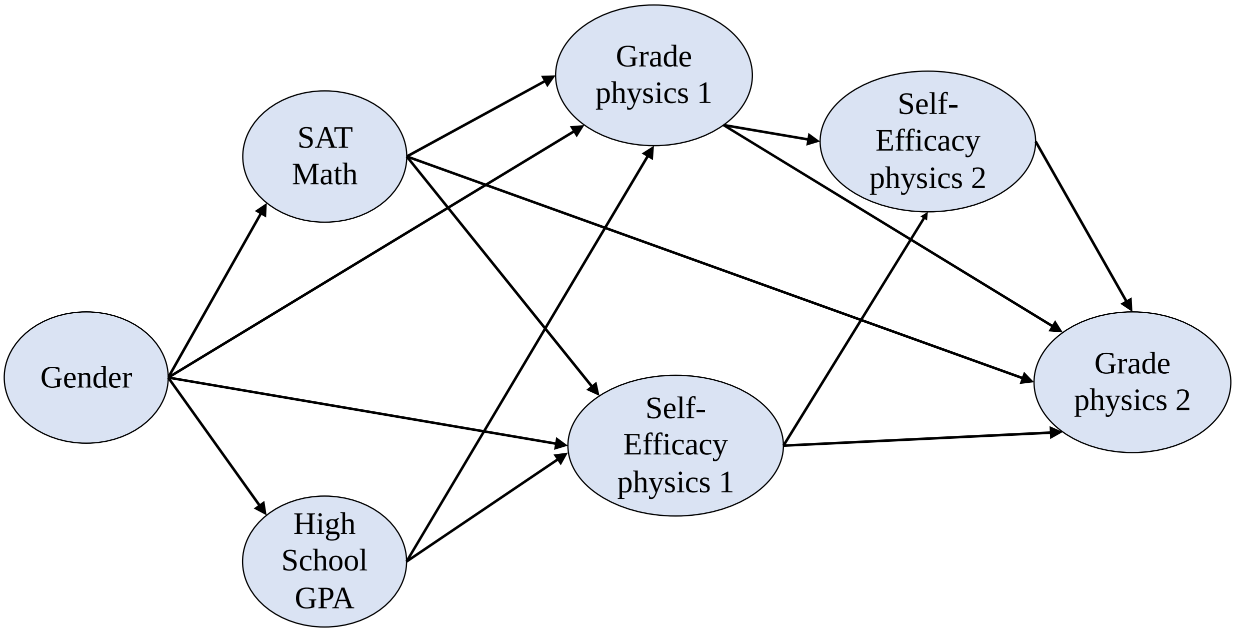


Figure 4 Schematic representation of the model and how self-efficacy mediates the relation between gender and grade in the physics 2 course for bioscience majors in which women are not underrepresented. From left to right, all possible paths were considered (including one from gender to grade in physics 2), however, only some of the paths are shown for clarity.

4.2 Methodology

In this study, a validated survey covering self-efficacy and other motivational constructs (not discussed here) was administered to students at a large public research university in the U.S. The survey was given at the beginning (pre) and end (post) of the semester in the two introductory algebra-based physics courses over the course of two years. This course is primarily taken by junior or senior bioscience majors or chemistry majors on the bioscience track for whom a two-semester physics course sequence is mandatory. We analyzed the data for 564 students who completed the survey at the end of the introductory physics 1 and 2 classes. The University provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the University data, the participants were 36% male and 64% female students. We recognize that gender is fluid and not a binary construct; however, the data collected by the institution is in binary terms. Thus, we use binary data provided by the university in this study. Less than 1% of the students did not choose male or female and thus were not included in the study.

4.2.1 Instrument validity

This study measured the physics self-efficacy of students enrolled in both of the introductory algebra-based physics courses primarily for bioscience majors. The validated survey involving several motivational beliefs including the self-efficacy items was adapted from previous research [170, 175] and re-validation of the survey at our institution involved conducting one-on-one student interviews [26], Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and calculation of the Pearson Correlations and Cronbach alpha. The *self-efficacy* questions on the survey covered students' belief in their ability to succeed in the physics course [27, 28]. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173]. A lower score is indicative of a negative endorsement of the survey construct while a higher score is related to a positive belief of the construct. Two of the questions were reverse-coded. The survey questions for self-efficacy and factor loadings for each question from the CFA are given in Table 5. The Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.79 for the self-efficacy questions which is considered reasonable [176].

Table 5 Survey questions for each of the motivational constructs along with factor loadings (Lambda) from the Confirmatory Factor Analysis (CFA) result for all students (N = 564). The rating scale for the physics self-efficacy questions was not at all true, a little true, somewhat true, mostly true, and completley true. All *p*-values (of the significance test of each item loading) are *p* < 0.001.

Physics Self-Efficacy Construct and Item	Lambda
I am able to help my classmates with physics in the laboratory or in recitation.	0.60
I understand concepts I have studied in physics.	0.69
If I study, I will do well on a physics test.	0.82
If I encounter a setback in a physics exam, I can overcome it	0.70

4.2.2 Analysis

Initially, we analyzed the descriptive statistics and compared female and male students' mean scores on the self-efficacy questions and students' grades for statistical significance using *t*-tests and to investigate the effect size using Cohen's d [182]. Cohen's d is $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students and σ_{pooled} is the pooled standard deviation (or weighted standard deviation for men and women) for all students. In general, d = 0.20 indicates a small effect size, d = 0.50 indicates a medium effect size, and d = 0.80 indicates a large effect size [182].

To quantify the statistical significance and relative strength of our frameworks' path links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. SEM is an extension of multiple regression and allows one to conduct several multiple regressions simultaneously between variables in one estimation model. This is an improvement over multiple regression since it allows us to calculate the overall goodness of fit and allows for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and SRMR and RMSEA < 0.08 [139].

The model estimates were performed using gender moderation analysis to check whether any of the relations between variables show differences across gender by using "lavaan" to conduct multi-group SEM [187]. 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 [187]. Since we did not find significant moderation by gender, we tested the theoretical model in mediation analysis, using gender as a variable directly predicting items to examine the resulting structural paths between constructs.

4.3 Results

4.3.1 Gender differences in predictors and outcomes

We find that women had statistically significantly lower mean values in their SAT math scores, physics 1 grade, physics 2 grade, and self-efficacy in physics 1 and physics 2 (see Table 6) while men had lower mean values in high school GPA (see Table 6). In addition, we report the percentages of students who selected each choice for each self-efficacy survey item in Appendix B (page 290), which shows consistent results with the descriptive statistics shown in Table 6.

Table 6 Mean predictor and outcome values by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender.

Predictors and Outcomes (Score Range)	Mean		*p-value*	Cohen's *d*
	Male	Female		
High School GPA (0-5)	4.11	4.16	0.170	-0.12
SAT math (200-800)	684	662	0.001	0.32
Self-efficacy physics 1 (1-4)	2.98	2.73	<0.001	0.50
Self-efficacy physics 2 (1-4)	2.93	2.71	<0.001	0.42
Physics 1 Grade (0-4)	3.40	3.24	0.005	0.25
Physics 2 Grade (0-4)	3.20	3.03	0.023	0.20

4.3.2 SEM path models

We used SEM to investigate the relationships between the constructs and to unpack students' self-efficacy contribution to explaining the physics 2 grade of women and men. In order to understand the contribution from high school factors (SAT math and high school GPA) and grade in physics 1, we also ran SEM models without those constructs included. We initially conducted gender moderation analysis between variables using multi-group SEM to investigate if any of the relationships between the constructs were different across gender. There were no group differences at the levels of weak and strong measurement invariance and the level of regression coefficients. Thus, there was no moderation effect by gender. Therefore, we proceeded to gender mediation to investigate the extent to which gender differences in physics 1 grade at the end of the course were mediated by differences in students' pre-college academic measures (high school GPA and SAT Math) and grade and self-efficacy in physics 1.

4.3.2.1 SEM model 1 with high school factors

The result of the path analysis part of the SEM is presented visually in Figure 5 for model 1. The model fit indices indicate a good fit to the data: CFI = 0.952 (> 0.90), TLI = 0.936 (0.90), RMSEA = 0.057 (<0.08), and SRMR = 0.041 (<0.08). Gender had direct connections to SAT math (β = 0.16) and self-efficacy in physics 1 (β = 0.23). However, gender did not have any direct connections to the grade or self-efficacy in physics 2. Instead, gender was mediated through students' high school factors and self-efficacy in the physics 1 course.

Students' grade in physics 2 had direct connections to SAT math (β = 0.16) as well as self-efficacy in physics 2 (β = 0.13), and grade in physics 1 (0.58). The variance explained by the grade in physics 2 is 0.510

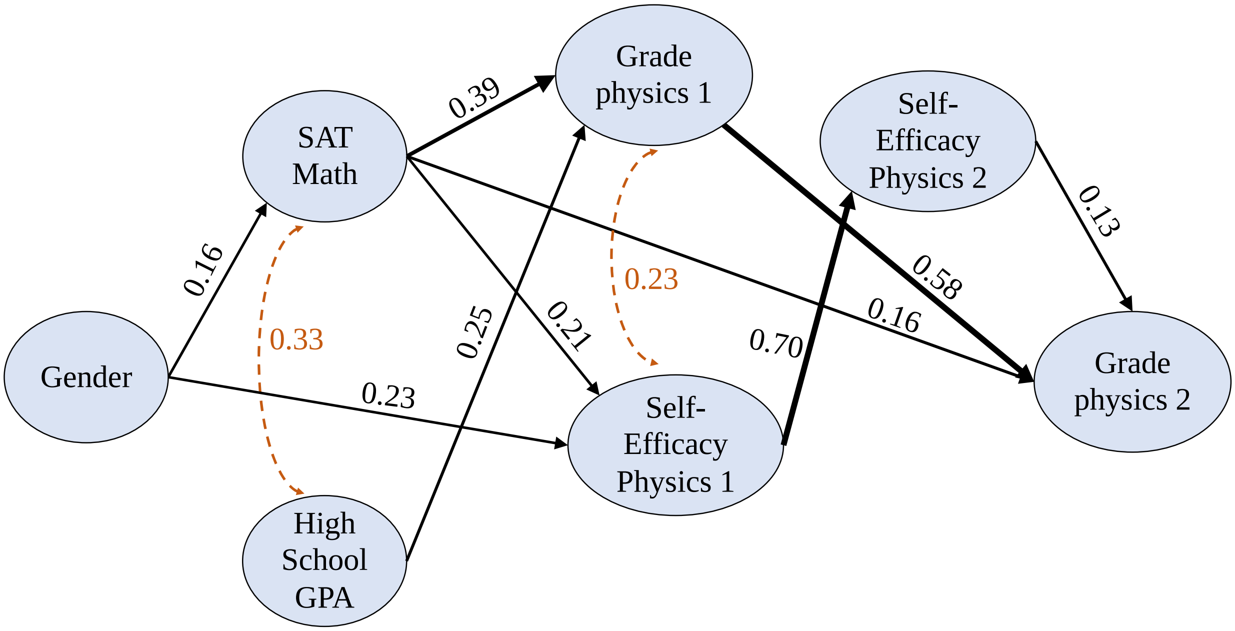


Figure 5 Result of the path analysis part of the SEM between gender and physics 2 grade through students' self-efficacy for model 1. The line thickness indicates the relative magnitude of β values. The dashed lines indicate covariances between constructs. All *p*-values are p < 0.001. Gender does not directly predict the physics grade.

4.3.2.2 SEM model 2 without SAT math

The result of the path analysis part of the SEM without SAT math is presented visually in Figure 6 for model 2. The model fit indices indicate a good fit to the data: CFI = 0.954 (> 0.90), TLI = 0.940 (0.90), RMSEA = 0.059 (<0.08), and SRMR = 0.039 (<0.08). Gender had direct connections to grade in physics 1 (β = 0.14) and self-efficacy in physics 1 (β = 0.29). However, gender did not have any direct connections to the grade or self-efficacy in physics 2.

Similar to model 1, students' grade in physics 2 had direct connections to self-efficacy in physics 2 (β = 0.15), and grade in physics 1 (0.66). The variance explained by grade in physics 2 is 0.494, which is similar to the variance explained by the grade in physics 2 for model 1.

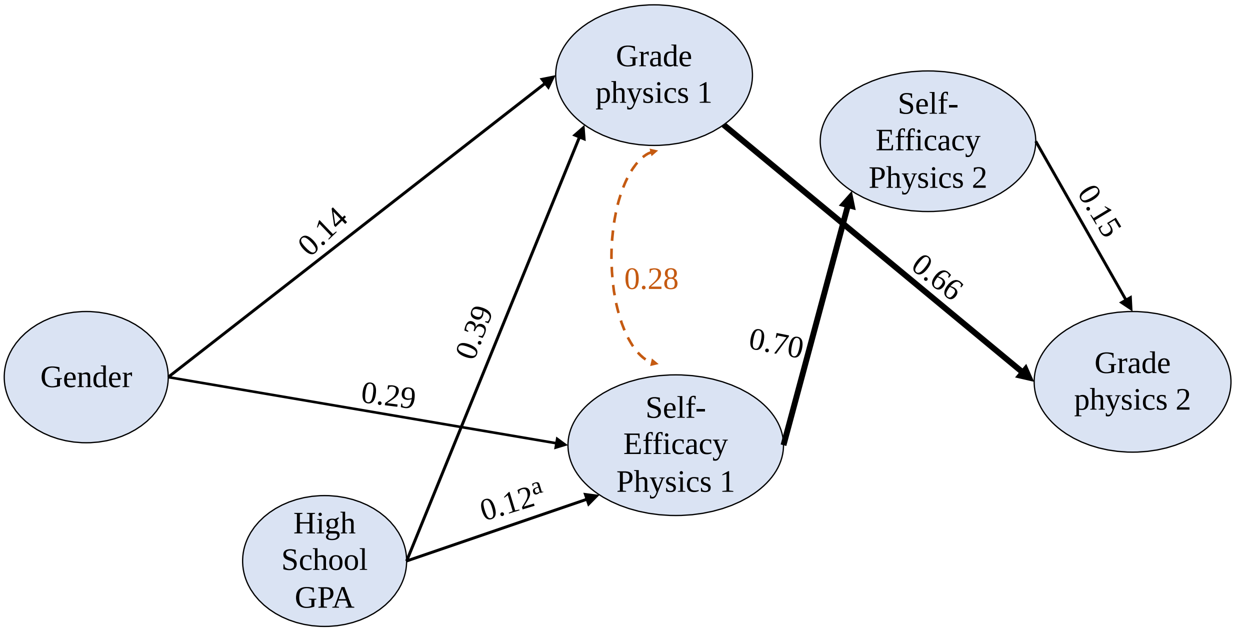


Figure 6 Result of the path analysis part of the SEM without SAT math between gender and physics 2 grade through students' self-efficacy for model 2. The line thickness indicates the relative magnitude of β values. The dashed lines indicate covariances between constructs. *P*-values are indicated by no superscript for $p <$ 0.001 and superscript "a" for $p = 0.008$. Gender does not directly predict the physics 2 grade.

4.3.2.3 SEM model 3 without high school GPA

The result of the path analysis part of the SEM without high school GPA is presented visually in Figure 7 for model 3. The model fit indices indicate a good fit to the data: CFI = 0.949 (> 0.90), TLI = 0.932 (0.90), RMSEA = 0.063 (<0.08), and SRMR = 0.041 (<0.08). Gender had direct connections to SAT math (β = 0.14) and self-efficacy in physics 1 (β = 0.24). Similar to previous models, gender did not have any direct connections to the grade or self-efficacy in physics 2.

Similar to model 1, students' grade in physics 2 had direct connections to self-efficacy in physics 2 (β = 0.13), and grade in physics 1 (0.59). In addition, grade in physics 2 had a direct connection to SAT math (β = 0.16) The variance explained by grade in physics 2 is 0.509, which is similar to the variance explained by the grade in physics 2 for model 1 and model 2.

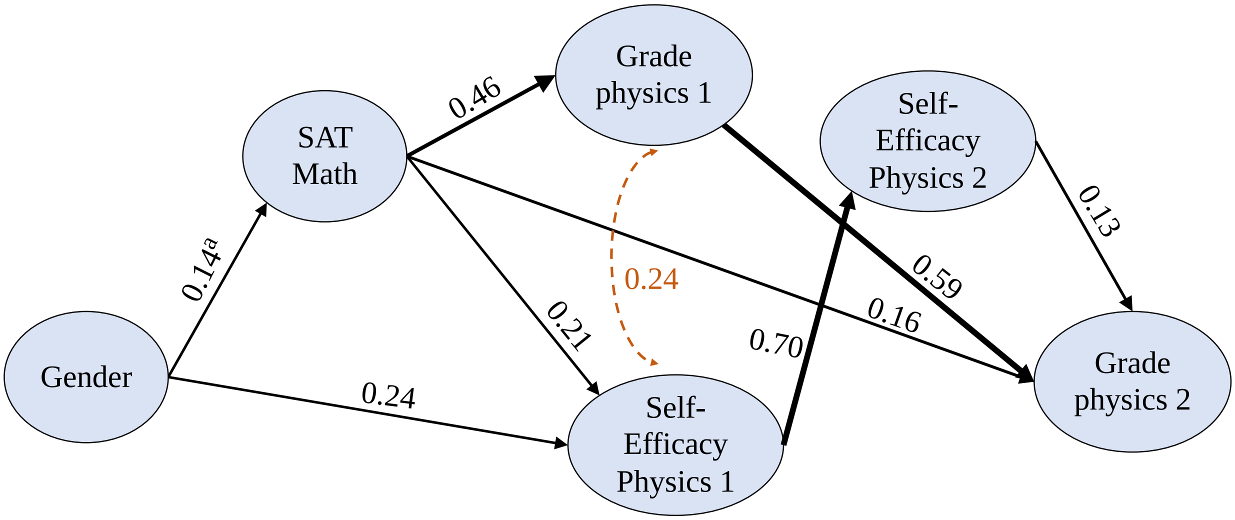


Figure 7 Result of the path analysis part of the SEM without high school GPA between gender and physics 2 grade through students' self-efficacy for model 3. The line thickness indicates the relative magnitude of β values. The dashed lines indicate covariances between constructs. The *p*-values are indicated by no superscript for $p < 0.001$ and superscript "a" for $p = 0.002$. Gender does not directly predict the physics 2 grade.

4.3.2.4 SEM model 4 without high school GPA or SAT math

The result of the path analysis part of the SEM without high school GPA or SAT math is presented visually in Figure 8 for model 4. The model fit indices indicate a good fit to the data: CFI = 0.952 (> 0.90), TLI = 0.935 (0.90), RMSEA = 0.065 (<0.08), and SRMR = 0.040 (<0.08). Gender had direct self-efficacy in physics 1 (β = 0.28) and grade in physics 2 (β = 0.11).

Similar to model 1, students' grade in physics 2 had direct connections to self-efficacy in physics 2 (β = 0.15), and grade in physics 1 (0.65). In addition, grade in physics 2 had a direct connection to gender (β = 0.11) The variance explained by grade in physics 2 is 0.492, which is similar to the variance explained by the grade in physics 2 for model 1, model 2, and model 3 which shows that the high school factors do not play a large role in explaining students' grade at the end of physics 2.

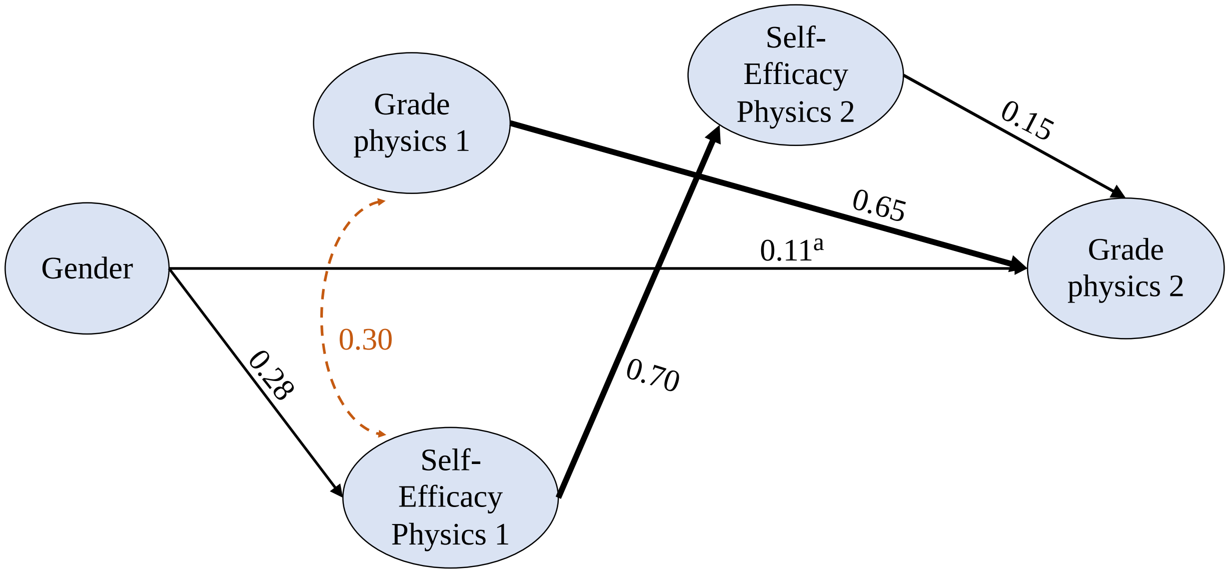


Figure 8 Result of the path analysis part of the SEM without high school GPA and SAT math between gender and physics 2 grade through students' self-efficacy for model 4. The line thickness indicates the relative magnitude of *β* values. The dashed lines indicate covariances between constructs. *P*-values are indicated by no superscript for $p < 0.001$ and superscript "a" for $p = 0.007$.

4.3.2.5 SEM model 5 without grade in physics 1

The result of the path analysis part of the SEM without grade in physics 1 is presented visually in Figure 9 for model 5. The model fit indices indicate a good fit to the data: CFI = 0.954 (> 0.90), TLI = 0.939 (0.90), RMSEA = 0.055 (<0.08), and SRMR = 0.037 (<0.08). Gender had direct connections to SAT math (β = 0.26) self-efficacy in physics 1 (β = 0.25) and grade in physics 2 (β = 0.11).

Similar to previous models, students' grade in physics 2 had direct connections to self-efficacy in physics 2 (β = 0.22). In addition, grade in physics 2 had a direct connection to SAT math (β = 0.36) and), and high school GPA (0.19). The variance explained by grade in physics 2 is 0.288, which is about 0.30 lower than the variance explained in past models which show that the students' grade in physics 1 plays an important role in explaining students' grade at the end of physics 2.

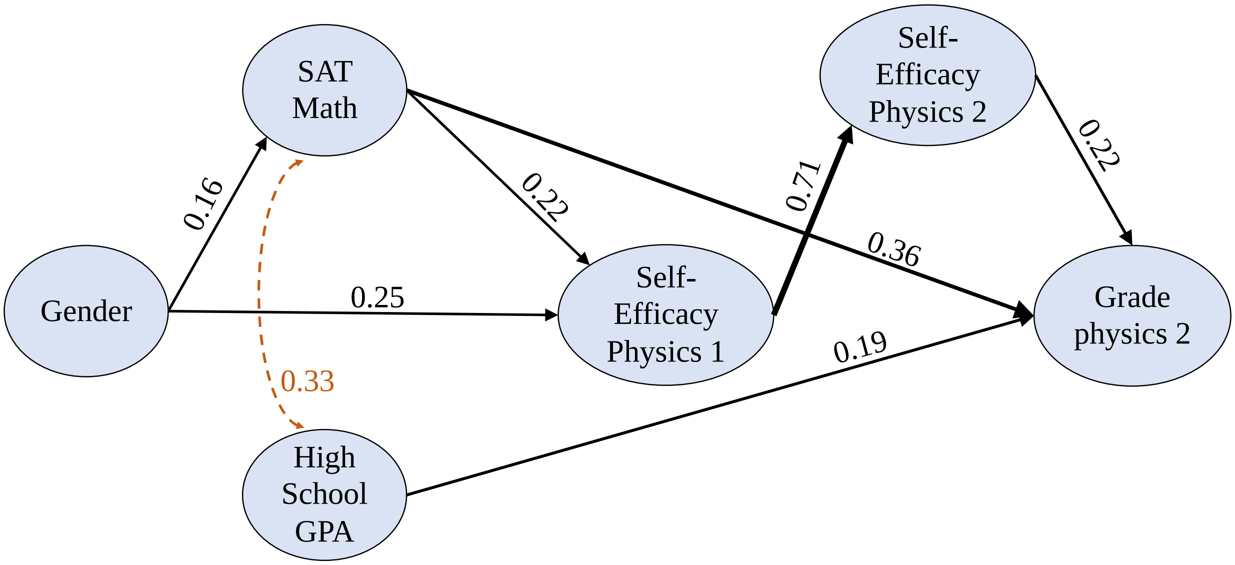


Figure 9 Result of the path analysis part of the SEM without grade in physics 1 between gender and physics 2 grade through students' self-efficacy for model 5. The line thickness indicates the relative magnitude of β values. The dashed lines indicate covariances between constructs. All *p*-values are for $p < 0.001$.

4.3.2.6 Variance explained by each model

After constructing each model, we calculated the coefficient of determination (Adjusted R^2) of grade in physics 2 in every model to investigate the proportion of variance explained by each model (see **Error! Reference source not found.**). We found that model 1 and model 3 e xplained the same amount of variance in physics 2 grade (51%) while model 2 and model 4 explained slightly less of the variance (49%). However, the amount of variance in physics 2 grade decreased to 29% in model 5 when students' physics 1 grade was removed from the model.

Table 7 Adjusted coefficients of determination (Adjusted R2) for physics 2 grade in the five different models. All *p*-values are <0.001.

Model	Adjusted R^2 physics 2 grade
Model 1: All Factors	0.51
Model 2: No SAT math	0.49
Model 3: No high school GPA	0.51
Model 4: No high school GPA and SAT math	0.49
Model 5: No physics 1 grade	0.29

4.4 Summary and Discussion

In this research, we find gender gaps in physics grades and self-efficacy disadvantaging women in the physics 2 course for bioscience majors. Our SEM model shows that students' self-efficacy plays an important role in predicting students' grades at the end of the physics course. However, on average, women have lower self-efficacy and grades than men even in these courses in which they are not underrepresented. In addition, the variance in physics 2 grade only decreased by 0.02 when high school GPA and SAT math scores were taken out of the model. In addition, recent studies have shown that SATs could disadvantage students from minority and low-income groups [188-191]. Therefore, removing SAT requirements for colleges may not have a large impact on students' learning outcomes.

It is important to note that students' self-efficacy is an important motivational belief itself beyond how it predicts grade in the physics course. Students' self-efficacy also predicts identity [5, 59] and persistence on challenging tasks [36]. Equitable and inclusive pedagogical strategies are necessary to bridge the gender gap in self-efficacy and raise the self-efficacy of all students. However, active learning pedagogies could decrease women's self-efficacy if not implemented with equity and inclusion as central constructs. Some interactive engagement physics courses have been shown to decrease women's self-efficacy [29] whereas others have been shown to improve women's self-efficacy [99]. In one study in engineering, women working in mixed-gender cooperative groups tended to be undermined by their male classmates and played a less active role in those groups [15]. Additionally, in a study in physics, students in courses that utilized active engagement strategies performed better than students in lecture-based courses, however, the gender gap in performance on conceptual surveys persisted and even became worse in some cases [11]. Therefore, the instructor/TA needs to ensure that all students feel they belong and have high self-efficacy, and can contribute equally. One strategy physics instructors can use to encourage equal contribution in a group is to assign each student a role that rotates throughout the course. However, this approach alone is unlikely to be effective in bridging the self-efficacy gender gap if the learning environment is otherwise not equitable and inclusive.

One strategy to make the learning environment equitable and inclusive is through brief social-psychological classroom interventions, e.g., focusing on improving students' sense of belonging and mindset [152, 185, 186]. They have been found to create a more inclusive learning environment that can mitigate students' doubts about belonging in college and improve the grades of racial and ethnic minority students [192] and women [74, 151] in STEM fields. These types of interventions have the potential to also increase the self-efficacy of students from these marginalized groups in physics such as women. It is important that instructors focus on increasing women's self-efficacy and other motivational beliefs such as sense of belonging in the introductory physics courses since they can influence students' persistence and engagement in the course and can also impact later choices of courses and careers. These interventions can be one step towards creating an equitable and inclusive learning environment where all students have higher motivational beliefs including sense of belonging and self-efficacy and can excel.

5.0 Feeling less recognized by instructors and TAs as a physics person predicts female students’ grades in a physics course for bioscience majors

5.1 Introduction and Theoretical Framework

In the past few decades, several research studies have focused on the experiences and participation of women in many science, technology, engineering, and math (STEM) fields [115, 145, 147-150, 153]. Research shows that motivational beliefs in academic domains can influence students’ continuation in related courses, majors, and careers [26, 32, 39, 41, 43, 54, 155-159]. Some of these studies that focus on women and ethnic and racial minority students in physics show that they, in general, have lower motivational beliefs than men that may be due to pervasive societal stereotypes and biases about who belongs in physics and can excel in it that students internalize over their lifetime and learning environments being inequitable and non-inclusive. A student’s perceived recognition by others as a science person has been shown to play an important role in predicting students’ other science motivational beliefs [50, 55, 140, 165, 193].

Prior studies show that a student’s perception of recognition by others, e.g., as a science person, plays an important role in their persistence and academic performance in STEM fields [194], and it is especially true for underrepresented students [195]. When women enter STEM fields, they do not feel validated and recognized as often as men [196-198]. Our prior individual interviews suggest that students’ perceived recognition as a physics person by instructors and teaching assistants (TAs) predicts their self-efficacy and interest in physics (e.g., see [52, 53]). Prior research also suggests that, in physics, recognition by others as a physics person predicts students' physics identity, i.e., whether they see themselves as a person who can excel in physics

[5, 59] and their career choices [6, 199]. Godwin et. al. found that in a sample of a general population of college students in English composition classes (a majority of students were in their first year), students' identity as a physics person (based upon pre-college exposure to physics) was predicted by their perceived recognition by others as a physics person [6]. However, it is unclear whether perceived recognition as a physics person plays a similar role in a different population of students who are taking a physics course required for their major, e.g., an introductory physics course for bioscience majors, and how these students' perceived recognition predicts their physics grades. Our study focuses on this important issue for female and male bioscience majors and the physics course is important for them not only because it is mandatory for their major but also because it is required for medical school that many of these students are aspiring to get into and information in the physics course is part of the Medical College Admission Test (MCAT). Recognition from others, especially instructors and TAs, as a physics person can be vital for student success in the physics course as well as their career aspirations.

In addition, in order for a student to feel recognized by instructors/TAs as a physics person, instructors/TAs need to both explicitly recognize the student (e.g., by verbally acknowledging their progress and achievements) and implicitly recognize the student (e.g., by setting high standards for all of their students and making it clear to them that they all have what it takes to excel in physics if they work hard and take advantage of all of the resources and they are there as instructors/TAs to support them as needed all the way through) [103]. Hazari et. al. found that in high school physics classes, students' physics identity, predicted by students' recognition as a physics person by others, was influenced by teachers' social cues, or how teachers' actions obscured social boundaries between teachers and students [200]. In another study [201], students who received praise for their intelligence were more likely to have a fixed mindset and lower levels of task persistence, enjoyment, and performance than students who received praise for their effort. On the other hand, students who received praise for their effort were more likely to have an interest in learning to master the material and attribute low performance to low effort rather than low intelligence [201].

While recognition as a physics person, who has what it takes to excel in the physics course, can help students excel in physics courses, negative recognition can be detrimental for students at any stage. This is especially true for marginalized students in physics such as women. For instance, Eileen Pollock, who graduated summa cum laude with a B.S. in physics from Yale in 1987, did not fulfill her dream of getting a Ph.D. in physics. In her memoir, she explains, "not a single professor—not even the adviser who supervised my senior thesis—encouraged me to go to graduate school." She continues, "Certainly this meant I wasn't talented enough to succeed in physics" [202]. She goes on to talk about other women's experiences. One woman painfully recalled her experience in her high school physics course where "the teacher announced that the boys would be graded on the "boy curve," while the one girl would be graded on the "girl curve"; and when asked why, the teacher explained that he couldn't reasonably expect a girl to compete in physics on equal terms with boys" [202]. This is an example of negative recognition. Studies have shown that without intentional strategies to create an equitable and inclusive learning environment, female students do not feel recognized appropriately even before they enter college [1, 4, 24], at least partly due to the societal stereotypes and biases about who belongs in physics and can excel in it that women are bombarded with over their lifetime and differential treatment of men and women by pre-college instructors, counselors, and other stakeholders.

Although a majority of prior studies have focused on physics courses in which women are underrepresented, the societal stereotypes and biases about who can excel in physics courses can negatively impact women even in physics courses in which they are not underrepresented, e.g., mandatory physics courses for bioscience majors. One common stereotype is that only high achievers or geniuses can succeed in physics [23]. However, genius is often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius [24]. Moreover, as these students get older, norms in the science curriculum become increasingly masculine and the existing curricula tend to represent the interest and values of female students less [1]. Also, teachers and school counselors often pay more attention to students from the dominant group, e.g., male students, and counselors give advice based on race and gender to students regarding high school physics and math courses to take as well as majors to pursue when in college [203, 204].

These stereotypes and biases are also prevalent at the university level. One study found that college biology and physics faculty members rated a male student as significantly more competent than a female student when presented with a hypothetical scenario for hiring a student for lab work with either a male or female name [51]. These stereotypes and biases and the differential treatment of women and men, e.g., by faculty members, can influence female students' perceptions about their ability to do physics before they enter the classroom. Thus, it is possible that although women are the majority in algebra-based physics courses primarily for bioscience majors, these societal stereotypes and biases can still influence their perceived recognition as a physics person, which can impact their performance as well as motivational beliefs in the physics classes, unless instructors make an explicit effort to create an equitable and inclusive learning environment.

This study examined the difference between male and female students with regard to the role of their perceived recognition as a physics person by instructors, TAs as well as friends and family in predicting performance outcomes at the end of a mandatory first semester of an algebra-based introductory physics course sequence for bioscience majors, controlling for their high school GPA, SAT math scores, and perceived recognition at the beginning of the course. We controlled for students' past performance in high school in order to isolate the effect of perceived recognition as a physics person on student outcomes. A visual representation of the path analysis of our final model is shown in Figure 10. All paths were considered from left to right in our model. However, only some of the paths are shown in Figure 10 for clarity. Our research questions are as follows:

RQ1. Are there gender differences in students' high school GPA and SAT math scores, as well as grades and perceived recognition as a physics person, at the beginning and end of the mandatory physics course?

RQ2. To what extent does the perceived recognition as a physics person predict physics course grade?

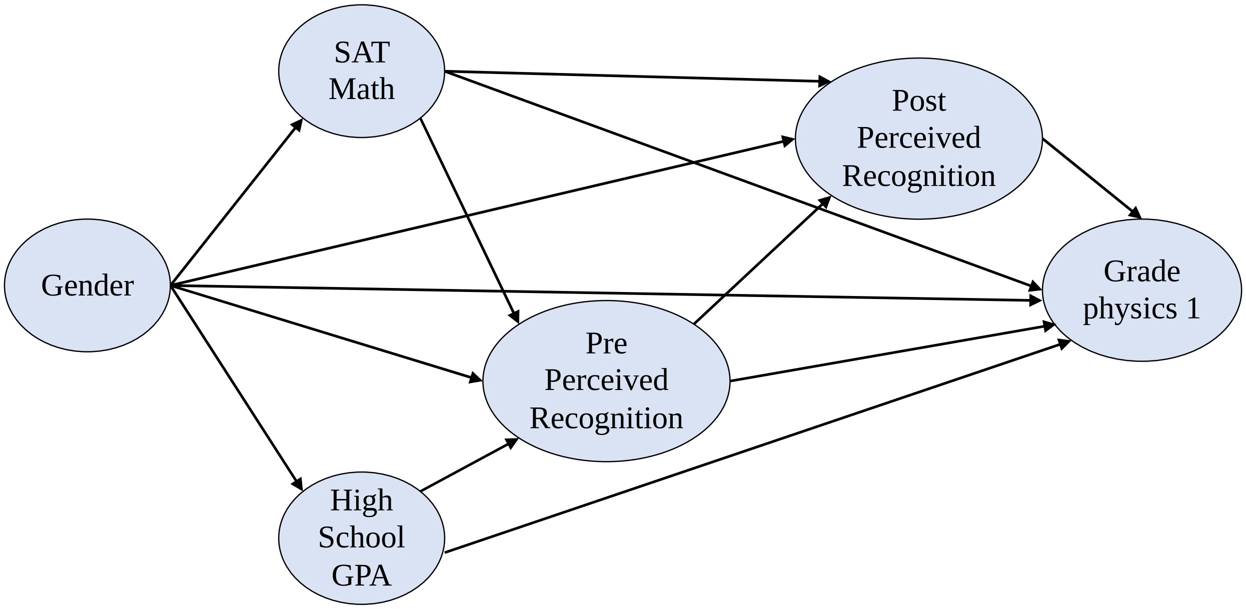


Figure 10 Schematic representation of the model and how perceived recognition mediates the relation between gender and grade in physics 1 course for bioscience majors. From left to right, all possible paths were considered (including the one from gender to grade in physics 1), however, only some of the paths are shown here for clarity.

5.2 Methodology

In this study, a validated survey covering perceived recognition and other motivational constructs (not discussed in this manuscript) [62] was administered to students at a large public research university in the U.S. The survey was given at the beginning (pre) and end (post) of the semester in the first introductory algebra-based physics courses over the course of two years. This course is primarily taken by junior or senior bioscience majors for whom a two-semester physics course sequence is mandatory. We analyzed the data for 827 students who completed the survey in the introductory physics 1 traditionally taught lecture style course. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the university data, the participants were 33% male and 67% female students. We recognize that gender is fluid and not a binary construct; however, the data collected by the institution was in binary terms. We use binary data provided by the university in this study. Less than 1% of the students did not choose male or female and thus were not included in the study.

5.2.1 Instrument validity

This study focused on the perceived recognition as a physics person of students enrolled in the first of the two introductory algebra-based physics courses primarily for bioscience majors. The validated survey involving several motivational beliefs, including the perceived recognition items, was adapted from previous research [6, 170, 175] and revalidated at our institution. The survey asked the students to answer all of the questions in the context of the physics course they were enrolled in. Specifically, the perceived recognition questions were taken from Godwin et. al. [6] and measure the extent to which a student believes that other people see them as a physics person or a person who can excel in physics [5, 59]. These physics recognition items have been used in prior studies in the general population of college students in English composition classes (a majority of students were in their first year) [5, 6, 199, 205] and in high school physics courses [200] as recognition from friends, family, and teachers.

Re-validation of the survey at our institution involved conducting one-on-one student interviews to ensure that the students interpreted the questions correctly [26], Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and calculation of the Pearson Correlations and Cronbach alpha. Our prior individual interviews also suggest that students interpreted the perceived recognition part of the question "see me as a physics person" to imply that they can excel in physics in the course they were enrolled in (the survey asked students to answer all of the questions in the context of the physics course). The questions in the study are on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173]. A lower score is indicative of a negative endorsement of the survey construct while a higher score is related to a positive endorsement. The survey questions for perceived recognition and factor loadings for each question from the CFA are given in Table 8. We note that all factor loadings are significant, but the factor loadings are larger for family and friends questions since they are more similar to each other and share a greater variance. The Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.85 for these perceived recognition as a physics person questions which is considered very reasonable [176].

Table 8 Physics perceived recognition survey questions along with factor loadings (Lambda) from the Confirmatory Factor Analysis (CFA) result for all students (N = 827). The rating scale for the perceived recognition as a physics person questions is strongly disagree, disagree, agree, strongly agree. All *p*-values (of the significance test of each item loading on to the construct or factor) are $p < 0.001$.

Physics Perceived Recognition	Lambda
My family sees me as a physics person.	0.88
My friends see me as a physics person.	0.91
My physics instructor and/or TA see me as a physics person.	0.65

5.2.2 Analysis

Initially, we analyzed the descriptive statistics by comparing female and male students' mean scores on the perceived recognition questions and students' grades for statistical significance using t-tests, and investigated the effect size using Cohen's d [182]. Cohen's d is $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students and σ_{pooled} is the pooled standard deviation for all students (or weighted standard deviation for men and women). In general, $d = 0.20$ indicates a small effect size, $d = 0.50$ a medium effect size, and $d = 0.80$ a large effect size [182].

To quantify the statistical significance and relative strength of our frameworks' path links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. The SEM is an extension of multiple regression and allows one to conduct several multiple regressions simultaneously between variables in one estimation model. This is an improvement over multiple regression since it allows us to calculate the overall goodness of fit and also allows for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [139].

The model estimates were performed using gender moderation analysis to check whether any of the relations between the variables show differences across gender by using "lavaan" to conduct multi-group SEM [187]. Initially, we tested different levels of measurement invariance. 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 [187]. Since we did not find significant moderation by gender, we tested the theoretical model in mediation analysis, using gender as a variable directly predicting items to examine the resulting structural paths between constructs.

5.3 Results

5.3.1 Gender differences in predictors and outcomes

We find that in the traditionally taught lecture style introductory physics course women had statistically significantly lower mean values in their SAT math scores, physics 1 grade, pre perceived recognition, and post perceived recognition as a physics person (see Table 9) while men had statistically significantly lower mean value in high school GPA (see Table 9). In addition, the gender gap in men and women's perceived recognition as a physics person increased from the beginning (pre: Cohen's *d* = 0.27) to the end (post: Cohen's *d* = 0.39) of the course.

Table 9 Mean predictor and outcome values by gender as well as statistical significance (*p*-values) and effect sizes (Cohen's *d*) by gender.

Predictors and Outcomes (Score Range)	Mean		*p-value*	Cohen's *d*
	Male	Female		
High School GPA (0-5)	4.00	4.17	<0.001	-0.36
SAT Math (200-800)	677	664	0.027	0.19
Pre Perceived Recognition (1-4)	2.13	1.97	<0.001	0.27
Post Perceived Recognition (1-4)	2.16	1.91	<0.001	0.39
Physics 1 Grade (0-4)	3.22	3.02	<0.001	0.27

5.3.2 SEM path model

We used SEM to investigate the relationships between the constructs and to unpack the students' perceived recognition by others as a physics person contribution to explaining the physics

1 average grade of women and men. We initially conducted gender moderation analysis between variables using multi-group SEM to investigate if any of the relationships between the constructs were different across gender. There were no group differences at the levels of weak and strong measurement invariance and the level of regression coefficients. Thus, there was no moderation effect by gender. Therefore, we proceeded to gender mediation to investigate the extent to which gender differences in physics 1 grade at the end of the course were mediated by differences in students' pre-college academic measures (high school GPA and SAT Math) and perceived recognition in physics.

The result of the path analysis part of the SEM is presented visually in Figure 11. The model fit indices indicate a good fit to the data: CFI = 0.958 (> 0.90), TLI = 0.932 (0.90), RMSEA = 0.077 (<0.08), and SRMR = 0.037 (<0.08). Gender had direct connections to SAT math (β = 0.09), high school GPA (β = -0.16), pre perceived recognition (β = 0.14), post perceived recognition (β = 0.09) and grade in physics 1 (β = 0.11).

Students' grade in physics 1 had direct connections to gender (β = 0.11), the high school factors including SAT math (β = 0.33) and high school GPA (β = 0.30) as well as post perceived recognition in physics (β = 0.21). Women also had lower perceived recognition as a physics person than men at the start and end of the physics course.

In addition, Appendix C (page 291) shows the same model as in Figure 11 except for the perceived recognition as a physics person only includes the question " My physics instructor and/or TA see me as a physics person". That model is very similar to the model in Figure 11 and students' post perceived recognition as a physics person predict their grades in physics 1 with a similar β value except high school GPA has a direct connection to post perceived recognition, gender did not have a direct connection to pre perceived recognition, and the β value from pre to post perceived recognition as a physics person was 0.38.

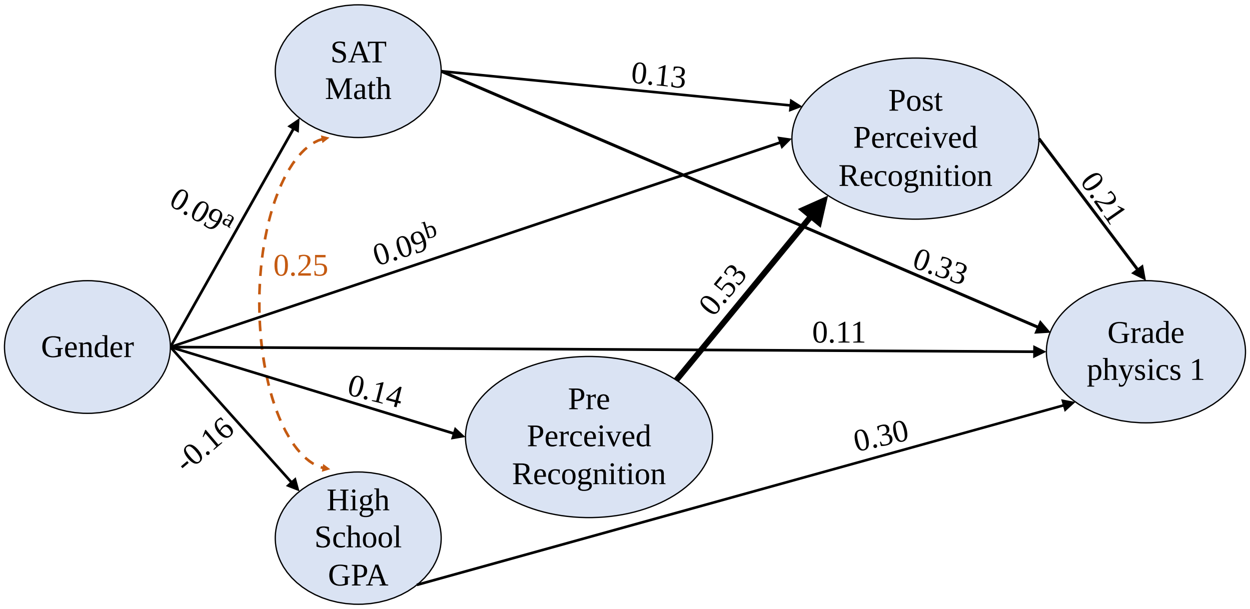


Figure 11 Result of the path analysis part of the SEM between gender and physics 1 grade through students' perceived recognition. The line thickness indicates the relative magnitude of β values. The dashed line indicates covariances between constructs. All p-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p = 0.021$, and superscript "b" for $p = 0.004$.

5.4 Summary and Discussion

In this research involving both descriptive and inferential analyses, we find gender gaps in physics grades and perceived recognition by others as a physics person disadvantaging women in the mandatory introductory physics 1 course for bioscience majors which was taught primarily using traditional lectures. We note that in the lecture style courses, students may receive recognition from their instructor or TA in different ways, e.g., the instructor or TA praises students who answer or ask questions during lecture or recitation. Our SEM models show that recognition by instructors and TAs as a physics person plays an important role in predicting students' grades at the end of the physics course. However, on average, women had lower grades and perceived recognition as a physics person at the beginning and end of the course than men even in these physics courses in which they are not the numerical minority in the class. This is concerning since women and men are receiving different cues from instructors and TAs about whether they can excel in the physics course. Since students' perceived recognition by others as a physics person predicts their grade at the end of the course, instructors/TAs must recognize all students as those who can excel in the physics course (and it is particularly important for students that have been traditionally marginalized in physics such as women).

The persistent gender gaps in perceived recognition by others as a physics person even at the end of this course may at least partly be due to the physics learning environment not being equitable and inclusive, with physics instructors and teaching assistants unwittingly reinforcing gender stereotypes about physics and communicating lower expectations for women. The instructors and TAs must reflect upon the fact that what is important is not what their intentions are but the impact they are having on the students. In particular, TAs and instructors must find meaningful ways to explicitly validate and recognize all students in their classes as physics people, those who can excel in the physics course. If instructors only recognize male students who are eager to ask questions and answer questions posed by instructors, this situation would amount to negative recognition for other students particularly those from stereotyped groups such as women even if they are not a numerical minority in the class [103]. For example, in a prior study, one student from the dominant group was selected to explain problems to the class more often than other students which led to other students feeling like they did not know the material [103]. Not receiving recognition may be more detrimental to women and other marginalized students in physics due to societal stereotypes and biases about who belongs in physics and can excel in it.

Moreover, if a student goes to an instructor after struggling with homework problems to ask for help and the instructor starts by saying that the problem is “trivial”, this type of comment can be perceived as negative recognition and insult. Again, this is especially true for students who are from historically marginalized groups, such as women (even if they are not a numerical minority in class), because they are always questioning whether they have what it takes to excel in physics due to societal stereotypes. For example, in an interview with women in physics courses, one woman mentioned that when some male physics professors consider her question trivial, she wonders, “ are they being condescending because they think I can’t do it, or because I am a woman?” [53]. If the students have worked hard on the problem before coming for help, such statements from the instructor about the problem being trivial are more likely to trap those students into thinking that they must not have what it takes to excel in physics and they also may infer (even if it is not intended) that the instructor must think that they are not capable.

One strategy to increase positive recognition as a physics person in the classroom includes instructors having high expectations for all students, but making it clear to students that all of them have what it takes to excel in the physics course if they work hard and use effective approaches to learning while instructors are there to support them to reach their goals as needed. Therefore, students should be supported so that they feel safe taking advantage of all of the resources including the instructor and TA office hours. Moreover, opportunities for student-centered learning where students can serve as leaders in problem solving and encouraging students to persist in their efforts by normalizing struggle and framing struggle as an important stepping stone to learning and developing a solid grasp of physics would help [60]. Other productive approaches to validating and recognizing students are for instructors/TAs to make themselves available to students as needed, discuss with students their own struggles when they were students, and affirm students when they make progress. In addition, in a study of a calculus-based physics 1 course, students' perceived recognition by instructors and TAs predicted their self-efficacy (students' belief about whether they can excel in physics) [62]. Therefore, if instructors can increase students perceived recognition, then students' self-efficacy may also be positively impacted. Additionally, it may be easier for instructors/TAs to create opportunities for students to receive positive perceived recognition, as opposed to increasing students' self-efficacy, which is critical for supporting students (particularly those from marginalized groups such as women). Due to societal stereotypes about who belongs in physics and excels in it, positive recognition as a person who can excel in physics and affirmation by instructors for making progress may be especially beneficial for women and other marginalized students in physics courses [206].

Instructors must create an equitable and inclusive learning environment in which all students feel validated and recognized as people who can excel in physics. In particular, if instructors create a learning environment where all students feel recognized by their instructors and TAs as those who can excel in physics, marginalized students such as women may benefit most and be able to more effectively engage in the course material, tackle challenging physics problems, and group work may become more productive with positive interdependence among all students.

6.0 Students' sense of belonging in introductory physics course for bioscience majors predicts their grade

6.1 Introduction and Theoretical Framework

Motivational beliefs in academic domains can influence students' continuation in related courses, majors, and careers [26, 32, 39, 41, 54, 155-159]. In particular, students' sense of belonging has been shown to be important in their academic outcomes and future careers [207, 208]. Students' sense of belonging in their college or university has been shown to predict students' intentions to persist through college [209, 210]. In science, technology, engineering, and math (STEM) fields, studies have shown that belonging is linked to students' engagement in the courses [211], grade [212], persistence [213, 214], and lower expectations of dropping out of their STEM major [215]. For example, in one study, students' sense of belonging in a math class was found to predict their intention to pursue math in the future [216]. However, it has also been found that women have a lower sense of belonging than men, potentially due to stereotypes related to who belongs and can excel in STEM fields [216, 217].

These stereotypes about who can excel in STEM courses could impact women even in physics courses in which they are not underrepresented, e.g., mandatory college physics courses for bioscience majors. One common stereotype is that genius and brilliance are important factors to succeed in physics [23]. However, genius is often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius [24]. Moreover, as these students get older, norms in the science curriculum hold less relevance for girls since the existing curricula tend not to represent the interests and values of girls as much [1]. Furthermore, teachers and school counselors often pay more attention to male students and counselors give gendered advice to students regarding which high school physics and math courses to take and majors to pursue when in college. All these factors which are associated with societal stereotypes about who belongs in physics and can excel in it can influence female students’ perceptions about their ability to do physics even before they enter the college classroom. Also, it is possible that although women are the majority in algebra-based physics courses primarily for bioscience majors, these societal stereotypes can still influence their sense of belonging and grades in the physics class unless instructors make an explicit effort to create an equitable and inclusive learning environment.

One mechanism by which societal stereotypes and biases negatively affect female students is via lowering their sense of belonging in fields in which they are underrepresented. For example, in one study, women STEM graduate students had the perception that they exerted more effort than their peers to succeed, and this perception predicted their sense of belonging [218]. Another study showed that female participants in a Physics Olympiad competition who endorsed negative stereotypes about female talent for physics felt a lower sense of belonging in physics [105]. In addition, when active learning is implemented in STEM courses in which learning environments are not equitable or inclusive, men have been shown to dominate responding to questions from instructors in classes [219]. Students’ sense of belonging has been shown to correlate with their retention and self-efficacy in school [50, 63, 220]. In addition, it is shown to be a predictor of students’ physics identity for senior physics majors [58]. Several prior studies have shown performance gaps between men's and women’s grades and scores on conceptual tests in calculus-based and algebra-based physics courses [8-13] that some hypothesize may be due to gender gap in prior preparation or motivational beliefs, such as sense of belonging [12-22]. However, there has not been research conducted on students' sense of belonging and how it predicts students' grades in physics courses in which women are not underrepresented.

This study examined the difference between male and female students with regard to the role of their sense of belonging in predicting performance outcomes at the end of a mandatory first semester of an algebra-based introductory physics course sequence for bioscience majors, controlling for their high school GPA, SAT math scores, and sense of belonging at the start of the class. We controlled for students' past performance in high school in order to isolate the effect of students' sense of belonging on their course outcomes. A visual representation of our final model is shown in Figure 12. All paths were considered from left to right in our model. However, only some of the paths are shown for clarity. Our research questions are as follows:

RQ1. Are there gender differences in students' high school GPA and SAT math scores, as well as grades and sense of belonging at the beginning and end of the mandatory physics course?

RQ2. To what extent does physics sense of belonging predict physics course grade in the model shown schematically in Fig. 1?

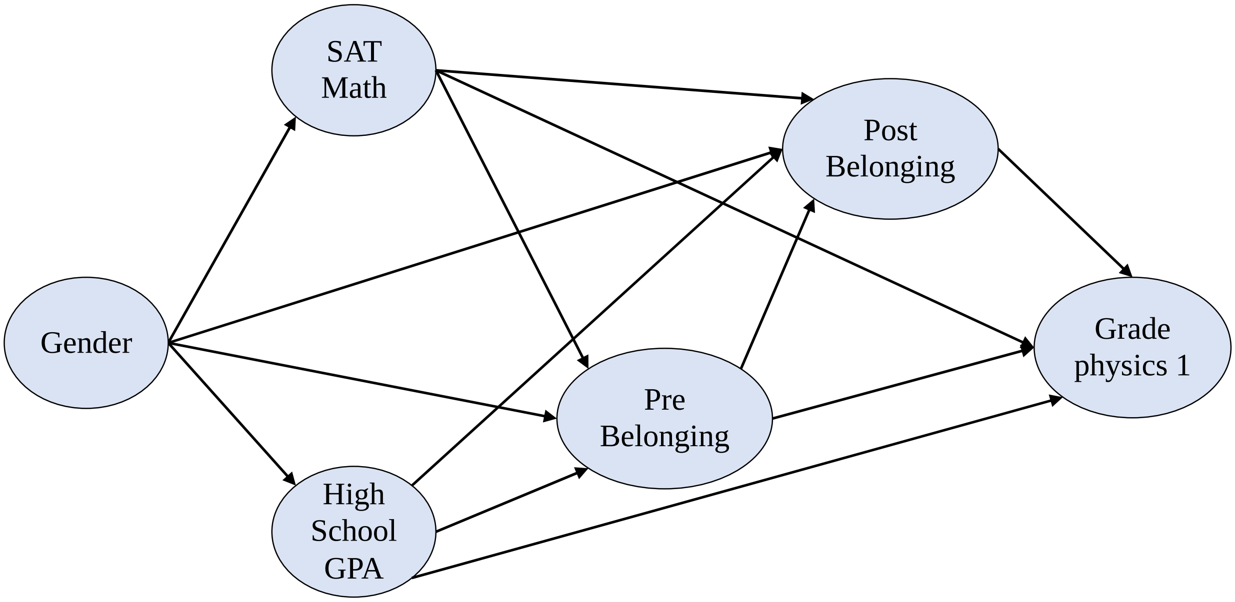


Figure 12 Schematic representation of the path analysis of the model and how sense of belonging mediates the relation between gender and grade in the physics 1 course for bioscience majors in which women are not underrepresented. From left to right, all possible paths were considered (including the one from gender to grade in physics 1), however, only some of the paths are shown here for clarity.

6.2 Methodology

In this study, a validated survey covering sense of belonging and other motivational constructs (not discussed here) was administered to students at a large public research university in the U.S. The survey was given at the beginning (pre) and end (post) of the first semester in the introductory algebra-based physics classes over the course of two years that were taught in the traditional in-person lecture style before the pandemic. This course is primarily taken by junior or senior bioscience majors for whom a two-semester physics course sequence is mandatory. We analyzed the data for 814 students who completed the survey in the introductory physics 1 class. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the university data, the participants were 36% male and 64% female. We recognize that gender is fluid and not a binary construct; however, the data collected by the institution is in binary terms. We use binary data provided by the university in this study. Less than 1% of the students did not choose male or female and thus were not included in the study.

We note that there were six sections and four instructors of the course within the investigated time period. However, we found that the intraclass correlation coefficient (ICC), which measures the proportion of the variance between instructors, e.g., in the student sense of belonging to be 4%. This is below 10%, the usual threshold cited for warranting the use of multi-level models and thus we grouped all the students together, regardless of instructor.

6.2.1 Instrument Validity

This study measured the physics sense of belonging of students enrolled in the first of the two introductory algebra-based physics courses primarily for bioscience majors. The survey with multiple motivational constructs was administered to the students [62], however, we focus on the sense of belonging questions here. The validated survey involving several motivational beliefs, including the sense of belonging items, was adapted from previous research [170, 175]. Re-validation of the survey at our institution involved conducting one-on-one student interviews [26], Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and calculation of the Pearson Correlations and Cronbach alpha. The sense of belonging items on the survey measured whether students felt like they belonged in the introductory physics class [50, 175]. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 5 (high endorsement) [173]. A lower score is indicative of a lower sense of belonging while a higher score is indicative of a higher sense of belonging. Two of the items were reverse-coded. The survey items for the sense of belonging factor and factor loadings for each item from the CFA, performed on the entire survey, are given in Table 10. The Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.85 for the sense of belonging items, which is considered reasonable [176].

Table 10 Survey items for students' sense of belonging along with factor loadings (Lambda) from the Confirmatory Factor Analysis (CFA) result for all students (N = 814). The rating scale for the items was not at all true, a little true, somewhat true, mostly true, and completely true. All *p*-values (of the significance test of each item loading) are $p < 0.001$.

Physics Belonging Items	Lambda
I feel like I belong in this class.	0.77
I feel like an outsider in this class.	0.77
I feel comfortable in this class.	0.77
Sometimes I worry that I do not belong in this class.	0.75

6.2.2 Analysis

Initially, we analyzed the descriptive statistics by comparing female and male students' mean scores on the belonging questions and students' grades for statistical significance using t-tests and investigated the effect size using Cohen's d [182]. Cohen's d is $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students and

σ_{pooled} is the pooled standard deviation (or weighted standard deviation for men and women) for all students. In general, d = 0.20 indicates a small effect size, d = 0.50 indicates a medium effect size, and d = 0.80 indicates a large effect size [182].

To quantify the statistical significance and relative strength of our framework's path links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. SEM is a statistical method consisting of two parts that are completed together; a measurement part which consists of CFA and a structural part that consists of path analysis. Path analysis can be considered an extension of multiple regression analysis, but it allows one to conduct several multiple regressions simultaneously between variables in one estimation model and allows us to predict multiple outcomes simultaneously. SEM also allows us to calculate the overall goodness of fit and for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. Thus, we are able to test more complicated models than we would with multiple regression analysis. A full SEM model combines this path analysis with CFA, allowing researchers to test the validity of their constructs (using CFA) and the connections between these constructs (using path analysis) in a single model with a single set of fit indices. We report the model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [139].

Initially, we performed gender moderation analysis by conducting multi-group SEM, i.e., the model estimates were performed separately for men and women to check whether any of the relations between variables show differences across gender by using "lavaan" [56]. In particular, our moderation analysis was similar to our mediation model in Fig. 1 but there was no link from gender, instead, multi-group SEM was performed separately for women and men simultaneously.

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

When the model is more complex than the preceding example of smoking and lung cancer as in our SEM model (which has a measurement part involving CFA and a structural part involving path analysis), checking to make sure there are no gender moderation effects involves checking that there are no gender moderation effects for both the measurement and structural parts. For the measurement part, to check for measurement invariance in each step of gender moderation analysis, we fixed different elements of the measurement part of the model to equality across gender and compared the results to the previous step when they were allowed to vary between groups (i.e., for women and men) separately using the Likelihood Ratio Test [187]. A non-significant p-value at each step indicates that the fit of this model is not appreciably worse than that of the model in the previous step, so the more restrictive invariance hypothesis (when the parameters are set to the same values for women and men) is retained. Therefore, setting those different elements of the measurement part of the model to equality across gender is valid, which means that estimates are not statistically significantly different across groups (i.e., women and men).

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

Similar to “weak” and “strong” measurement invariance for the measurement part, when testing moderation effect in path analysis, the predictive relationship (regression path) between two variables is tested for the two groups (e.g., women and men) simultaneously. If the predictive relationship is different for the groups (i.e., the values of the regression coefficients (β) are not the same for the predictive relationship between the two constructs for women and men), then there is a gender moderation effect in the model. If moderation does not show differences by gender in any of these steps (measurement invariance holds and testing for regression coefficients shows that they can be set equal for women and men), we can utilize a gender mediation model (see Figure 12). In other words, we can interpret our model the same way for both men and women, and any gender differences can be modeled using a separate gender variable.

In our multi-group SEM model, we found a non-significant p-value in each step, and thus measurement invariance holds and the regression coefficients for women and men can be set equal, i.e., there are no moderation effects by gender (for men and women) in our models. Thus, we concluded that our SEM model can be interpreted similarly for men and women and we can use gender mediation analysis (instead of doing moderation by gender). Therefore, we tested the theoretical model in mediation analysis, using gender as a variable (1 for male and 0 for female) directly predicting items to examine the resulting structural paths between constructs (a schematic representation of the path analysis for the gender mediation model is shown in Figure 12). In the mediation analysis, if there are paths from gender to any of the constructs as we found in our results discussed in the next section, it implies that women and men did not have the same average value for those constructs controlling for all constructs to the left. However, it is important to note that all of the item factor loadings and regression coefficients between the constructs are the same for women and men (as found from the gender moderation analysis which preceded the mediation analysis).

6.3 Results

6.3.1 Gender differences in predictors and outcomes

We find that women had statistically significantly lower mean values in their SAT math scores, physics 1 grade, pre belonging, and post belonging (see Table 11) while men had statistically significantly lower mean values in high school GPA (see Table 11). In addition, the gender gap in men and women's sense of belonging increased from the beginning (pre: Cohen’s d = 0.29) to the end (post: Cohen’s d = 0.44) of the course. We also investigated the difference between men's and women’s sense of belonging from the pre-test to the post-test. We found that women’s sense of belonging was not statistically significantly different from the beginning of the class to the end of the class (Cohen's d = 0.06, p = 0.340) while men’s increase in their sense of belonging was statistically significant (Cohen's d = 0.27, p = 0.001). We note that Cohen’s d indicates a small to medium effect size. However, small and medium effect sizes can still be important for instructors and researchers to focus on considering this is in a physics course in which women make up two-thirds of the class and women had higher high school GPAs than men. While the mean values and effect sizes (Cohen’s d) indicate that women have a lower sense of belonging than men, Table 11 cannot be used to make inferences about the relationships between various factors in the table. Therefore, we used SEM in the next section to show the relationships between the constructs.

Additionally, in Appendix D (page 292), we provide the percentage of men and women who selected each response to the items. This provides a sense of how students shifted their answers from the beginning to the end. From Table 43 in Appendix D (page 292), we can see that in general, the percentage of women who selected 1 or 5 increased from pre to post and the percentage of men who selected 5 increased from pre to post, while the percentage of men who selected 3 decreased. In addition, more women than men selected choices 1 (not at all true) and 2 (a little true), while men were more likely to select the answers 4 (mostly true) and 5 (completely true).

Table 11 Mean predictor and outcome values by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender.

Predictors and Outcomes	Mean		*p-value*	Cohen's *d*
(Score Range)	Male	Female		
High School GPA (0-5)	3.96	4.13	<0.001	-0.34
SAT Math (200-800)	682	656	<0.001	0.38
Pre Belonging (1-5)	3.52	3.28	<0.001	0.29
Post Belonging (1-5)	3.74	3.34	<0.001	0.44
Physics 1 Grade (0-4)	3.24	3.06	0.003	0.22

6.3.2 SEM path model

We used SEM to investigate the relationships between the constructs and to unpack the sense of belonging contribution in explaining the physics 1 grade of women and men. We initially conducted gender moderation analysis between variables using multi-group SEM to investigate if any of the relationships between the constructs were different across gender. There were no group differences at the levels of weak and strong measurement invariance and the level of regression coefficients. Thus, there was no moderation effect by gender. Therefore, we proceeded to gender mediation to investigate the extent to which gender differences in physics 1 grade at the end of the course were mediated by differences in students' pre-college academic measures (high school GPA and SAT Math) and sense of belonging in physics.

The result of the path analysis part of the SEM is presented visually in Figure 13. The model fit indices indicate a good fit to the data: CFI = 0.934 (> 0.90), TLI = 0.903 (0.90), RMSEA = 0.078 (<0.08), and SRMR = 0.033 (<0.08). Gender had direct connections to SAT math (β = 0.16), high school GPA (β = -0.17), pre belonging (β = 0.14) and post belonging (β = 0.18). However, gender did not have any direct connections to the grade in physics 1. Instead, gender was mediated through students' high school factors and their sense of belonging in the course. To expand further, the statistically significant path from gender to pre belonging means that men are predicted to have higher mean values in their sense of belonging than women controlling for their SAT math scores and high school GPA. The reason that there is no direct path from gender to students' grades in physics 1 is that the gender differences in students' grades (Table 11) are statistically non-significant when controlling for the other constructs in the model (SAT math, high school GPA, pre and post belonging).

Students' grade in physics 1 had direct connections to their high school factors of high school GPA (β = 0.25) and SAT math (β = 0.29) as well as post sense of belonging in physics (β = 0.33). To clarify, a standardized regression coefficient of β = 0.33 between post sense of belonging and physics grade in Figure 13 implies that for one standard deviation increase in the post sense of belonging, the grade would show 0.33 standard deviation change controlling for gender, SAT scores, high school GPA and pre belonging at the beginning of the course. Women also have lower scores in sense of belonging than men at the start and end of the physics course. Since the students' sense of belonging can predict students' grade at the end of the course, it is important for instructors to improve students' sense of belonging in the course by making their classes equitable and inclusive.

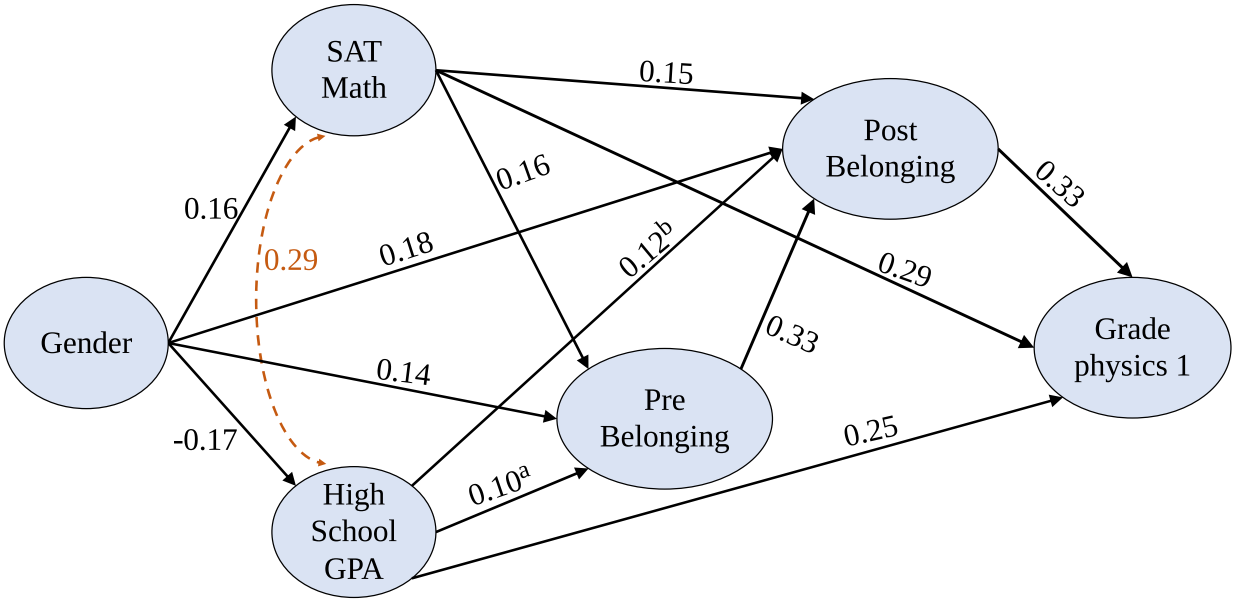


Figure 13 Result of the path analysis part of the SEM between gender and physics 1 grade through students' sense of belonging. The line thickness indicates the relative magnitude of β values. The dashed line indicates covariances between constructs. The gender variable was coded as 1 for men and 0 for women, so paths from gender with $\beta > 0$ indicate a higher mean for men while $\beta < 0$ indicate a higher mean for women in the predicted variable. All *p*-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p = 0.014$, and superscript "b" for $p = 0.001$. Gender does not directly predict the physics grade.

6.4 Summary and Discussion

In this research involving both descriptive and inferential quantitative analyses, we find gender gaps in physics grades and sense of belonging (Table 11) disadvantaging women in the mandatory introductory physics 1 course for bioscience majors in which women are not outnumbered by men. This is supported by a statistically significant value of Cohen's d between women and men for both sense of belonging and grade with small to medium effect sizes (Table 11). Our SEM model shows that students' sense of belonging plays an important role in predicting students' grades at the end of the physics course. Gender also directly predicts students' SAT Math scores, high school GPA, pre belonging, and post belonging. In addition, while men's sense of belonging statistically significantly increases from the beginning to the end of the course, women's sense of belonging does not statistically significantly increase by the end of the course despite women outnumbering men in the course. Since students' sense of belonging can have important implications not only for students' grades in the class but also for their retention in their major and future career trajectories, instructors teaching these physics courses must create an equitable and inclusive learning environment and increase women's sense of belonging.

Instructors have a central role to play in increasing students' sense of belonging in the introductory physics courses. In order to increase students' sense of belonging particularly for students who are underrepresented, some recommendations researchers have made for physics instructors are to send messages that concerns about belonging are normal and fade with time, identify and temper cues that perpetuate the "geeky" scientist stereotype, and openly endorse effort and hard work over brilliance [106]. One study showed that teacher practices, including their encouragement of cooperative activities in an inclusive environment, were related to students' engagement in the course and a high sense of belonging in the classroom [221]. Instructors could provide time for students to work in groups during class or recitations in a supportive environment, making sure that all student voices are heard equally while discussing problems. These types of pedagogical approaches could in turn help increase students' sense of belonging in the classroom if they observe that their ideas are respected by other students in the classroom. This is especially important when implementing active learning pedagogies in the classroom. If not implemented using teaching strategies that are equitable and inclusive, men have been shown to dominate responding to questions from instructors in STEM classes and women have reported lower scientific self-efficacy [219].

In addition, values-affirmation interventions that typically have students write about core personal values can have positive effects on students in academic settings [222]. The values-affirmation interventions may be more beneficial for students from underrepresented groups [223]. Moreover, belonging interventions that focus on improving the sense of belonging of underrepresented students in STEM courses have been found effective. They have been found to create an inclusive learning environment that can mitigate their doubts about social belonging in college and improve the grades of racial and ethnic minority students [192] and women [74, 151] in STEM fields. Lastly, inclusive mentoring, as well as contact with experts and peers who share similar demographic characteristics in academic fields, may be able to protect people, especially individuals who are underrepresented in the field, from the negative effects of stereotype threat and has the potential to increase their sense of belonging [73].

7.0 Women have lower physics self-efficacy and identity even in courses in which they outnumber men: A sign of systemic inequity?

7.1 Introduction and Theoretical Framework

While prior research has investigated issues pertaining to women's representation in science, technology, engineering, and math (STEM) fields [19, 77, 114, 115, 145-153, 163, 224], most studies in college physics courses have focused on courses in which women are outnumbered by men. Moreover, in explaining participation and learning in STEM disciplines, student motivational beliefs have been shown to play an important role [54, 218, 225-227]. However women have been shown to have lower motivational beliefs than men [26, 160, 228, 229]. In particular, students' identity has been shown to play an important role not only in students' in-class participation and performance but also in their choices of future courses and careers [5, 49, 54-57]. Identity in this context, e.g., physics identity, refers to students' views about whether they see themselves as a "physics person" or a person who can excel in physics [2, 5, 61, 230].

Here we use a version of the physics identity framework developed by Hazari et al. [5], which adapted the science identity framework by Johnson et al. [131]. The science identity framework by Johnson et al. [55] includes three dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others"). Hazari et al. modified the framework specifically for physics. "Competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a good physics student. Lastly, a fourth dimension, interest, was added to the framework since students have highly varying levels of interest in physics [60, 61]. In future studies by Hazari for introductory students, performance and competence are combined into one variable [58]. In a slightly reframed version of Hazari's physics identity framework by Kalender et al. [59], the framework we use in our study, performance/competence was framed as self-efficacy (closely related to competency belief). Additionally, recognition was framed explicitly as "perceived recognition" by students for clarity.

Self-efficacy in a discipline refers to students' belief in their ability to accomplish tasks or solve problems [27, 28]. It has been shown to influence students' engagement, learning, and persistence in science courses, in addition to contributing to students' science identity [14, 15, 29-42, 231, 232]. For example, when tackling difficult problems, students with high self-efficacy tend to view the problems as challenges that can be overcome, whereas those with low self-efficacy tend to view them as personal threats to be avoided [27]. However, in introductory physics courses in which women are underrepresented, studies have found a gender gap in self-efficacy disadvantaging women that widens by the end of the course, even in interactive engagement courses [29, 43]. Additionally, self-efficacy has been shown to predict students' engagement, learning, and persistence in science courses [14, 15, 27, 29-42].

Another motivational belief that influences physics identity is interest. Interest in a particular discipline may affect students' perseverance, persistence, and achievement [16, 38, 41, 44-46, 233]. One study showed that changing the curriculum to stimulate the interest of female students helped improve all of the students' understanding at the end of the year [47]. Within expectancy-value theory, interest and self-efficacy are constructs that predict students' academic outcomes and career expectations [48]. We focus on intrinsic interest or an individual's personal interest and enjoyment in engaging with physics concepts in this study.

The third belief, perceived recognition by others, has been shown to play an important role in a student's identity [49] and motivation to excel [50]. In a study on students' perception of support, teacher support was more strongly linked to the motivation and engagement of girls than boys [50]. However, studies have shown that female students are not recognized appropriately in many STEM disciplines even before they enter college [1, 4, 24]. One study found that science faculty members in biological and physical sciences exhibit biases against female students by rating male students significantly more competent [51]. Moreover, our prior research suggests that students' perceived recognition by instructors and teaching assistants can impact their self-efficacy and interest in physics and students' physics self-efficacy can impact their interest in physics (e.g., see [52, 53]).

Prior studies have shown that women have lower physics motivational beliefs, including identity, than men [1-3]. However, most of the studies in the college context concerning physics motivational beliefs are conducted in classes in which women are underrepresented [4, 5]. Nevertheless, pervasive societal stereotypes and biases about who can excel in physics that women are bombarded with from a young age could impact their physics motivational beliefs, including their identity, even in physics courses in which they outnumber men, e.g., introductory physics courses for bioscience majors in which women are not underrepresented. The students' physics motivational beliefs in these courses could influence not only their participation and performance in the physics courses, which is very important in itself, but also in other STEM disciplines in which they are intending to major or may be considering a career. For example, past studies have shown that the factors that influence students' physics identity influence their engineering career choices [6] and predict students' engineering identity [7] when the survey was administered in college English composition courses (where the majority of students were first year students with intentions to major in a variety of disciplines). Similarly, these physics identity factors could influence students in the physics course for bioscience and pre-health majors. In particular, we focus here on students who are bioscience majors primarily interested in health-related professions, for whom two physics courses are mandatory. The reason these physics courses are mandatory for them is not only because the foundational knowledge of physics is central to developing deep understanding of bioscience, but also because the kind of reasoning skills students develop in physics can help them in their majors and future careers. In fact, a majority of these bioscience majors are on the pre-health track and want to become health professionals. Higher physics motivational beliefs including an identity as a person who can excel in physics can help them engage and perform better, e.g., on the Medical College Admission Test (MCAT) in which 5-8% of the material consists of physics concepts, to realize their career goals.

Explicit and implicit societal stereotypes and biases that women internalize about physics could influence their motivational beliefs, including their physics identity. One common societal stereotype is that genius and brilliance are important factors to succeed in physics [23]. However, genius is often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius [24]. Studies have found that by the age of six, girls are less likely than boys to believe they are “really really smart” and less likely to choose activities that are made for “brilliant people” [24]. These types of stereotypes may continue to negatively impact women from childhood through college physics courses. In formal (K-12 education in which teachers and counselors often treat men and women differently and give them differential advice about courses to take) and informal situations (e.g., while interacting with family, friends, media and others), women are often made to feel that they cannot excel in physics-related fields. Additionally, negative societal stereotypes and biases may lead to a “chilly climate” for women in science classes, and lack of attention to making science curricula relevant to the interests of many female students [25]. These issues emanating from structural societal inequities pertaining to physics can lower the physics motivational beliefs of women compared to men.

Moreover, in prior investigations, researchers have often hinted at the fact that underrepresentation of women in those courses is likely related to their lower physics motivational beliefs than men and the gender gap increasing from the beginning to the end of those courses [26]. However, it is important to investigate whether in physics courses in which women are not numerically underrepresented, a similar gender gap in physics motivational beliefs exists and whether these gender gaps get worse from the beginning to the end of the courses. In particular, it is important to investigate whether numerical overrepresentation of women, e.g., in physics courses for bioscience majors we focus on here, would eliminate or significantly reduce the gender gap in physics motivational beliefs at the beginning and end of these courses. If this is not the case, it may signify that classroom representation alone will not change the pernicious effects of systemic gender inequities in physics perpetuated by society and bolstered further by the physics learning environments. It may point to the need to view these gender-gaps as more deep-rooted than simply due to women's underrepresentation in a physics course, as well as the need to address inequities in physics learning environments that exacerbate these gender gaps. Since physics motivational beliefs can impact student engagement and performance in physics courses and have the potential to impact their career choices, the deterioration of a gender gap in these courses would be markers of inequity that must be addressed.

Thus, our study focuses on the hypothesis that male and female students' physics motivational beliefs including their identity may be different even in courses in which women are not underrepresented and the physics learning environment in this situation may also exacerbate the situation and make the gender gap worse. Figure 14 shows the schematic representation of the identity framework for this investigation (identical to the one used for calculus-based introductory physics courses in which women are outnumbered by men [59]). In this framework shown in Figure 14, the relation between gender and physics identity is mediated by students' perceived recognition, physics self-efficacy, and interest [3, 6, 31, 234, 235]. We note that we have tested models which also have a direct path from gender to physics identity and we did not find it to be statistically significant. Therefore, that direct path from gender to physics identity is not shown for clarity.

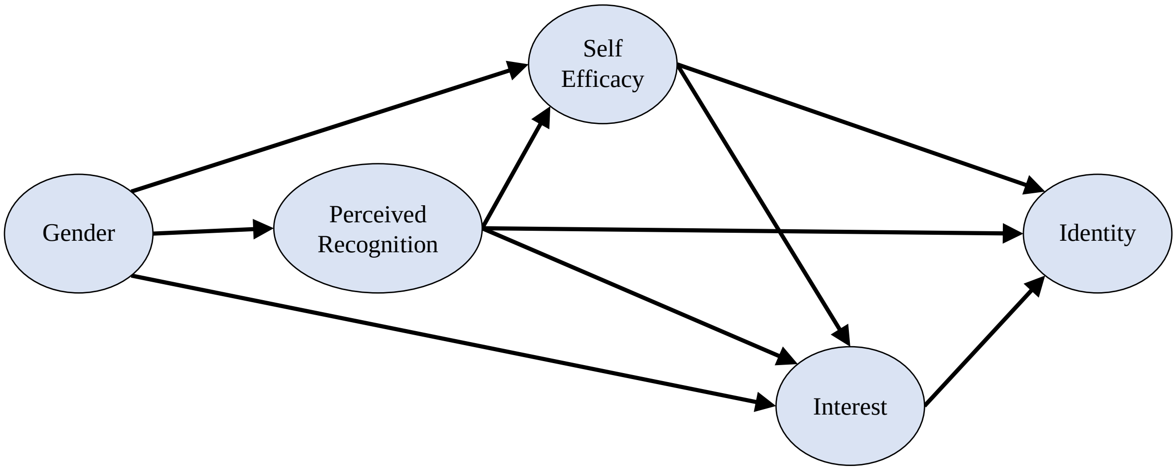


Figure 14 Schematic representation of the physics identity model and how perceived recognition, self-efficacy, and interest mediate the relation between gender and identity. From left to right, all possible paths were considered (including the one from gender to identity, although we do not show it here since we did not find it to be statistically significant).

In accordance with our framework, we investigated male and female students' self-efficacy, interest, and perceived recognition and how these motivational beliefs predict identity in introductory physics courses for bioscience majors in which women outnumber men. We used structural equation modeling (SEM) to investigate how the motivational beliefs that comprise physics identity relate to each other in this model. As noted earlier, it is not clear in the context of an introductory physics course in which there are majority women whether there are gender differences in these motivational beliefs and how perceived recognition, self-efficacy, and interest mediate male and female students' physics identity. From a statistical point of view, structural equation modeling can establish correlations among factors but cannot establish causal effects [187]. While past models have theorized the directionality between the constructs in Figure 14 in multiple ways [4, 6, 58], we chose this model because it has the ability to empower instructors so that they adopt effective practices, understand their role in empowering students, and recognize and affirm their work. Additionally, physics interest is not fixed and can increase or decrease depending on students' perceived recognition and self-efficacy, which in turn are dependent on how inclusive and equitable the physics learning environment is. Below, we first delineate the research questions based upon our framework, describe the methodology, then provide results and discussion and finally conclude with instructional implications and future directions. The following research questions were answered by analyzing data from a validated survey administered to students in large algebra-based physics courses at a research university in the US in which women outnumber men and using mediation analysis via SEM:

RQ1. Are there gender differences in the physics motivational beliefs (self-efficacy, interest, perceived recognition, and identity) and how do they change from the pre-test (at the beginning of the course) to the post-test (at the end of the course)?

RQ2. Can gender differences in students' physics identity be explained with gender differences in physics perceived recognition, self-efficacy, and interest at the end of an introductory algebra-based physics sequence?

7.2 Methodology

7.2.1 Participants

The participants who were administered the validated written survey were students at a large, public, research university in the U.S. The students were administered the survey at the beginning (pre) and end (post) of the first semester of a two-semester sequence in a traditionally taught introductory algebra-based physics course in which women outnumber men. We used data from 501 students who completed the survey on paper in the first week and the last two weeks of recitation class of about 30 students. The students completed the posttest in the week preceding the final exam for the course. These courses are typically taken by students primarily majoring in bioscience in their junior or senior year of undergraduate studies, with approximately 50%-70% of students expressing a desire to pursue future careers in health professions. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. The gender data provided by the university include only binary options of “male” and “female”. We recognize that gender is a socio-cultural and a nonbinary construct; however, the data provided by the university only included binary options (less than 1% of the students did not provide this information and thus were not included in this study). Based on the university data from the participants, 35% identified as male and 65% identified as female students. Thus, female students outnumber male students significantly in this physics class.

7.2.2 Instrument Validity

The survey items were constructed from items validated by others [170, 175, 236] and re-validated in our own context using one-on-one student interviews [26], exploratory factor analysis (EFA), confirmatory factor analysis (CFA) [182], analyzing the Pearson correlation between different constructs [182], and using Cronbach alpha [176] At the beginning of the survey, students were instructed to answer the questions on the survey with regard to the physics course they were in. The survey items asked about different motivational beliefs at the beginning and end of the course. These motivational constructs included students' physics identity (1 item), self-efficacy (4 items), interest (4 items), and perceived recognition (3 items). The *physics identity* question focuses on whether students see themselves as a physics person [5, 230, 237]. *The physics self-efficacy* questions measure students' confidence in their ability to understand and answer physics problems [6, 170-172, 230, 236]. The *interest in physics* questions measure students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *perceived recognition* questions measure the extent to which a student believes that other people see them as a physics person [230]. The interest questions were adapted from the Activation Lab Survey on science fascination/interest [171]. The first self-efficacy question was taken from the Peer Instruction self-efficacy instrument [172]. Three other self-efficacy questions were adapted from Godwin et. al.'s performance/competence questions with minor changes based upon individual interviews during the validation of the survey instrument at our institution [6]. The physics identity and perceived recognition questions were adapted from Godwin et. al. with no fine-tuning required based upon our interviews as students interpreted the questions as intended [6].

After performing EFA to ensure that the items factored according to different constructs as envisioned, a CFA was conducted to establish a measurement model for the constructs and used in SEM. The square of CFA factor loadings (lambda) indicates the fraction of variance explained by the factor. The model fit indices were good and all of the factor loadings (lambda) were above 0.50, which indicate good loadings [182]. The results of the CFA model are shown in Table 12. The Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.79 for the self-efficacy questions, 0.76 for interest questions, and 0.89 for perceived recognition questions which are considered reasonable [176].

Table 12 Survey questions corresponding to each of the motivational constructs, along with factor loadings from the Confirmatory Factor Analysis (CFA) for all students (N = 501). The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173]. The rating scale for some of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree. All p-values (of the significance test of each item loading) are $p < 0.001$.

Construct and Item	Lambda
Physics Identity	
I see myself as a physics person.	1.00
Physics Self-Efficacy	
I am able to help my classmates with physics in the laboratory or recitation.	0.59
I understand concepts I have studied in physics.	0.75
If I study, I will do well on a physics test.	0.76
If I encounter a setback in a physics exam, I can overcome it.	0.72
Physics Interest	
I wonder about how physics works.	0.57
In general, I find physics.†	0.75
I want to know everything I can about physics.	0.74
I am curious about recent discoveries in physics.	0.64
Physics Perceived Recognition	
My family sees me as a physics person.	0.89
My friends see me as a physics person.	0.92
My physics instructor and/or TA sees me as a physics person.	0.70

† the rating scale for this question was very boring, boring, interesting, very interesting.

Zero-order pair-wise Pearson correlations *r* are shown in Table 13. These Pearson r values signify the strength of the relationship between constructs. The inter-correlations vary in the strength of their correlation, but none of the correlations are so high that the constructs cannot be considered separate, consistent with prior studies [59]. The highest inter-correlation was between physics identity and perceived recognition (0.78), which is consistent with prior studies in calculus-based courses. Perceived recognition questions ask about external identity (perception of whether other people recognize an individual as a physics person), whereas the physics identity question asks about internal identity (whether an individual sees oneself as a physics person) so there tends to be a high correlation between these constructs. However, the correlation is low enough that they can be considered separate constructs.

Table 13 Pearson inter-correlations are given between all the predictors and outcomes for the post-test in Physics 1. All p-values < 0.001.

Pearson Correlation Coefficient				
Observed Variable	1	2	3	4
1. Perceived Recognition	--	--	--	--
2. Self-Efficacy	0.61	--	--	--
3. Interest	0.60	0.60	--	--
4. Physics Identity	0.78	0.59	0.62	--

In addition, we preformed item response theory (IRT) in order to check the response option distances for the survey constructs [177]. The parametric grades response model using STATA was used to test the measurement precision of our response scale. The two response scales were “NO! no yes YES!” and “strongly disagree, disagree, agree, strongly agree”. The parametric grades item response model calculates the location parameter for each response and calculates the difference between the locations. For the first group (NO! no yes YES!), the response scale discrimination values were 1.54 and 2.08. For the second set of answers (strongly disagree, disagree, agree, strongly agree), the response scale discrimination values were 1.51 and 1.26. The numerical values for the location differences need to be similar, as they are for both of these scales so we can use the means for these values [178]. In addition, we estimated IRT-based domain scores [177]. The correlation coefficient between the Likert mean scale values for each observed variable and the IRT-based domain scores were calculated to be 0.99 between interest questions, 0.98 between self-efficacy questions and 0.99 between perceived recognition and identity questions [177, 178, 181]. Since the factor scores derived from the IRT are so highly correlated with the mean score, using mean values for the scores is acceptable when analyzing the data [177, 179, 180]. In particular, because the psychological distance between adjacent response items and across items was approximately similar and complex factor scores derived from IRT or CFA are so highly correlated with the mean scores, it is reasonable to use mean scores [177, 178, 181]. Moreover, we have included frequencies of student responses for each question in Appendix E.

7.2.3 Analysis

Initially, we analyzed data using descriptive statistics and compared female and male students' mean scores on various constructs for statistical significance using *t*-tests and computed the effect sizes using Cohen's *d* [182]. In general, $d = 0.20$ indicated a small effect size, $d = 0.50$ indicates a medium effect size, and $d = 0.80$ indicates a large effect size [182]. For predictive relationships between different motivational constructs, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. The SEM is an extension of multiple regression. It conducts several multiple regressions simultaneously between different latent variables (or factors or constructs) in one estimation model and can have multiple outcome variables. This is an improvement over multiple regression since it can calculate overall goodness of fit and it allows for all estimates to be standardized simultaneously. This enables a direct comparison among different structural components, along with calculations of factor loadings for all factors (or constructs or latent variables). The SEM also has an option to handle missing data using the full estimation maximum likelihood “ML” estimation feature which improves both power and generalizability since it imputes missing data so that students only missing some data are not dropped. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMS). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and SRMR and RMSEA < 0.08 [139].

The model estimates were performed using gender moderation analysis to check whether any of the relations between variables show differences across gender by using “lavaan” to conduct multi-group SEM. Initially, we tested different levels of measurement invariance in the multi-group SEM model with gender moderation. 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. Since we did not find statistically significant moderation by gender, we tested the theoretical model in a gender mediation analysis, using gender as a variable directly predicting all latent variables to examine the resulting structural paths between constructs.

7.3 Results and Discussion

Pertaining to **RQ1**, Table 14 and Table 15 show that women had statistically significantly lower mean values than men for all motivational beliefs in our model. For example, both women's and men's scores were low in physics identity with women scoring below the negative value (2). Additionally, when we compare the scores in the pre-test and post-test (Table 16), the scores significantly change for women’s self-efficacy, interest, and physics identity as well as for men’s self-efficacy from when they enter the class until the end of the course. These differences increase throughout the semester and the gender gap widens by the end as evident from the larger Cohen's *d* values for women in Table 16. This occurs despite the fact that women outnumber men in this course. In addition, in Appendix E, we provide the percentage of men and women who selected each response to the questions. This provides a sense of how students shifted their answers from the pre-test to the post-test. From Table 44 and Table 45, we can see that in general both men and women shift from higher values to lower values for each motivational factor. For example, in the perceived recognition and identity questions, most students selected response 2 in the pretest while most students selected response 2 in the posttest. Additionally from Table 44 and Table 45, for the self-efficacy and interest questions, the overall trend is that the percentage of students who answered 3 or 4 decreased from the pretest to the posttest and the percentage of students who answered 1 or 2 increased from the pretest to the posttest.

Table 14 Mean Pre Predictor and Outcome Values by Gender and Effect Sizes (Cohen's *d*). All p-values < 0.001.

Predictors and Outcomes	Mean		Cohen's *d*
	Men	Women	
Perceived Recognition	2.19	2.00	0.32
Self-Efficacy	3.07	2.83	0.58
Interest	2.74	2.45	0.58
Physics Identity	2.16	1.87	0.46

Table 15 Mean Post Predictor and Outcome Values by Gender and Effect Sizes (Cohen's *d*). All p-values < 0.001.

Predictors and Outcomes	Mean		Cohen's *d*
	Men	Women	
Perceived Recognition	2.21	1.92	0.43
Self-Efficacy	2.90	2.55	0.66
Interest	2.66	2.27	0.69
Physics Identity	2.03	1.64	0.59

Table 16 Cohen's *d* (*d*) for Men and Women's Outcomes from the Pre-Test to the Post-Test. Statistical significance refers to p-values.

Predictors and Outcomes	Men		Women	
	d	*p-value*	*d*	*p-value*
Perceived Recognition	0.05	0.686	0.13	0.128
Self-Efficacy	0.33	0.005	0.54	< 0.001
Interest	0.15	0.203	0.34	< 0.001
Physics Identity	0.18	0.137	0.34	< 0.001

Pertaining to **RQ2**, we used SEM to investigate the relationships between the constructs and to unpack each construct's contribution to explaining the physics identity of women and men. We initially tested gender moderation between different constructs using multi-group SEM (between male and female students) and investigated whether the relationships between the different motivational constructs were different across gender. There were no group differences at the level of weak and strong measurement invariance, including no difference at the level of regression coefficients. Therefore, we proceeded to gender mediation analysis to understand how gender mediates physics identity at the end of the semester in the introductory physics course.

The results of the SEM are presented visually in Figure 15. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): CFI = 0.981 (> 0.90), TLI = 0.974 (> 0.90), RMSEA = 0.043 (<0.08), and SRMR = 0.030 (< 0.08). All three of the mediating constructs (perceived recognition, self-efficacy, and interest) predict physics identity at the end of the physics course similar to past results [5, 59] in other contexts. Perceived recognition has the largest direct effect ($\beta = 0.59$) with smaller effects from self-efficacy ($\beta = 0.12$) and interest ($\beta = 0.19$). Additionally, gender predicts perceived recognition (P.R.), self-efficacy, and interest. However, the relation between gender and physics identity is mediated only by the mediating constructs, and after accounting for these indirect paths, there is no direct path from gender to identity. In other words, women appear to have a lower physics identity because they have lower perceived recognition, self-efficacy, and interest. These results are very similar to those for calculus-based introductory physics courses [59] in which women are severely underrepresented and the majority of students are engineering majors. In addition, in Appendix F, we present data for how the factors that predict students' physics identity also predict students' science identity. Although it is not one of our main research questions, Figure 37 shows that physics self-efficacy also predict students' science idenity.

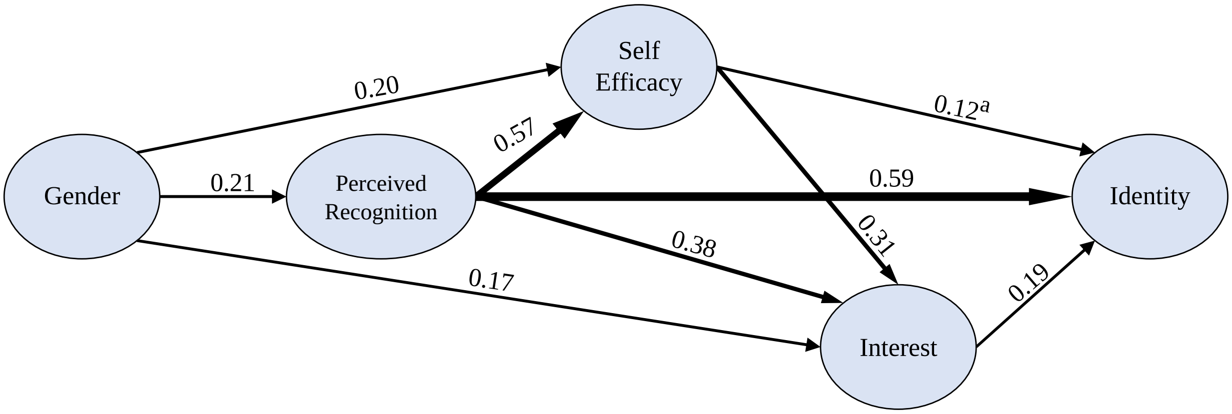


Figure 15 Result of the path analysis part of the SEM showing mediation between gender and physics identity through perceived recognition (P.R.), self-efficacy, and interest. The line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. All p-values for β are indicated by no superscript for $p < 0.001$ and "a" for $p = 0.013$. Gender does not directly predict physics identity.

7.4 Summary, Implications, and Future Directions

In this research involving both descriptive and inferential quantitative analyses, we find gender gaps in physics motivational beliefs disadvantaging women in mandatory introductory physics courses for bioscience majors in which women are not outnumbered by men similar to what has been found earlier in introductory calculus-based courses in which women are severely underrepresented [5, 59]. In particular, prior research in calculus-based courses in which women are underrepresented shows that compared to men, women have a greater decrease in their motivational beliefs, e.g., about whether instructors and TAs see them as people who can excel in physics [59]. We find that these findings are true even in algebra-based physics courses in which women are not underrepresented (e.g., both men and women have a mean recognition below the positive lower threshold, i.e., score of 3, and women score significantly lower than men). Table 14 and Table 15 show that women had a greater decrease in physics perceived recognition, self-efficacy, interest and identity compared to men. The inferential analysis using SEM in Figure 15 shows how perceived recognition, self-efficacy and interest mediate students' physics identity. We find that the gender differences in students' perceived recognition, self-efficacy and interest predict their identity as a person who can excel in physics and the correlations between these motivational beliefs predicting students' physics identity are similar to those in calculus-based introductory courses in which women are underrepresented [5, 59]. Moreover, there is no direct path from gender to physics identity, similar to the calculus-based courses [59].

Our model provides support for the physics identity framework in a new context and helps us understand the role of different motivational factors in predicting physics identity of women and men in algebra-based physics courses where women are not underrepresented. This is important since identity in a particular discipline is context-dependent and factors that influence physics identity can relate to each other differently in different contexts. Additionally, the model shows that TAs and instructors could play a critical role to increase students' self-efficacy, interest, and identity in physics. Our findings suggest that women feel less recognized by their instructors and TAs than men which influence women's self-efficacy, interest, and identity in physics.

The existence of a gender gap at the beginning of the physics course may be due to a variety of reasons including societal stereotypes, experiences in previous science courses and lack of female physics role models in media. The exacerbation of the gender gaps in motivational beliefs such as physics perceived recognition, self-efficacy, interest and identity from the beginning to the end of the physics course for bioscience majors may be a sign of inequity and can disadvantage women in terms of their participation, engagement and outcomes in the physics courses. These persistent gaps can also impact their choice of majors and future careers in STEM. These increased gender gap in these beliefs at the end of the physics courses may signify inequity and non-inclusive nature of learning environment. These trends in gender gap in physics motivational beliefs are highly troubling and require further investigation of their cause because they may signify that classroom representation alone will not change the pernicious effects of systemic gender inequities in physics perpetuated by society and bolstered further by the physics learning environments. For example, the increased motivational belief gaps may at least partly be due to physics instructors and teaching assistants unwittingly reinforcing gender stereotypes about physics and communicating lower expectations for women. If female students are not given the same type of positive feedback as male students, this could have a negative effect on perceived recognition and self-efficacy of women. Changing the narrative, increasing students' sense of belonging [186] and making the physics learning environment equitable and inclusive, e.g., by not letting men dominate the conversations in class and affirming students more when they make progress in the course has the potential to decrease the gender gap in student's perceived recognition and other motivation beliefs.

A limitation of our quantitative study focusing on descriptive and inferential quantitative analysis is that we can only gain insight into the relative values of motivational beliefs of men and women at the beginning and end of the course and how the relation between gender and physics identity is mediated by physics perceived recognition, self-efficacy and interest but we cannot establish causal effects. Therefore, future studies should investigate factors in the physics learning environment that exacerbate gender gaps in motivational beliefs even in these courses in which women outnumber men. In addition, future work can investigate approaches to improving student motivational beliefs in these types of physics courses in which women are not underrepresented and investigate whether approaches that are effective in these courses would also be effective in courses in which women are underrepresented. One potential approach for improving students' motivational beliefs is through brief social-psychological classroom interventions, e.g., mindset/sense of belonging interventions, that have been shown to be effective in boosting women's grades in some science courses [151, 152, 185, 186].

8.0 A longitudinal analysis of women and men’s beliefs in a two-semester introductory physics course sequence for students on the bioscience track

8.1 Introduction and Theoretical Framework

Domain specific beliefs, e.g., self-efficacy and identity, in science, technology, engineering, and math (STEM) domains can influence students' participation and persistence in those fields [43, 117, 118, 155, 156, 238-245]. For example, students were more likely to take courses or pursue a career in science if they have higher competency beliefs or self-efficacy [32, 39, 41, 155, 156], display higher interest in science [157], or have a higher science identity [54, 158, 159]. However, a gender gap disadvantaging women in student domain specific beliefs [26, 29, 97, 160, 229] has been observed in many STEM courses, including physics courses. Most of the research in physics has focused on courses in which women are underrepresented. However, in a study of introductory bioscience courses, in which women made up 60% of the course, women participated less and did not perform as well as men [246]. These trends may signify gender inequities in STEM learning environments even when women are not underrepresented. Since physics is a discipline with one of the worst stereotypes about who belongs in it and can succeed in it [23], it is particularly important to investigate inequitable gender gaps in beliefs (including self-efficacy, interest, recognition by others, and identity) in physics courses in which women are not underrepresented.

Domain-specific identity has been shown to play an important role, particularly in students’ participation in courses and their professional choices [5, 49, 54-57]. Identity in this context refers to identifying with an academic domain, like physics, or a student’s view about whether they see themselves as a “physics person” [2, 5, 61, 230]. Prior studies have shown that it can be more difficult for women to form a physics identity than men [1-3, 6]. However, most of the prior studies concerning physics identity and factors that influence it have been conducted in classes in which women are underrepresented. It is important to investigate physics identity in a variety of classes and contexts since there can be differences in each given context. For example, in one study on female undergraduate students, their sense of belonging predicted physics identity for upper-level students but not for introductory students [58]. Therefore, we investigated physics identity and student beliefs such as self-efficacy, interest, and perceived recognition (which have been shown to be related to physics identity in other contexts) in algebra-based physics courses for students on the bioscience track in which women are not underrepresented.

One factor hypothesized to influence physics identity is self-efficacy. Self-efficacy in a particular academic domain is defined as a person’s belief that they can succeed in a particular task, activity, or course related to that domain [27, 28]. According to Badura’s social cognitive theory, self-efficacy may be derived from four sources: mastery experiences, vicarious learning experiences, social persuasion experiences, and emotional state [27]. Mastery experience means that experiences with successful completion of a task should have a strong positive influence on one's confidence to complete a similar task. Vicarious learning experiences occur when observing someone else’s success on a task can influence students’ own belief in their ability to perform a similar task. Social persuasion experiences mean, e.g., that verbal suggestions from others such as words of encouragement can positively impact one’s self-efficacy. Lastly, one’s emotional state can act as a mediating source to amplify or undermine one’s confidence in one’s ability. Sawtelle et. al. found that predicting the probability of passing an introductory physics course for women relies primarily on the vicarious learning experiences source while it primarily relies on mastery experiences for men [31]. Students' self-efficacy has been shown to impact students' engagement, learning, and persistence in science courses as well as contribute to students' science identity [14, 15, 29-42]. For example, when tackling difficult problems, students with high self-efficacy tend to view the problems as challenges that can be overcome whereas people with low self-efficacy tend to view them as personal threats to be avoided [27]. However, in introductory physics courses in which women are underrepresented, studies have found a gender gap in self-efficacy favoring men that widens by the end of the course even in interactive engagement courses [29, 43].

Another factor hypothesized to influence physics identity is interest. According to the four-phase model of interest development [247], people's interest in a field is triggered and maintained by external factors first and then becomes a sustained individual interest. Interest in a particular discipline may affect students' perseverance, persistence, and achievement [16, 38, 41, 44-46]. One study showed that changing the curriculum to stimulate the interest of the female students helped improve all of the students' understanding at the end of the year [47]. Within expectancy value theory, interest and competency beliefs (closely related to self-efficacy) are connected constructs that predict students' academic outcomes and career expectations [48]. We focus on intrinsic interest, which is an individual's personal interest and enjoyment in engaging with a topic.

The third factor hypothesized to influence physics identity, perceived recognition, has been shown to be an important factor [50]. Our prior individual interviews suggest that students' perceived recognition as a physics person by instructors and teaching assistants (TAs) predicts their self-efficacy and interest in physics (e.g., see [52, 53]). When women enter STEM fields, they do not feel validated and recognized as often as men [196-198]. In a study on students' perception of support, teacher support was more strongly linked to the motivation and engagement of girls than boys [50]. Hazari et. al. found that in high school physics classes, students' physics identity, predicted by students' recognition as a physics person by others, was influenced by teachers' social cues, or how teachers' actions obscured social boundaries between teachers and students [200]. In another study [201], students who received praise for their intelligence were more likely to have a fixed mindset and lower levels of task persistence, enjoyment, and performance than students who received praise for their effort. Prior studies have also shown that female students are not recognized appropriately even before they enter college [1, 4, 24]. For example, one of the stereotypical views in science is that it is for high achievers or naturally gifted students [1], and in general, being a genius or exceptionally smart is attributed to boys [169].

In this study, we use a physics identity framework to investigate whether the relation between gender and physics identity was mediated by self-efficacy, interest, and perceived recognition by others longitudinally in introductory physics 1 and physics 2, which are mandatory for students on the bioscience track. We build on the physics identity framework developed by Hazari et al., which adapted the science identity framework by Carlone and Johnson [55]. The science identity framework by Carlone and Johnson [55] includes three dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others"). Hazari et al. modified the framework specifically for physics. "Competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a physics person. Lastly, a fourth dimension, interest, was added to the framework since students can have highly varying levels of interest in physics [60, 61]. In recent studies by Hazari et al. of introductory students, performance and competence are combined into one variable [58]. In a slightly reframed version of Hazari et al.'s physics identity framework by Kalender et al. [59] that we use, performance/competence was framed as self-efficacy (closely related to competency belief). Figure 16 shows the schematic representation of the model used here. In this framework, students’ physics identity has been shown to be influenced by their self-efficacy, interest, and perceived recognition [3, 6, 31, 234, 235].

Stereotypes and biases about who belongs in physics and can excel in it are one of the absolute worst out of all of the STEM disciplines [23] and this could be impacting women even in physics courses in which they are not underrepresented. Boys and girls are exposed to these fixed intelligence views starting from an early age [24]. Prior studies have found that by the age of six, girls are less likely than boys to believe they are “really really smart” and less likely to choose activities that are made for “brilliant people” [24]. As these students get older, these stereotypes can negatively impact women. One study found that science faculty members in biological and physical sciences exhibit biases against female students by rating a male student with an identical resume to a female student significantly more competent and being more willing to hire the male student and pay him more [51]. All of these stereotypes and biases can influence female students' perception of their ability to engage in physics problem solving even before they enter the physics classroom. Therefore, it is possible that although women are the majority in algebra-based physics courses for students on the bioscience track, these societal stereotypes can still influence their outcomes and identity in these courses.

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

Therefore, we administered a validated survey about physics beliefs in algebra-based physics courses and used mediation analysis in SEM to investigate progression in physics identity and beliefs (self-efficacy, perceived recognition, and interest) that mediate the relation between gender and physics identity in introductory physics 1 and physics 2. We recognize that there can be multiple statistically equivalent SEM models with the same constructs mediating the relationship between gender and physics identity. We chose our particular statistically equivalent SEM model in which perceived recognition precedes self-efficacy and interest in which the potential instructional implications (positive recognition can play a role in supporting students' self-efficacy and interest) can empower instructors/teaching assistants to positively recognize students and make their classes more equitable and inclusive. Thus, our model puts recognition by instructors and TAs first before self-efficacy and interest. Figure 16 shows the schematic representation of the physics identity framework for this investigation since no gender moderation effects were found (identical to the one used for calculus-based introductory physics courses [59]). We fit the model with the data collected at the end of physics 1 and physics 2, and compare the predictive relationships among the constructs in the two courses. Our research questions are as follows for the study spanning the two-semester course sequence:

- **RQ1.** Are there gender differences in the student physics beliefs (self-efficacy, interest, perceived recognition, and identity) and do they change from physics 1 to physics 2 for bioscience students?
- **RQ2.** Can gender differences in students' physics identity at the end of physics 1 and physics 2 for bioscience students be explained with gender differences in physics perceived recognition, self-efficacy, and interest?
- **RQ3.** How do the predictive relationships between the student beliefs such as self-efficacy, interest, and perceived recognition mediate the relation between gender and physics identity and how do they change between physics 1 and physics 2 for bioscience students?

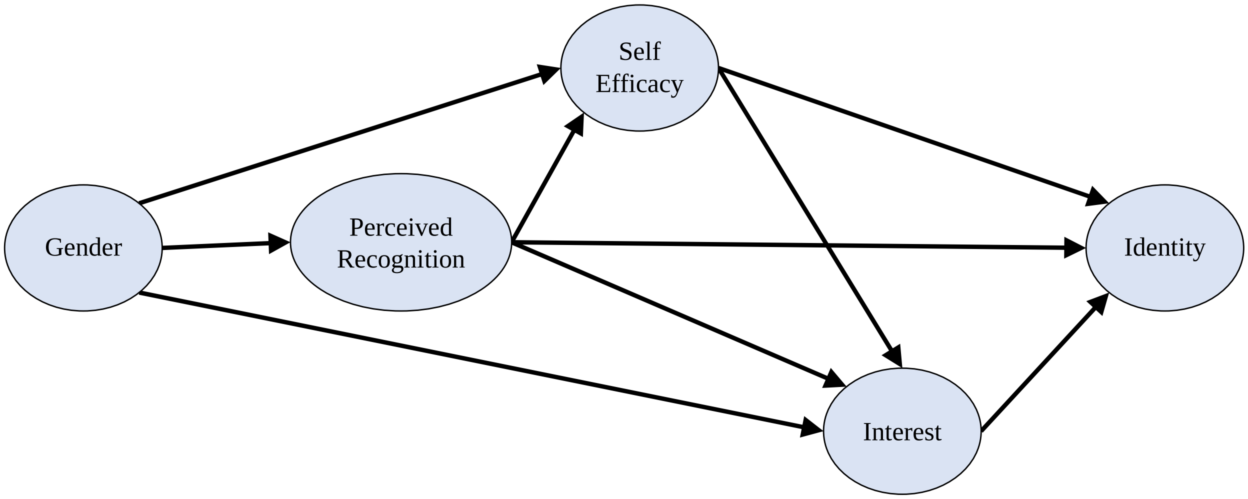


Figure 16 Schematic representation of the model based on the theoretical framework. From left to right, all possible paths were considered. Some, but not all, of the regression paths are shown.

8.2 Methodology

8.2.1 Participants

We analyzed data from a physics belief survey of 563 matched students (the same students took the survey at the end of physics 1 and 2) that was administered at a large public research university at the beginning and end of two semesters of introductory algebra-based physics 1 and 2 over the course of two years. The lecture-based physics courses were taught by 5 male instructors and 1 female instructor in physics 1 and 3 male instructors in physics 2. The classes consisted of 3 hours of lecture per week taught by the instructor and 1 hour of recitation per week taught by a teaching assistant (TA) (in which, typically, the students could ask questions about the homework or course material and complete group work). The classes were similar in terms of their grading policy in that students' grades heavily consisted of 2-3 midterm exams and a final exam involving primarily quantitative problems. However, using hierarchical modeling, we found the intraclass correlation coefficient (ICC) [248], which measures the proportion of the variance in the different constructs investigated in our model between instructors to be at most 4%. This is below 10%, the usual threshold cited for warranting the use of multi-level models and thus we grouped all the students together, regardless of instructor. We also note that the difference between the course taught by the female instructor and male instructors was comparable to those between different male instructors.

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

8.2.2 Instrument Validity

The survey items were constructed from items validated by others [170, 175, 236] and re-validated in our own context using one-on-one student interviews [26], exploratory factor analysis (EFA), confirmatory factor analysis (CFA) [182], analyzing the Pearson correlation between different constructs [182], and using Cronbach alpha [176]. At the beginning of the survey, students were instructed to answer the questions on the survey with regard to the physics course they were in. The survey items asked about different beliefs at the beginning and end of the course. These constructs included students' physics identity (1 item), self-efficacy (4 items), interest (4 items), and perceived recognition (3 items). The *physics self-efficacy* questions measured students' confidence in their ability to answer and understand physics problems [170-172, 230, 236]. The self-efficacy questions are mainly drawing on students' mastery experiences. The *interest in physics* questions measured students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *perceived recognition* questions measured the extent to which the students thought that other people see them as a physics person [230]. The *physics identity* question evaluated whether the students see themselves as a physics person [5, 230, 237]. The physics identity instrument only included one question, which is consistent with past studies since it has been challenging to make other questions that factor in this category in exploratory factor analysis [5, 6, 249, 250].

The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173]. A lower score is indicative of a negative endorsement of the survey construct while a higher score is related to a positive belief of the construct. Some of the questions were reverse coded (e.g., I feel like an outsider in this class). A CFA was conducted to establish a measurement model for the constructs and used in SEM. The square of CFA factor loadings (lambda) indicates the fraction of variance explained by the factor. The model fit indices were good and all of the factor loadings (lambda) were above 0.50, which indicates good loadings [182]. The results of the CFA model are shown in Table 17. The survey questions for each construct and factor loadings for each question from the CFA are given in Table 17. The Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.79 for the self-efficacy questions, 0.76 for interest questions, and 0.89 for perceived recognition questions, which are considered reasonable [176].

Table 17 Survey questions corresponding to each of the constructs, along with factor loadings from the Confirmatory Factor Analysis (CFA) for all matched students ($N = 563$) in physics 1 and physics 2. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173]. The rating scale for most of the self-efficacy and interest questions was: NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was: strongly disagree, disagree, agree, strongly agree. All p-values (for the significance test of each item loading) are $p < 0.001$.

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

† the rating scale for this question was very boring, boring, interesting, very interesting.

Zero-order pair-wise Pearson correlations of the average values of each construct are given in Table 18. Pearson's *r* values signify the strength of the relationship between variables. The inter-correlations vary in the strength of their correlation, but none of the correlations are so high that the constructs cannot be separately examined. The only high inter-correlation was the value between physics identity and perceived recognition (0.83). Perceived recognition questions ask about external identity, whereas physics identity asks about internal identity so there tends to be a high correlation between the constructs; however, the correlation is low enough that they can be considered separate constructs.

Table 18 Pearson inter-correlations are given between all the predictors and outcomes for the post survey in physics 2. All *p*-values < 0.001.

Pearson Correlation Coefficient				
Observed Variable	1	2	3	4
1. Perceived Recognition	--	--	--	--
2. Self-Efficacy	0.56	--	--	--
3. Interest	0.57	0.59	--	--
4. Physics Identity	0.83	0.56	0.54	--

8.2.3 Analysis

Initially, we compared female and male students' mean scores for each construct for statistical significance using *t*-tests and for the effect sizes using Cohen's *d* [182]. To quantify the significance and relative strength of our framework's links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. SEM is a statistical method consisting of two parts that are completed together; a measurement part which consists of CFA and a structural part which consists of path analysis. Path analysis can be considered an extension of multiple regression and conducts several multiple regressions simultaneously between variables in one estimation model in addition to the measurement part involving CFA. This is an improvement over multiple regression since it allows us to calculate the overall goodness of fit and allows for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [139].

Initially, we performed gender moderation analysis by conducting multi-group SEM, i.e., the model estimates were performed separately for men and women to check whether any of the relations between variables show differences across gender by using “lavaan” [187]. In particular, our moderation analysis was similar to our mediation model in Figure 16 except there was no link from gender, instead, multi-group SEM was performed separately for women and men simultaneously.

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

When the model is more complex than the preceding example of smoking and lung cancer as in our SEM model (which has a measurement part involving CFA and a structural part involving path analysis), checking to make sure there are no gender moderation effects involves checking that there are no gender moderation effects for both the measurement and structural parts. For the measurement part, to check for measurement invariance in each step of gender moderation analysis, we fixed different elements of the measurement part of the model to equality across gender and compared the results to the previous step when they were allowed to vary between groups (i.e., for women and men) separately using the Likelihood Ratio Test [187]. A non-significant *p*-value at each step indicates that the fit of this model is not appreciably worse than that of the model in the previous step, so the more restrictive invariance hypothesis (when the parameters are set to the same values for women and men) is retained. Therefore, setting those different elements of the measurement part of the model to equality across gender is valid, which means that estimates are not statistically significantly different across groups (i.e., women and men).

First, we tested for “weak” measurement invariance, which determines if survey items have similar factor loadings for men and women. We compared two models, one in which the factor loadings (which represent the correlation between each item and its corresponding construct) for women and men were predicted independently, and the other in which the factor loadings were forced to be equal between the groups (i.e., for women and men). Next, we tested for “strong” measurement invariance, which determines if survey items have similar factor loadings as well as similar intercepts [187] for men and women. Similar to weak invariance testing, we compared the models in which these factors were allowed to vary between groups separately for women and men and when they were set equal for women and men. If measurement invariance passes the weak and strong invariance test, i.e., there is no statistically significant difference between models when those parameters for women and men are set equal, then we must check for differences in the path analysis part, i.e., regression coefficients (β) among different latent variables in the model between women and men. This is because differences between the groups could occur at the factor (latent variable) level in regression coefficients (β).

Similar to "weak" and "strong" measurement invariance for the measurement part, when testing moderation effect in path analysis, the predictive relationship (regression path) between two variables is tested for the two groups (e.g., women and men) simultaneously. If the predictive relationship is different for the groups (i.e., the values of the regression coefficients (β) are not the same for the predictive relationship between the two constructs for women and men), then there is a gender moderation effect in the model. If moderation does not show differences by gender in any of these steps (measurement invariance holds and testing for regression coefficients shows that they can be set equal for women and men), we can utilize a gender mediation model (see Figure 16). In other words, we can interpret our model the same way for both men and women, and any gender differences can be modeled using a separate gender variable as in Figure 16.

In our multi-group SEM model, we found a non-significant *p*-value in each step, and thus measurement invariance holds and the regression coefficients for women and men can be set equal, i.e., there are no moderation effects by gender (for men and women) in our models. Thus, we concluded that our SEM model can be interpreted similarly for men and women and we can use gender mediation analysis (instead of doing moderation by gender). Therefore, we tested the theoretical model in mediation analysis, using gender as a variable (1 for male and 0 for female) directly predicting items to examine the resulting structural paths between constructs (a schematic representation of the path analysis for the gender mediation model is shown in Figure 16). In the mediation analysis, if there are paths from gender to any of the constructs as we found in our results discussed in the next section, it implies that women and men did not have the same average value for those constructs controlling for all constructs to the left. However, it is important to note that all of the item factor loadings and regression coefficients between the constructs are the same for women and men (as found from the gender moderation analysis which preceded the mediation analysis).

8.3 Results

To answer **RQ1** we analyzed the means of the beliefs in our model. According to Table 19 and Table 20, women had statistically significantly lower mean values than men in all constructs in our model. Both women and men scored lowest in physics identity and women scored below 2. Although we focus on the beliefs at the end of the course, we have included data for students' beliefs at the beginning (pre) and end (post) of the course in Appendix G (page (note the pre-survey in physics 1 did not include the identity or perceived recognition constructs and thus in order to match pre and post survey responses, only data for the second year of survey administration for physics 1 are included). In addition, the percentages of men and women who selected each choice for each survey item are included in Appendix H.

In order to investigate the effect of the introductory physics course on the gender differences in student beliefs, multiple One-way Analysis of Covariance (ANCOVAs) were conducted to determine whether there was a statistically significant gender difference in each post belief controlling for the pre belief. The results show that there was a significant effect of gender on students' post interest (F(1, 383) = 1.85 (p = 0.002)), post self-efficacy (F(1, 383) = 1.25 (p = 0.029)), and post perceived recognition (F(1, 382) = 1.96 (p = 0.010)) in physics 1 and on post interest (F(1, 419) = 0.84 (p = 0.019)) and post identity (F(1, 419) = 1.92 (p = 0.014)) in physics 2 controlling for the pre values in each case.

Table 19 Mean post predictor and outcome values in physics 1 by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender. 563 matched students are included with 361 women and 202 men. All p-values < 0.001.

Predictors and Outcomes	Mean		Cohen's *d*
	Male	Female	
Perceived Recognition	2.17	1.89	0.45
Self-Efficacy	2.98	2.73	0.50
Interest	2.81	2.37	0.73
Physics Identity	2.08	1.78	0.39

Table 20 Mean post predictor and outcome values in physics 2 by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender. 563 matched students are included with 361 women and 202 men. All p-values < 0.001.

Predictors and Outcomes	Mean		Cohen's *d*
	Male	Female	
Perceived Recognition	2.24	1.98	0.39
Self-Efficacy	2.93	2.71	0.42
Interest	2.77	2.29	0.78
Physics Identity	2.24	1.85	0.46

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

The results of the SEM for physics 1 are presented visually in Figure 17. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): CFI = 0.977 (> 0.90), TLI = 0.969 (> 0.90), RMSEA = 0.048 (<0.08), and SRMR = 0.033 (< 0.08). All three of the intervening variables (perceived recognition, self-efficacy, and interest) predict physics identity at the end of the physics course similar to past models [5, 59]. Perceived recognition has the largest direct effect with smaller effects from self-efficacy and interest. In addition to the direct effect, perceived recognition also indirectly predicts physics identity through self-efficacy and interest and self-efficacy indirectly predicts physics identity through interest.

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

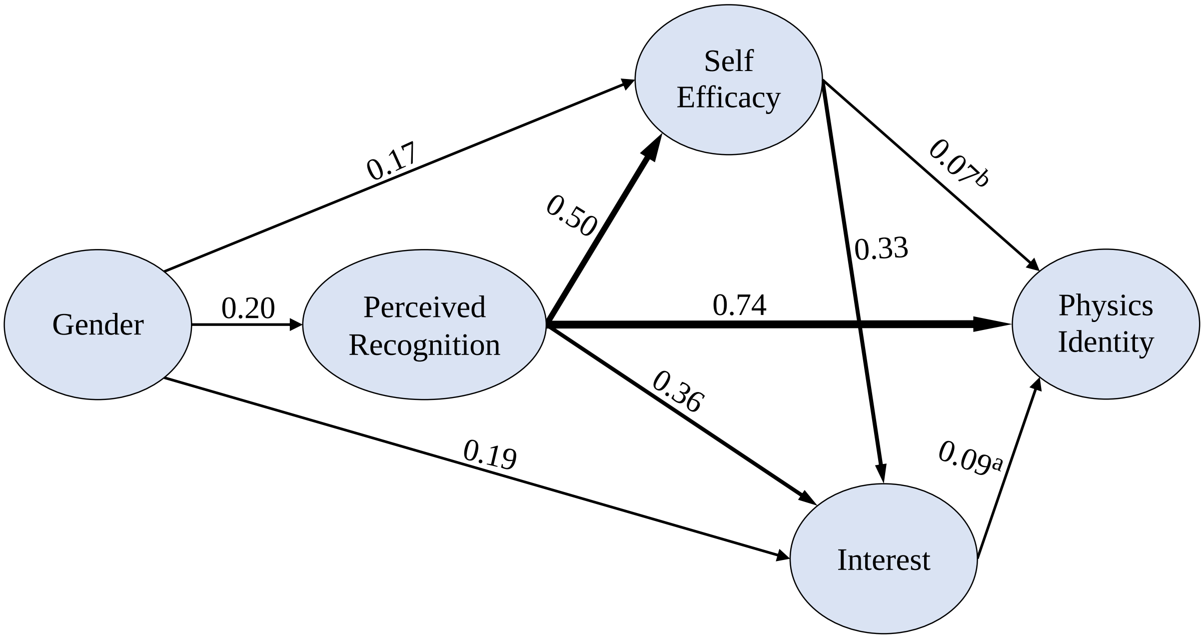


Figure 17 Result of the path analysis part of the SEM with the relationship between gender and physics identity being mediated through perceived recognition, self-efficacy, and interest for physics 1. The gender variable was coded as 0 for women and 1 for men. The line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. All *p*-values for β are indicated by no

superscript for *p* < 0.001, “a” for *p* = 0.037, “b” and for *p* = 0.114. Gender does not directly predict physics identity.

In order to understand how these relationships change over the introductory physics course sequence, the results of the SEM for physics 2 are presented visually in Figure 18. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): CFI = 0.972 (> 0.90), TLI = 0.963 (> 0.90), RMSEA = 0.055 (<0.08), and SRMR = 0.038 (< 0.08). Similar to the model in physics 1, all three of the intervening variables (perceived recognition, self-efficacy, and interest) predict physics identity at the end of the yearlong physics course sequence. The students' perceived recognition is the strongest predictor of physics identity ($\beta = 0.73$). In addition to the direct effect, perceived recognition also indirectly predicts physics identity through self-efficacy and interest and self-efficacy indirectly predicts physics identity through interest.

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

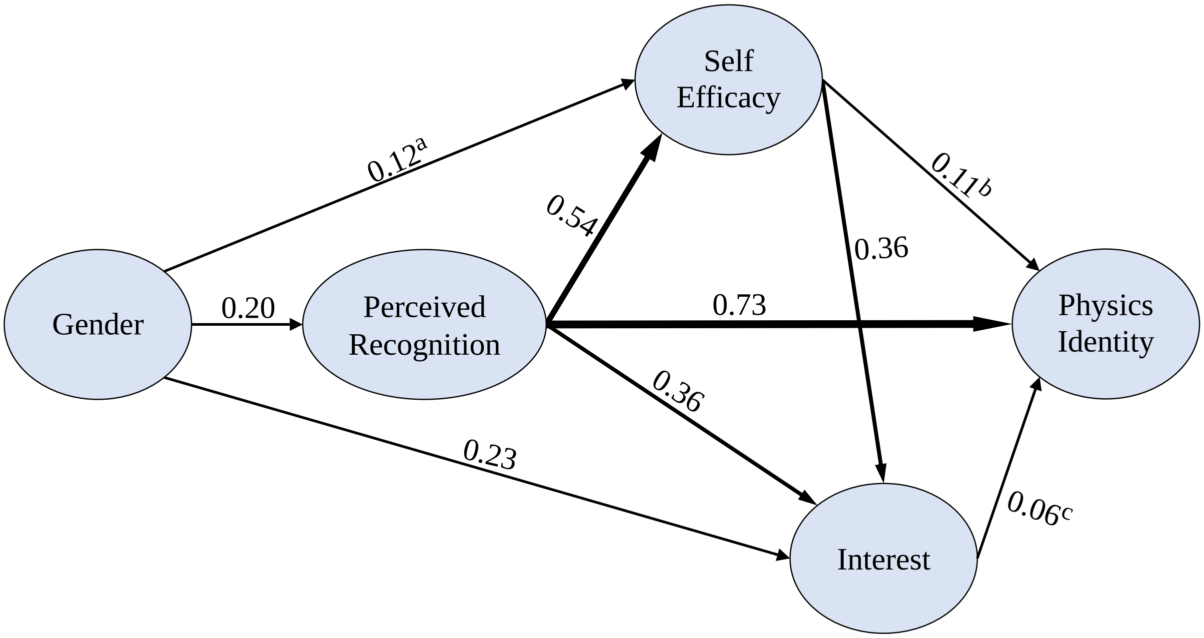


Figure 18 Result of the path analysis part of the SEM with the relationship between gender and physics identity being mediated through perceived recognition, self-efficacy, and interest for physics 2. The gender variable was coded as 0 for women and 1 for men. The line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. All p-values for β are indicated by no superscript for $p < 0.001$, "a" for $p = 0.003$, "b" for $p = 0.007$ and "c" for $p = 0.148$. Gender does not directly predict physics identity.

To answer **RQ3**, we compared the path analysis results of the SEM model in physics 1 (Figure 17) and physics 2 (Figure 18). Most of the pathways in each model did not change or changed by a small amount. The exception was that the effect of physics self-efficacy on identity becomes larger in physics 2 while the effect of physics interest on identity becomes smaller and nonsignificant in physics 2. We also ran a multigroup SEM analysis using the bootstrap technique for physics 1 and 2 and compared the regression paths in physics 1 and physics 2 using confidence intervals. The confidence intervals for each regression pathway are shown in Table 21. A prior study [251] suggests that the difference is statistically significant ($p \leq 0.05$) when the overlap of the 95% confidence intervals is no more than about half the average margin of error, that is when proportion overlap is about 0.5 or less. We use this rule to compare the regression paths in physics 1 and physics 2. For example, when comparing the confidence intervals for the regression pathway from perceived recognition (P.R.) to identity, the confidence interval is (0.79, 0.98) for physics 1 and (0.76, 0.93) for physics 2. The midpoint of the first confidence interval is 0.89, which is lower than the upper bound of the second confidence interval. Thus, the proportion of overlap between the confidence intervals is more than 0.5 and the regression pathways between physics 1 and 2 are not statistically significantly different from one another. None of the differences in the confidence intervals of the regression pathways (shown in Table 21) are statistically significantly different from one another and thus none of these regression pathways for physics 1 and physics 2 are statistically significantly different from one another.

Table 21 Confidence intervals for each regression pathway in physics 1 and physics 2. C.I is the confidence interval, P.R. is perceived recognition, Lower is lower bound and Upper is the upper bound. The proportion of overlap is above 0.5 for each of the regression pathway confidence intervals, and thus the regression pathways are not statistically significantly different from one another.

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

8.4 Summary and Discussion

In this study, we investigated women's and men's physics beliefs longitudinally in a two-semester introductory physics course sequence for students on the bioscience track. While other studies have investigated physics identity and other student beliefs in calculus-based physics courses in which women are underrepresented [59, 62], physics identity has not been investigated in physics courses in which women are not underrepresented. This new context is important since students' physics identity is context dependent and the factors that predict physics identity in a calculus-based physics course may not be the same in a course for students on the bioscience track. For example, in one recent study on female undergraduate students' physics identity, sense of belonging predicted physics identity for upper level students but not for introductory students [58]. Therefore, researchers have called for this type of research to be conducted in multiple contexts [58].

We find gender gaps in the beliefs disadvantaging women in both semesters of the physics course sequence. This finding is similar to prior research in calculus-based introductory physics courses in which women are underrepresented [59]. However, men's and women's beliefs do not significantly decrease over the course of the physics course sequence for students on the bioscience track, unlike in calculus-based physics courses (in which women's beliefs decrease more than men's) [26]. One potential reason is that women are not underrepresented in the physics courses for students on the bioscience track. However, gender inequity is maintained in these traditional lecture-based courses in which no explicit equity-related efforts are made, signifying that the learning environment is not equitable and inclusive. In addition, controlling for students' pre belief, gender differences were found in students' self-efficacy, interest, and perceived recognition at the end of physics 1 and in students' identity and interest at the end of physics 2.

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

Moreover, we find that our identity model using SEM for the two-semester introductory physics sequence suggests that perceived recognition, self-efficacy, and interest predict students' physics identity in a qualitatively similar manner in physics 1 and physics 2. This model is similar to models in past studies [5] and calculus-based physics courses in which women are underrepresented [58, 59, 62]. Our findings also suggest that perceived recognition as a physics person by others is the strongest predictor of physics identity and gender differences in perceived recognition disadvantage women's physics identity in both physics 1 and physics 2.

The strong contribution of perceived recognition in predicting physics identity throughout the two-semester course sequence shows that TAs and instructors can play a critical role to increase students' physics self-efficacy, interest, and identity. We note that the physics 1 and physics 2 courses in this study were traditionally taught lecture-based physics courses in which student grades heavily depended on two or three midterm exams and a final exam. The courses consisted of 3 hours of lecture per week taught by the instructor and 1 hour of recitation per week taught by a TA. It is important to recognize that even in these traditional lecture-based courses, students often receive feedback from their instructors in multiple ways, including receiving praise for asking or answering a question in class (which often advantages male students since they dominate these situations) and their interactions with students during office hours or over email. In addition, students interact with their TAs in recitation by asking questions about the homework or class material at the start of recitation, when completing group work during recitation, and during the TA’s office hours. However, the gender differences in these beliefs throughout the two-semester course sequence suggest that physics instructors and TAs are not doing enough to create an equitable and inclusive learning environment. Our view of an equitable and inclusive learning environment is that it should provide adequate support to all students and close the initial gaps, e.g., in the beliefs of students from different demographic groups such as female and male students discussed here. The focus on inclusivity of the learning environment is important since one study showed that the student perception of the inclusivity of the physics learning environment, consisting of students’ perceived recognition, sense of belonging, and perception of the effectiveness of interaction with their peers predicted students’ physics beliefs at the end of the course [62]. Thus, instructors and TAs should strive to make the learning environment in their classes equitable and inclusive.

One study showed that incorporating discussion sessions about the underrepresentation of women in physics improved women’s physics identities [252]. In addition, explicitly holding everyone to the same standards, not letting men dominate the conversations in class or office hours, and explicitly praising women when they do well or make progress on various components of the class could help decrease the gender gap in students' perceived recognition while also boosting their self-efficacy and interest. One strategy to increase students' recognition by instructors is to offer opportunities for student-centered learning where students can serve as leaders in problem solving and encouraging students to persist in their efforts by normalizing struggle and framing struggle as an important stepping stone to learning and developing a solid grasp of physics [60].

It is especially important for instructors to focus on equity and inclusion when implementing active-learning pedagogy. For example, one study showed that teacher practices, including their encouragement of cooperative activities in an inclusive environment, were related to students' engagement in the course [221]. In particular, if active-engagement pedagogies are not implemented using teaching strategies that are equitable and inclusive, men have been shown to not only dominate responding to questions in class but also while working in groups which can lower women's self-efficacy [219]. Additionally, instructors can improve equity and inclusion in their courses by making their courses student-centered, e.g., by adopting pedagogy that focuses on societal implications of physics [47] in addition to providing mentoring/support for students who are underrepresented [253]. Short social-psychological classroom interventions, e.g., sense of belonging and mindset interventions, have also been shown to positively impact students from underrepresented groups including women [151, 152, 185, 186]. Beyond the physics classroom, informal science activities, like participation in science fairs or talking science with family and friends could increase students' physics identity in college [254].

In summary, even in physics courses in which women are not underrepresented, women have lower physics beliefs, including physics identity than men. While having a larger cohort of women helps these constructs stay relatively similar over the two-semester course sequence compared to calculus-based courses [26], negative stereotypes about who belongs in physics and can excel in it disadvantage women throughout the yearlong physics sequence, hinting at lack of effort to create an equitable and inclusive learning environment in these physics classes. Instructors and TAs must make concerted efforts to not let men dominate in class and create a learning environment that emphasizes recognizing and validating all their students in these physics courses, particularly women and other underrepresented students who have been stigmatized due to societal stereotypes and biases about physics for too long.

8.5 Limitations

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

9.0 Self-efficacy and perceived recognition by peers, instructors, and teaching assistants in physics predict bioscience majors' science identity

9.1 Introduction

Students' science, technology, engineering, and mathematics (STEM) related motivational beliefs, e.g., self-efficacy and identity, have implications for their short and long-term outcomes in those disciplines regardless of their performance. Identity frameworks in STEM disciplines have been used to explore undergraduate students' participation in college classes and career outcomes [5, 49, 55-57, 255-257]. For example, in a study by Stets et.al., students with a higher science identity were more likely to enter a science occupation after college [54]. In addition, students on the pre-medical or pre-health track are more likely to have higher STEM identity and perceived recognition by others than their peers [165]. However, students on these tracks and bioscience majors, in general, are required to take physics courses for their major, and prior research suggests that it can be more challenging for women to form a physics identity than men [1, 3, 6]. While women are not underrepresented in these physics courses for bioscience majors, there may still be a gender gap in the motivational beliefs of students in the course. In a study on introductory bioscience courses, where women made up 60% of the enrollment of the courses, women participated less and underperformed on exams compared to the men in the courses [246]. Moreover, many bioscience majors take physics courses in their junior or senior year and the evolution in students' motivational beliefs in the physics courses could also influence their overall science identity aligned with their major as well as future science pathways. Therefore, it is important to understand the connection between bioscience majors' motivational beliefs in mandatory science courses, e.g., physics, and their overall science identity in order to contemplate multiple pathways to boost their science identity.

Science identity in this context refers to identifying with science or a student's view about whether they see themselves as a "science person"[2, 5]. The science identity framework by Carlone and Johnson [55] includes three dimensions: competence ("I think I can"), performance ("I am able to do"), and recognition ("I am recognized by others"). Hazari et al. modified the framework specifically for physics. "Competence" and "performance" were defined as students' beliefs in their ability to understand the subject and students' belief in their ability to perform physics tasks. Additionally, recognition was framed as perceived recognition by others as students who can excel in physics. Lastly, a fourth dimension, interest, was added to the identity framework [60, 61]. Also, competence and performance were later combined into a single construct, self-efficacy or competency belief. In this framework, students' physics identity was found to be influenced by their self-efficacy, interest, and perceived recognition [3, 4, 6, 62, 234]

Self-efficacy (closely related to competency belief) is defined as a person's motivational belief that they can succeed in a particular activity or course [27, 28]. It has been shown to impact students' engagement, learning, and persistence in science courses [30, 31, 33, 34, 40, 42]. In particular, students with high self-efficacy are more likely to enroll in more challenging classes because they view difficult problems as challenges they can overcome as opposed to threats to be avoided [27]. However, in introductory physics courses in which women are underrepresented, studies have found a gender gap in self-efficacy favoring men that widens by the end of the course [29, 43].

Another motivational belief hypothesized to influence identity in a particular discipline, interest, may also affect students' perseverance, persistence, and achievement [16, 38, 41, 44-46].

One study showed that changing the curriculum to stimulate the interest of the female students helped improve all students' understanding at the end of the year [47]. Within expectancy-value theory, interest and competency beliefs are connected constructs that predict students' academic outcomes and career expectations [48].

The third factor influencing identity, perceived recognition by others, has been shown to be an important factor in motivation [50]. In one study, students' perception of teacher recognition and support was more strongly linked to the motivation and engagement of girls than boys [50]. Studies have shown that female students are not recognized appropriately even before they enter college [1, 4, 24]. For example, one of the stereotypical views is that science is for high achievers or naturally gifted students [1], and in general, being a genius or exceptionally smart is attributed to boys [169]. Boys and girls are exposed to these fixed intelligence views starting from an early age [24]. These views impact recognition by others. One study found that science faculty members in biological and physical sciences exhibit biases against female students by rating male students significantly more competent even when the accomplishments of the hypothetical female and male students were identical [51]. Thus the science identity of women in college science courses can be negatively impacted by their lower perceived recognition by others [140].

All three of these factors (self-efficacy, interest, and perceived recognition) play an important role in identity formation. However, identity can be context dependent [56] and thus a student's physics identity is not the same as their overall science identity aligned more with their disciplinary major. For example, bioscience majors may answer survey questions about their science identity based on their identity aligned primarily with their disciplinary major while physics identity is likely to be aligned with their views of physics and associated courses. However, since physics is also a science domain, it is possible that a positive or negative change in students' physics identity in a course will also impact students' science identity. For example, in one study, first-year college students' math and physics identities were important predictors of their engineering critical agency and identity [6]. Since physics self-efficacy, perceived recognition by others, and interest that contribute to physics identity could also impact students' overall science identity aligned with students' disciplinary major, these connections are important to investigate as they may provide additional pathways for boosting students' overall science identity.

Here we use the framework in which self-efficacy, interest, and perceived recognition play a central role in identity formation. Figure 19 shows a schematic representation of the path analysis part of the model in this study. In particular, we administered a motivational beliefs survey to students at the end of an algebra-based physics course sequence for bioscience majors and used mediation analysis in structural equation modeling (SEM) to investigate how physics self-efficacy, interest, and perceived recognition predict physics identity as well as the overall science identity aligned more with students' disciplinary major. In this study, we answer the following research questions.

RQ1. Are there gender differences in students' motivational beliefs (physics self-efficacy, interest, perceived recognition, and identity, and overall science identity)?

RQ2. To what extent can the components of student physics identity (i.e., self-efficacy, perceived recognition, and interest) predict students' overall science identity?

RQ3. What is the covariance between the physics identity and overall science identity?

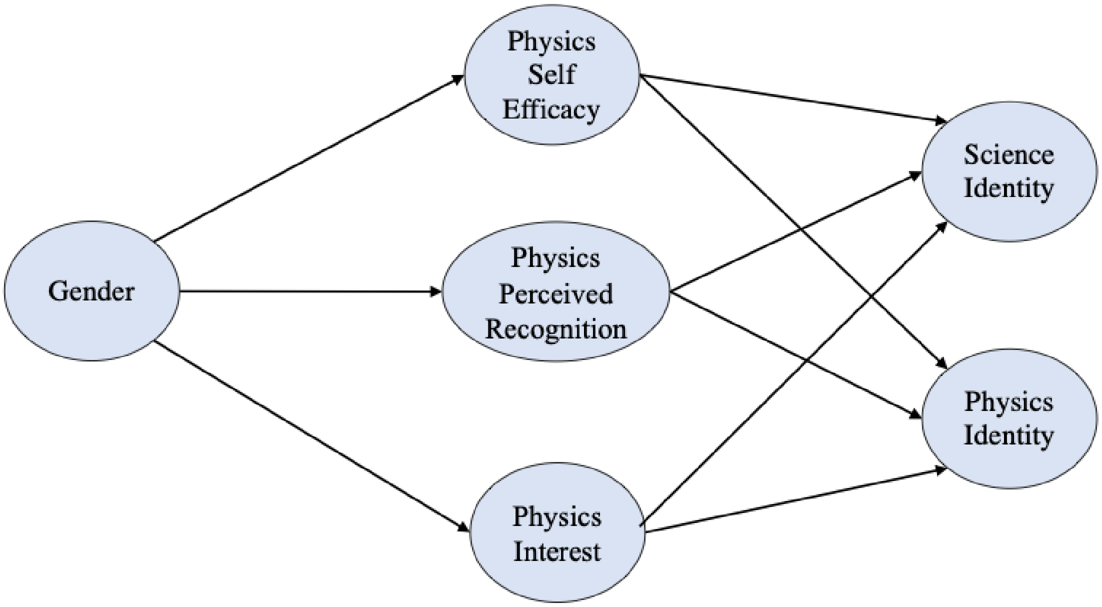


Figure 19 Schematic representation of the path analysis part of the model based on the theoretical framework. From left to right, all possible paths were considered. The regression paths from gender to physics identity or science identity are not shown since we did not find them to be statistically significant in our analysis.

9.2 Methods

9.2.1 Participants

The study was conducted at a large, public research university. We analyzed the results from responses to motivational beliefs survey administered to 873 students who took the survey at the end of the semester in an introductory algebra-based physics 2 course over two years. This research was carried out at the University of Pittsburgh in accordance with the principles outlined in the institutional review board (IRB) ethical policy. All interviewed participants provided oral consent for this research. This class is taken by bioscience majors for whom both physics courses in this sequence are mandatory and is usually taken in their junior or senior year of undergraduate studies. Both physics courses in this sequence were traditionally taught with lectures as the primary mode of instruction with a few clicker questions per week. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. The gender data provided by the university includes only binary options of "male" and "female" for students. We recognize that gender is a socio-cultural and nonbinary construct; however, we are limited to the binary data provided by the university in this study. Based on the university data, 38% of the participants identified as male and 62% as female students.

9.2.2 Instrument Validity

The motivational belief survey instrument used in this study is a validated instrument. This study focused on students' physics self-efficacy, interest, perceived recognition, and identity as well as their overall science identity for the students enrolled in the physics 2 course using the validated survey. The *physics self-efficacy* questions measured students' confidence in their ability to solve and understand physics problems [170-172, 230]. The *interest in physics* questions measured students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *perceived recognition* questions measured the extent to which the students thought that other people see them as a physics person [230]. The *science identity* question evaluated whether the students see themselves as a science person. The *physics identity* questions evaluated whether the students see themselves as a physics person [230]. The physics and science identity instruments only included one question, which is consistent with past studies since it has been difficult to make other questions that factor in this category in exploratory factor analysis [5, 6, 249, 250]. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) [173].

The survey was adapted from previous research [170, 175] and questions were re-validated in our own context using Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), one-on-one student interviews [26], and Pearson Correlations. In order to test the reliability of the survey, we calculated Cronbach alpha [176]. The survey questions for each construct and factor loadings (lambda) for each question are given in Table 22. During one-on-one interviews, bioscience majors noted that their overall science identity was mostly aligned with their disciplinary major.

Table 22 Survey questions for each of the motivational constructs along with factor loadings (lambda) from the Confirmatory Factor Analysis (CFA) result for all students (N = 873). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree. All p-values (showing statistical significance for each factor loading) are < 0.001.

Construct and Item	Lambda
Science Identity	
I see myself as a scientist.	1.00
Physics Identity	
I see myself as a physics person.	1.00
Physics Self-Efficacy	
I am able to help my classmates with physics in the laboratory or recitation.	0.61
I understand concepts I have studied in physics.	0.70
If I study, I will do well on a physics test.	0.74
If I encounter a setback in a physics exam, I can overcome it.	0.69
Physics Interest	
I wonder about how physics works. ††	0.69
In general, I find physics. †	0.79
I want to know everything I can about physics.	0.78
I am curious about recent discoveries in physics.	0.72
Physics Perceived Recognition	
My family sees me as a physics person.	0.91
My friends see me as a physics person.	0.91
My physics instructor and/or TA sees me as a physics person.	0.68

† The rating scale for this question was very boring, boring, interesting, very interesting.

†† The rating scale for this question was never, once a month, once a week, every day.

The pair-wise Pearson correlations are given in Table 23. Pearson's *r* values signify the strength of the relationship between variables. The inter-correlations vary in the strength of their correlation, but none of the correlations are so high that the constructs cannot be separately examined. The highest intercorrelation was the value between Physics Identity and Perceived Recognition (0.82). Perceived recognition questions ask about external identity (perception of whether other people recognize an individual as a physics person), whereas the physics identity question asks about internal identity (whether an individual sees oneself as a physics person) so past research has found a high correlation between these constructs [5, 6, 59, 62]. However, the correlation is low enough that they can be considered separate constructs [258]. Furthermore, the correlations between the physics-specific constructs (perceived recognition, self-efficacy, interest, and physics identity) are stronger than the correlations between them and science identity. In addition, Cronbach alpha was used to measure the internal consistency of the items. The Cronbach alpha is 0.78 for the self-efficacy questions, 0.83 for interest questions, and 0.87 for perceived recognition questions which are considered reasonable [176].

Table 23 Pearson inter-correlations are given between all constructs based upon student responses to the motivational beliefs survey at the end of physics 2.

Pearson Correlation Coefficient				
Observed Variable	1	2	3	4
1. Perceived Recognition	--	--	--	--
2. Self-Efficacy	0.57	--	--	--
3. Interest	0.57	0.58	--	--
4. Physics Identity	0.82	0.58	0.56	--
4. Science Identity	0.23	0.26	0.24	0.21

9.2.3 Analysis

We first compared female and male students' mean scores for all predictors (self-efficacy, perceived recognition, and interest) and outcomes (physics identity and overall science identity) for statistical significance using t-tests and for the effect size using Cohen's d [182]. Cohen's d is $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students, and σ_{pooled} is the pooled standard deviation for all students. To quantify the statistical significance and relative strength of our framework's links (see Figure 14), we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) [259]. The SEM is an extension of the multiple regression, which affords conducting several multiple regressions simultaneously between variables in one estimation model (with the possibility of multiple outcomes, e.g., both physics and science identities). This is an improvement over multiple regression since it allows us to calculate the overall goodness of fit and allows for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMS). Commonly used thresholds for the goodness of fit are as follows: CFI and TLI > 0.90, and SRMR and RMSEA < 0.08 [139].

The model estimates were performed using moderation analysis to check whether any of the relations between variables show differences across gender by using "lavaan" to conduct multi-group SEM [187] Initially we tested different levels of measurement invariance in the multi-group SEM model with gender moderation. 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 [187]. Since we did not find significant moderation by gender, we tested the theoretical model in mediation analysis, using gender as a variable directly predicting all constructs (physics self-efficacy, perceived recognition, interest and identity, and overall science identity) to examine the resulting structural paths between constructs.

9.2.4 Results and discussion

In relation to **RQ1**, women had statistically significantly lower mean values than men for all constructs in our model, including science identity (Table 24). Both women and men scored lowest in physics identity and women scored in the negative range (i.e., a score of 2 which corresponds to "disagree"). One hypothesis for the gender gap in these constructs (indicated by their lower scores in Table 24) is that women may be affected by previous experiences, stereotypes, and biases about who belongs in physics and science in general and who can excel in it, which can accumulate over their lifetime. Table 24 shows that the gender gaps remain at the end of physics 2. It is also clear from Table 24 that both female and male students' science identity is on average higher than their physics identity. This is reasonable since interviews with some of the students suggest that bioscience majors' overall science identity is closely tied to their disciplinary major and they saw themselves as scientists primarily because of their interest in bioscience.

Table 24 Mean construct values by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender. No superscript means p-values are < 0.001 and superscript "a" means p-value = 0.006.

Predictors and Outcomes	Mean		Cohen's *d*
	Male	Female	
Perceived Recognition	2.24	1.98	0.39
Self-Efficacy	2.94	2.73	0.40
Interest	2.77	2.31	0.73
Physics Identity	2.18	1.86	0.44
Science Identity	3.02	2.86	0.19[a]

After analyzing descriptive statistics, we used SEM to investigate the relationships between the constructs and to unpack whether the constructs contributed toward explaining student physics identity and science identity. We initially tested gender moderation between different constructs using multi-group SEM (between male and female students) to see if the relationships between the motivational beliefs were different across gender. There were no group differences at the level of weak and strong measurement invariance including no difference at the level of regression coefficients. Therefore, we proceeded to gender mediation analysis to understand how gender mediates physics identity at the end of the yearlong introductory physics sequence.

The results of the SEM are presented visually in Figure 20. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): CFI = 0.964 (> 0.90), TLI = 0.952 (> 0.90), RMSEA = 0.058 (<0.08), and SRMR = 0.038 (< 0.08). The solid lines represent regression paths and the regression coefficients (β) represent the strength of the regression relation. All three of the intervening variables (perceived recognition, self-efficacy, and interest) predict physics identity at the end of the yearlong physics course sequence, similar to past models [5, 59]. Perceived recognition has the largest direct effect on physics identity with smaller effects from self-efficacy and interest. In response to **RQ2,** we find that students' science identity is predicted by physics self-efficacy and perceived recognition with student self-efficacy in physics having the largest direct effect on science identity (0.20). For bioscience majors, this physics 2 course is the second of the two physics courses they take, usually in their junior or senior year. In addition, gender is directly connected to perceived recognition (P.R.), self-efficacy, and interest. The relations between gender and both science identity and physics identity are mediated by the intervening variables and gender does not directly predict either science identity or physics identity. Thus, women have a lower physics and science identity because they have lower self-efficacy, perceived recognition, and interest. In response to **RQ3,** we investigated the covariance (dashed lines) between science and physics identity and find that science identity and physics identity covary with one another (0.21). Thus, while students' overall science identity and physics identity are not the same construct, they could influence one another in meaningful ways.

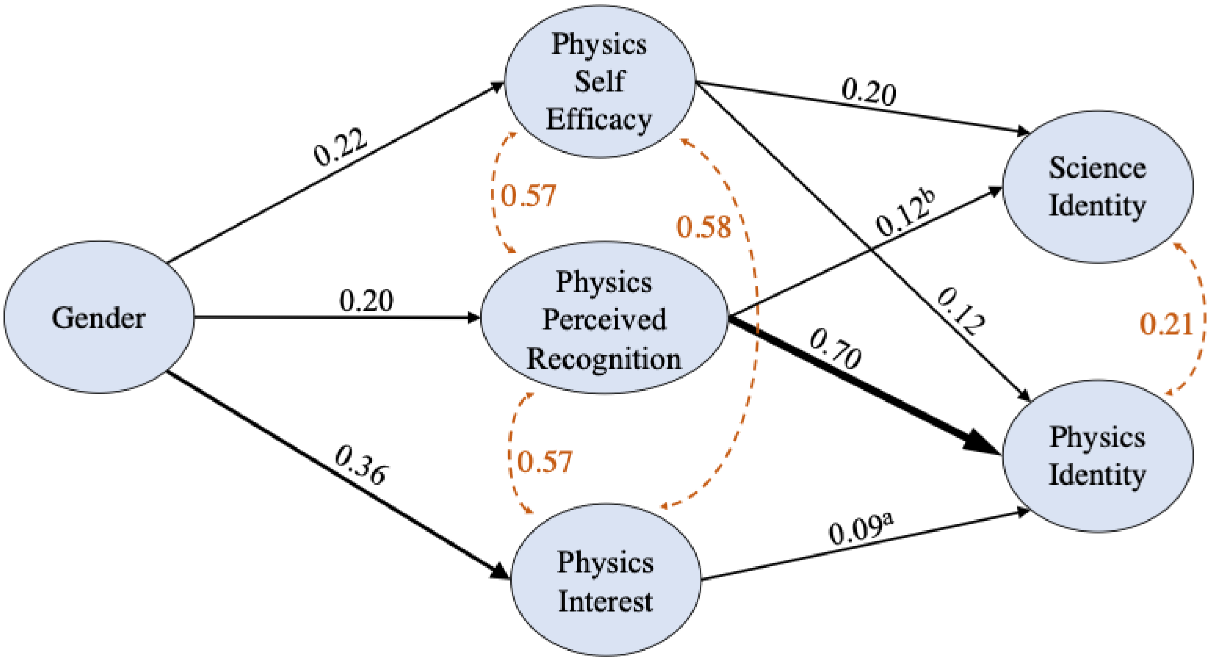


Figure 20 Result of the path analysis part of the SEM with mediation between gender and science/physics identity through physics perceived recognition (P.R.), self-efficacy, and interest. The line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. The dashed lines indicate covariance. All *p*-values for β are indicated by no superscript for $p < 0.001$, "a" for $p = 0.004$, "b" and for $p = 0.010$.

9.2.5 Summary and implications

In physics courses that bioscience majors are required to take for their major, women have lower mean values in their motivational beliefs than men (Table 24). It is important to note that there was a high correlation between students' physics perceived recognition and physics identity (Table 23), however, that is consistent with past research [5, 6, 59, 62]. In addition, the science and physics identity constructs included one question which is consistent with past studies since it has been challenging to make other questions that factor in this category in exploratory factor analysis [5, 6, 249, 250].

In our model, we found that some of the motivational beliefs that predict physics identity also predict science identity at the end of the course. The identity model in Figure 20 shows the relationship between the predictors (self-efficacy, perceived recognition, and interest) and students' physics and science identities. In particular, we find that at the end of a two-semester mandatory physics course sequence, bioscience majors' physics self-efficacy and perceived recognition not only predicted their physics identity but also their overall science identity primarily aligned with their disciplinary major. The model suggests that physics TAs and instructors can potentially play an important role in increasing not only students' physics identity but also their overall science identity, since perceived recognition by them in the physics course predicts science identity. It is important to note that physics self-efficacy also predicts science identity even though the students' overall science identity is more closely related to their bioscience identity. These relations between physics self-efficacy and perceived recognition and the overall science identity of bioscience majors suggest interdisciplinary connections that may provide additional pathways for boosting students' science identity aligned with their major, for example, by enhancing their self-efficacy and perceived recognition in their other mandatory science courses such as physics. Some ways instructors can build these interdisciplinary connections is by providing examples and problems in class that connect physics concepts with bioscience concepts. For example, in the introductory physics lab course, there are several labs about bioscience concepts including those focusing on the physics of human eyes, electrocardiogram (EKG or ECG), blood pressure, and viscosity measurements that help students build connections between physics and bioscience.

Since science identity has the potential to influence students' persistence in science careers, it is important for physics instructors to employ approaches that improve students' science identity.

We also find covariance (0.21) between the overall science identity and physics identity which may be useful for devising strategies for boosting students' science identity and career aspirations. Since physics 2 is usually taken in the students' junior or senior year, it is one of the last non-major STEM courses they take before they graduate. Therefore, physics instructors and TAs may have the opportunity to improve students' science identity by recognizing and affirming their students positively for making progress. Furthermore, at the end of a traditionally taught two-semester mandatory physics course sequence, we find that on average, women majoring in bioscience had lower physics self-efficacy, perceived recognition, physics identity, and overall science identity aligned with their disciplinary major than men even though women were not underrepresented in the physics course. One possible reason is that the societal stereotypes and biases pertaining to who can excel in physics can impact women who are exposed to these stereotypes and biases from an early age through differential upbringing, media coverage, K-12 education, and the culture of college physics classes. Moreover, although both men and women had an average perceived recognition below the positive lower threshold (score of 3 corresponding to "agree"), women's averages were lower than men's. The fact that women feel less recognized by their instructors and TAs than men could influence women's self-efficacy, interest, and identity in physics in particular and science in general.

Physics instructors may unwittingly reinforce gender stereotypes about physics and communicate lower expectations for female students in physics classes. By not letting men dominate the conversations in class and explicitly praising effort and affirming women when they do well/achieve in the class, the gender gap in students' perceived recognition could be decreased.

This is especially important when implementing active learning pedagogy in the classroom. If not implemented using teaching strategies that are equitable and inclusive, the gender gap in the courses may increase [92]. In prior studies, men have been shown to dominate responding to questions in class and women have reported lower scientific self-efficacy [219]. Moreover, instructors should be careful not to say that problems are “trivial”, “easy” or “obvious” when students’ ask them for help after trying their best because otherwise female students are more likely to feel disparaged (or negatively recognized). Another way to improve the learning environment in science courses is through classroom interventions, which have been shown to eliminate the gender gap in performance and also have the potential to positively impact the motivational beliefs of students [151, 152, 185, 186].

10.0 Role of inclusiveness of learning environment in predicting students' outcomes in courses in which women are not underrepresented

10.1 Introduction and Theoretical Framework

In the past few decades, there has been a focus on the experiences and participation of women in many science, technology, engineering, and math (STEM) fields [115, 145, 149, 153, 260, 261]. Some research studies have focused on students' beliefs in different STEM domains which can influence students' continuation in related courses, majors, and careers [43, 118, 156, 238, 240, 242, 245]. Some of these studies that focus on women and ethnic and racial minority students in physics show that they, in general, have lower beliefs and grades than men [1-3]. Inequitable outcomes like this may be a result of inequitable access to resources, inadequate support, and inequitable learning environments.

Our conceptualization of equity in learning includes three pillars: equitable opportunities to learn, equitable and inclusive learning environments, and equitable outcomes. Equity in learning would require all students to have equitable opportunities and access to resources and that students 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. Equitable learning outcome means that students from all demographic groups (e.g., regardless of their gender identity, etc.) 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 [72]. We note that equitable and inclusive learning environments and equitable outcomes are intricately tied to each other.

The formation and development of students' science identity have been a focus of many prior studies [4, 5, 262-265]. Science identity is defined as identifying with academic domains in science, i.e., whether students see themselves as science people [2, 5, 266] or those who can excel in science. Students' identity in STEM disciplines has been shown to play an important role in their in-class participation and choices of majors and careers [49, 54-56, 267]. Studies have shown that it can be more difficult for women to form a physics identity than men [1-3, 6]. However, most of the studies in the college context concerning physics identity and factors that influence it are conducted in classes in which women are underrepresented [4, 5]. Students' identity in a domain can be context dependent [56], therefore, it is important to examine the role of the inclusiveness of the learning environment on the physics identity of students in courses in which women outnumber men, e.g., in introductory physics courses for bioscience majors.

When examining equity in physics education and the role of inclusiveness of the learning environment, it is useful to examine the practices that lead to inequities [268-270]. Physics is a discipline with problematic stereotypes and biases about who belongs in it and can excel in it. One common stereotype is that only high achievers or geniuses can excel in physics [23]. However, genius is more often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius [24]. These stereotypes can continue to impact women as they get older. Teachers and school counselors pay more attention to male students and counselors give gendered advice to students regarding high school physics and math courses to take and majors to pursue when in college. These stereotypes and biases are also prevalent at the university level. One study found that biology and physics faculty members rated a male student as significantly more competent than a female student when presented with a hypothetical scenario for hiring a student for lab work with either a male or female name [51]. These highly problematic stereotypes and biases are founded in the historical marginalization of certain groups, e.g., women in physics, and continue to manifest today in many ways including, gendered beliefs and barriers to women excelling in physics when there is no explicit focus on making the learning environment equitable and inclusive. The pervasive stereotypes and biases can impact women's identity even in these physics courses in which they are not underrepresented in an inequitable and non-inclusive environment.

Since science identity is key to students' success, it is important to investigate students' beliefs that play a role in identity formation. Carlone and Johnson's science identity framework includes three interrelated dimensions: competence, performance, and recognition by others [131]. However, in a study of introductory physics students by Hazari et al., performance and competence were found to be highly correlated and a fourth dimension, interest was added to the model [5]. In a slightly reframed version of the physics identity framework [59], students' self-efficacy, interest, and perceived recognition by others have been shown to predict students' physics identity.

Self-efficacy in a particular discipline is the students' belief in their ability to succeed in a particular task or course [27]. It has been shown to impact students' engagement, learning, and persistence in science courses [14, 29-31, 34-36, 271]. Students with high self-efficacy in a domain are less likely to have anxiety that can rob them of their cognitive resources while learning and test taking, since the working memory during problem solving has limited capacity [166]. They are also more likely to engage in effective learning strategies and are less likely to procrastinate while engaging with learning tools. Thus, self-efficacy can be a predictor of student performance in that domain [33].

Similarly, interest in a particular discipline can affect students' perseverance and achievement in that discipline [16, 41, 46, 272]. One study showed that changing the curriculum to stimulate the interest of girls helped improve all the students' understanding at the end of the year [47]. Furthermore, according to expectancy-value theory, self-efficacy and interest are related constructs that predict student outcomes and career expectations [48]. Moreover, students' interest and self-efficacy can be connected to their interaction with and recognition by other people [46, 273].

A student's perceived recognition by others has also been shown to play an important role in a student's identity [49] and is a particularly important factor that influences women's other beliefs [50]. Studies have shown that female students do not feel recognized appropriately even before they enter college [1, 4, 24]. Our prior individual interviews suggest that students' perceived recognition by instructors and teaching assistants impacts their self-efficacy and interest in physics, e.g., see [52, 53]. In addition, other learning environment factors, such as students' sense of belonging and perception of the effectiveness of interaction with their peers are important. Students' sense of belonging in physics has been shown to correlate with their retention and self-efficacy [50, 63] and students' interaction with peers has been shown to enhance understanding and engagement in courses [274]. Since these factors can impact students' physics beliefs, it is important to investigate student perceptions of the inclusiveness of the learning environment and how they influence the outcomes pertaining to physics self-efficacy, interest, and identity.

This study examined students' perception of the inclusiveness of learning environment (including their sense of belonging, perceived recognition, and peer interaction) on their self-efficacy, interest, and identity at the end of an algebra-based introductory physics course controlling for students' gender, self-efficacy, and interest at the beginning of the course (Figure 21). Student perceptions of the inclusiveness of the learning environment are not only formed by classroom experiences but also experiences they have outside of the classroom. For example, the perceptions of the inclusiveness of the learning environment could also be shaped by students asking for help during office hours, their email correspondence with the instructor or teaching assistant (TA), and students studying together. Therefore, the perception of the inclusiveness of learning environment factors in our investigation include students' perceptions of their interaction with their peers (from now on simply referred to as peer interaction for brevity), sense of belonging (referred to as belonging for brevity), and perceived recognition by others (including friends, family, and their instructors/TAs). We control for students' gender, self-efficacy, and interest at the beginning of the course, which are related to students' beliefs about physics before they enter the classroom. A schematic diagram of our model with only the path analysis is shown in Figure 21.

We first delineate the research questions based upon our framework, then describe the methodology, show the results and discussion, and finally conclude with instructional implications and future directions. We use structural equation modeling (SEM) [259] to investigate how the perception of the inclusiveness of the learning environment predicts students' physics self-efficacy, interest, and identity at the end of the course. The following research questions were answered by analyzing data from a validated survey administered to students in algebra-based physics courses primarily for bioscience majors at a large public research university in the US in which women outnumber men and using mediation analysis via SEM:

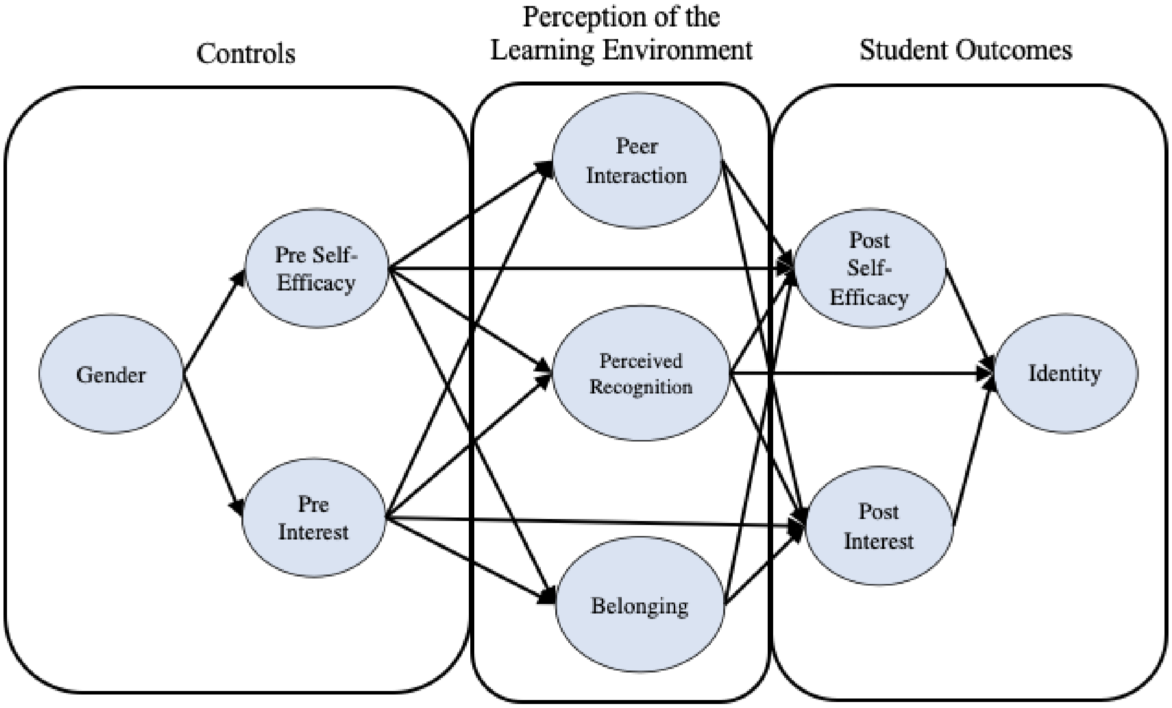


Figure 21 All regression paths were considered from left to right in our model; however, only some of the paths are shown for clarity.

RQ1. Are there gender differences in students' physics beliefs (self-efficacy, interest, and identity) at the end of the course and do interest and self-efficacy change from the beginning to the end of the course?

RQ2. Is there moderation by gender for any of the regression paths?

RQ3. If moderation does not affect any path, does gender mediate any of the controlled factors, perceived inclusiveness of learning environment factors, or student outcomes?

RQ4. How does the inclusiveness of the learning environment (including perceived recognition, belonging, and peer interaction) predict physics identity, physics self-efficacy, and physics interest at the end of the course?

10.2 Methodology

10.2.1 Participants

We administered a validated written survey at a large public research university in the U.S. to students at the beginning (pre) and end (post) of the first semester of an algebra-based introductory physics course (called physics 1). The class is typically taken as a mandatory requirement by students on the pre-medical or pre-health track (mainly biosciences majors) primarily in their junior or senior year of undergraduate studies. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. We analyzed data from 501 students who completed the pre/post survey on paper scantrons. From the university data, the participants were 35% male and 65% female students. We acknowledge that gender is not a binary construct. However, the data provided by the university only includes options of male and female and thus gender was used as a binary variable in our analysis. Less than 1% of the students did not provide this information and were not included in the analysis.

10.2.2 Instrument Validity

The items in the validated survey focused on different aspects of students' beliefs at the time the survey was administered (beginning and end of the course) that included their perception of the inclusiveness of the learning environment. In particular, the study focused on students' responses to validated survey items on physics identity, self-efficacy, interest, sense of belonging, perceived recognition, and peer interaction. The survey items were adapted from previously validated surveys [170, 175, 275] and re-validated in our own context using one-on-one student interviews, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) [26, 53], and Pearson correlations. Interviews with students suggest that they interpreted the items correctly. The CFA established a measurement model for the constructs and was used in SEM. In the CFA, the model fit indices were good and all the factor loadings were above 0.50, which indicates good loadings [182]. The results of the CFA model are shown in Table 25.

The *physics identity* items focus on whether the students see themselves as a physics person [230]. The *physics self-efficacy* items measure students' confidence in their ability to solve physics problems [170-172, 230]. The *interest in physics* items measure students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *sense of belonging* items evaluate whether students felt like they belonged in the introductory physics class [50, 175]. The *perceived recognition* items measure the extent the student thought that other people see them as a physics person [230]. Lastly, the *peer interaction* items measure whether students thought that working with their peers was beneficial, e.g., in increasing their confidence to do physics [276, 277]. The questions in the study were denoted on a Likert scale of 1 (low belief) to 4 (high belief) except for the sense of belonging questions which were designed on a scale of 1 to 5 to keep them consistent with the original survey [173].

Table 25 Survey questions for each of the constructs and factor loadings from the Confirmatory Factor Analysis (CFA) result for all students (N = 501). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity, peer interaction, and perceived recognition questions were: strongly disagree, disagree, agree, strongly agree. The rating scale for the physics belonging questions was: not at all true, a little true, somewhat true, mostly true, and completely true. All p-values are <0.001.

Construct and Item	Lambda
Physics Identity	
I see myself as a physics person.	1.00
Physics Self-Efficacy	
I am able to help my classmates with physics in the laboratory or recitation.	0.60
I understand concepts I have studied in physics.	0.74
If I study, I will do well on a physics test.	0.76
If I encounter a setback in a physics exam, I can overcome it.	0.72
Physics Interest	
I wonder about how physics works.	0.57
In general, I find physics.[†]	0.75
I want to know everything I can about physics.	0.74
I am curious about recent discoveries in physics.	0.64
Physics Perceived Recognition	
My family sees me as a physics person.	0.89
My friends see me as a physics person.	0.92
My physics instructor and/or TA sees me as a physics person.	0.71
Physics Belonging	
I feel like I belong in this class.	0.81
I feel like an outsider in this class.	0.74
I feel comfortable in this class.	0.82
I feel like I can be myself in this class.	0.62
Sometimes I worry that I do not belong in this class.	0.69

Physics Peer Interaction My experiences and interactions with other students in this class…	
made me feel more relaxed about learning physics.	0.75
increased my confidence in my ability to do physics.	0.93
increased my confidence that I can succeed in physics.	0.96
increased my confidence in my ability to handle difficult physics problems.	0.90

† the rating scale for this question was very boring, boring, interesting, very interesting.

Pair-wise Pearson's *r* values indicate the correlation between each pair of constructs without controlling for the influence of any other factors (values are given in Table 26). Inter-correlations vary in strength, but none of the correlations are so high that the constructs cannot be separately examined, consistent with prior studies. In other words, all the correlations are low enough that all the constructs can be considered separate. Since the construct is the same at two different points of time and the course does not affect students' interest significantly, the highest inter-correlation was the value between pre-interest and post-interest (0.86). Our past work in calculus-based introductory physics courses has also shown that there is a high correlation in interest over the course of the introductory classes [26].

Table 26 Pearson inter-correlations are given for each pair of constructs. All p-values are < 0.001

	Pearson Correlation Coefficient							
Observed Variable	1	2	3	4	5	6	7	8
1. Pre Self-Efficacy	--	--	--	--	--	--	--	--
2. Pre Interest	0.50	--	--	--	--	--	--	--
3. Perceived Recognition	0.41	0.41	--	--	--	--	--	--
4. Peer Interaction	0.21	0.24	0.46	--	--	--	--	--
5. Belonging	0.48	0.32	0.59	0.62	--	--	--	--
6. Post Self-Efficacy	0.54	0.35	0.61	0.62	0.81	--	--	--
7. Post Interest	0.34	0.86	0.61	0.47	0.51	0.60	--	--
8. Physics Identity	0.40	0.45	0.78	0.43	0.58	0.60	0.62	--

10.2.3 Analysis

We first analyzed descriptive statistics and compared female and male students' mean scores of the predictors and outcomes for statistical significance using *t*-tests and computed the effect size using Cohen's *d* [182]. The square of CFA factor loading (lambda) indicates the fraction of variance explained by the factor. For predictive relationships between different constructs, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [259]. SEM is an extension of multiple regression and conducts several multiple regressions simultaneously between variables in one estimation model. This allows us to calculate the overall goodness of fit and allows for all estimates to be performed simultaneously so there can be a direct comparison between different structural components. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [139].

The model estimates were performed using moderation analysis to check whether any of the relations between variables show differences across gender by using "lavaan" to conduct multi-group SEM. 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. Since we did not find significant moderation by gender, we tested the theoretical model in mediation analysis, using gender as a variable directly predicting items to examine the resulting structural paths between constructs.

10.3 Results and Discussion

10.3.1 Gender Differences in predictors and outcomes

To answer **RQ1**, the mean values of all constructs in our model were statistically significantly different disadvantaging female students (Table 27). This pattern is qualitatively similar to what was previously observed in calculus-based courses [4, 59] even though women make up the majority of students in the algebra-based course (65%). Furthermore, for both self-efficacy (gray) and interest (blue), constructs that have scores from both the pre test and the post test, we find statistically significant differences in the scores from the beginning to the end of physics 1 for both women and men. One possible hypothesis for this disparity is the previous experiences and stereotypes about physics that women have internalized from an early age before starting college physics (both in K-12 education setting and outside of it interacting with family, friends, media, etc.). Additionally, Table 27 shows that women also have more negative experiences in the classroom, indicated by their lower scores on the perception of the inclusiveness of the learning environment constructs, despite women making up the majority of students in the class.

Table 27 Mean predictor and outcome values and effect sizes (Cohen's d) by gender. All *p*-values are < 0.001.

Predictors and Outcomes (Score Range)	Mean		Cohen's *d*
	Male	Female	
Pre Self-Efficacy (1-4)	3.07	2.83	0.58
Pre Interest (1-4)	2.74	2.45	0.58
Perceived Recognition (1-4)	2.22	1.92	0.43
Peer Interaction (1-4)	2.90	2.58	0.43
Belonging (1-5)	3.58	3.01	0.60
Post Self-Efficacy (1-4)	2.92	2.55	0.66
Post Interest (1-4)	2.66	2.27	0.69
Physics Identity (1-4)	2.04	1.64	0.59

10.3.2 SEM path model

In relation to **RQ2,** we initially tested moderation analysis using multi-group SEM between female and male students to find if any of the relationships between the variables were different across gender. There were no group differences at the level of weak and strong measurement invariance at the level of regression coefficients, so we proceeded with mediation analysis.

Then, in relation to **RQ3,** Figure 22 shows that gender only directly predicts pre self-efficacy, pre interest, sense of belonging, and peer interaction. In particular, Figure 22 shows that gender does not directly predict any of the outcomes. We used mediation analysis to understand the extent to which gender differences in students' outcomes at the end of the introductory physics courses (self-efficacy or S.E., interest, and physics identity) were mediated by differences in students' initial self-efficacy, interest, and perception of the learning environment (perceived recognition, peer interaction, and belonging) in the course. The model is shown in Figure 22.

In relation to **RQ4,** we found that interest at the end of physics 1 was mainly predicted by interest at the beginning of the class (regression coefficient $\beta = 0.72$) with smaller effects from peer interaction ($\beta = 0.24$) and perceived recognition ($\beta = 0.21$). Although post interest appears to be strongly correlated with pre interest, it does not mean that interest cannot be changed throughout the class via evidence-based intentional design. For example, students' interest could increase through the learning environment from the positive contributions from peer interaction and perceived recognition. One possible way to increase students' interest in physics is to provide opportunities in class that relate to students' interests and career paths, so students can see how they could use physics in their careers. Relating physics to everyday life could be another effective approach to increasing interest [278]

Post self-efficacy has direct effects from pre self-efficacy ($\beta = 0.45$), belonging ($\beta = 0.49$), peer interaction ($\beta = 0.21$), and a smaller direct effect from perceived recognition ($\beta = 0.13$). Self-efficacy is important for students' persistence in the class and future careers. Since the learning environment can predict student self-efficacy, it is important for an instructor to try to improve the belonging, peer interaction, and perceived recognition of the students. For instance, instructors can positively influence the students' peer interaction with each other by providing time for the students to work together in class and making sure all voices are heard equally when discussing problems as a class.

Self-efficacy, interest, and perceived recognition influence students' physics identity directly, with perceived recognition having the largest direct effect (β = 0.59). This is consistent with past models [4, 5]. Table 27 shows us that both women and men have a mean perceived recognition below the positive threshold (score of 3) and women also have lower scores than men in all constructs, including their physics identity.

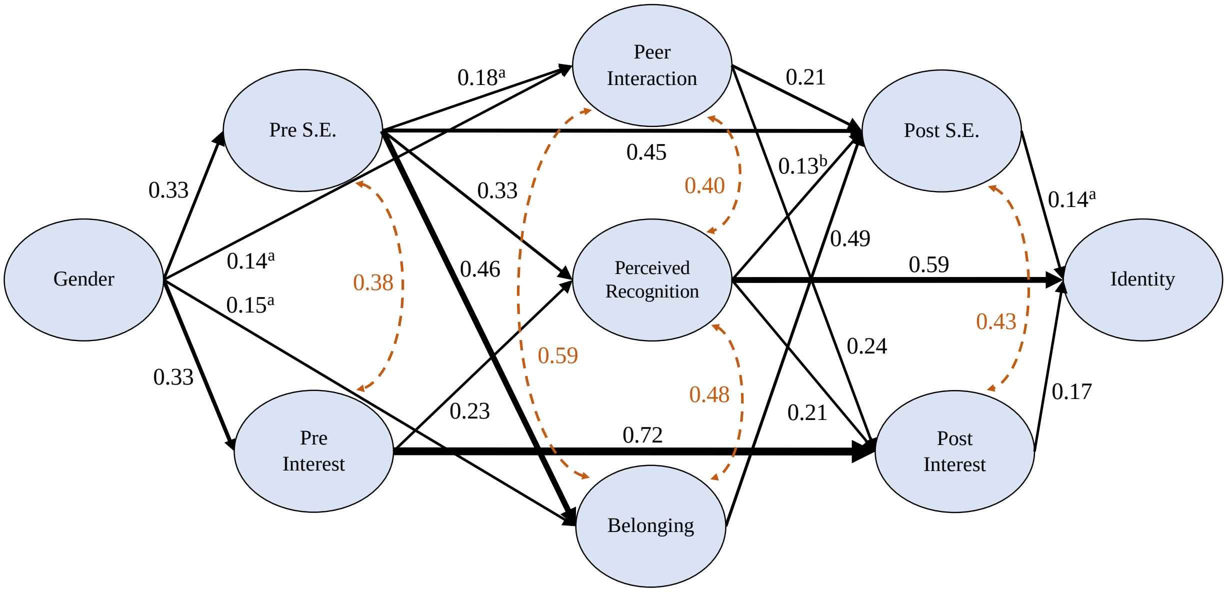


Figure 22 Result of the SEM between gender and outcomes in physics through various mediating factors. The line thickness is the relative magnitude of β values. All p-values are indicated by no superscript for p <0.001, "a" for $p < 0.005$, and "b" for $p < 0.05$ values. S.E. refers to students' self-efficacy. All regression paths that are statistically significant from left to right between any two constructs are shown. The dashed lines indicate covariance.

10.4 Summary, Implications, and Future Directions

In this research, we find a gender gap disadvantaging women in all of the students' beliefs studied, similar to what was found in calculus-based physics courses [59]. In addition, both men's and women's self-efficacy and interest scores dropped from the beginning to the end of the physics 1 course. Women also had a more negative perception of the inclusiveness of learning environment constructs (Table 27). These differences may result from structural inequities in the physics learning environment and marginalized students, e.g., women, not being adequately supported [68, 70, 279-281]

Our SEM model indicates that the perception of the inclusiveness of learning environment factors (belonging, peer interaction, and perceived recognition) explains student outcomes (self-efficacy, interest, and identity). While past research has investigated physics identity, no research has investigated factors that instructors can control in order to make their classes equitable and inclusive. Our findings suggest that it may be beneficial to implement structural changes at the classroom level that targets students' sense of belonging, peer interaction, and perceived recognition in order to improve equity and inclusion. These factors predict each other as well, so if instructors can provide support for some of the factors they can readily control (e.g., peer interaction or perceived recognition), they can make their classes more equitable, and they are likely to improve student outcomes in the process.

We note that the learning environment in many classes may not be providing the same opportunities for all students [282, 283], particularly those from marginalized groups such as women in physics. Approaches to improve the inclusivity of the learning environment need to be systemic. Structural changes at the institution level require centering marginalized students in the design of curricula and pedagogies. Departments and institutions can reward instructors for supporting equity and inclusion that eliminates the problematic gender gap discussed here.

In addition, instructors can improve the learning environment in their courses by adopting culturally responsive pedagogy and providing mentoring/support for students who are underrepresented [64]. Instructors should be careful not to say that problems are “trivial”, “easy” or “obvious” when students ask them for help because female students are more likely to feel disparaged (or negatively recognized) due to the stereotypes in physics. What is important for instructors to realize is that it is not their intention that matters but the impact they are having on students. Another way to improve the learning environment is through classroom interventions. Brief social-psychological classroom interventions have been shown to decrease or eliminate the gap in STEM contexts between students from marginalized and dominant groups [151, 152, 185, 186].

In summary, more should be done in the college physics classrooms to mitigate the stereotypes and past experiences women have had over their lifetime since otherwise even in a physics class in which they are not underrepresented, women will be disadvantaged. Physics instructors and TAs need to provide an equitable and inclusive learning environment that emphasizes recognizing students for making progress, allowing for positive peer interactions, and providing a space where all students can feel that they belong. From our analysis, these factors play a key role in predicting students’ self-efficacy, interest, and identity in physics. It is important to note that student perception of the inclusiveness of the learning environment is not shaped only by what happens in the classroom. Student interactions with each other while they do homework, students’ experiences in an instructor's or TA’s office hours, interactions between students and the instructor over email, and other circumstances all contribute to students’ perception of inclusiveness of learning environment and can affect students' physics identity, self-efficacy, and interest.

11.0 How the learning environment predicts female and male students' physics self-efficacy, interest, and identity in an introductory course for bioscience majors

11.1 Introduction and Theoretical Framework

Studies have shown that motivational factors such as students' identity, self-efficacy, and interest in a particular field are important for students' career interests [32, 54, 230, 284], learning [40, 285], and continuation in science, technology, engineering, and math (STEM) fields [39, 159, 286, 287]. For example, students were more likely to take courses or pursue a career in science if they had higher competency beliefs or self-efficacy [32, 39, 41, 155, 156], displayed higher interest in science [157], or had a higher science identity [54, 158, 159]. Moreover, a gender gap favoring men in motivational factors [26, 29, 97, 160, 229] and conceptual tests [84] have been observed in STEM courses. Specifically, in physics, many studies have been conducted in calculus-based physics courses, where women tend to be underrepresented, to understand and address the low diversity in those courses.

However, stereotypes about who can excel in physics could impact women even in the physics courses in which they are not underrepresented, e.g., mandatory two semester physics course sequence for bioscience majors. One common stereotype is that genius and brilliance are important factors to succeed in physics [23]. However, genius is often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius [24]. Studies have found that by the age of six, girls are less likely than boys to believe they are "really really smart" and less likely to choose activities that are made for "brilliant people" [24]. Moreover, as these students get older, norms in the science curriculum hold less relevance for girls since they tend not to represent the interest and values of girls [1]. All these stereotypes, biases, and related factors can influence female students' perceptions about their ability to do physics before they enter the physics classroom. Thus, it is possible that although women are the majority in algebra-based physics courses primarily for bioscience majors, these societal stereotypes and biases can still influence their motivational outcomes in the physics class unless instructors make an explicit effort to create an equitable and inclusive learning environment in which all students thrive and have high motivational outcomes.

Additionally, students' identity in STEM disciplines (e.g., whether they see themselves as a STEM person) has been shown to play an important role in their participation in classes and professional choices [5, 49, 54-57]. However, prior studies have shown that it can be more difficult for women to form a physics identity than men [1-3, 6]. Furthermore, students' physics identity has been shown to be influenced by their self-efficacy, interest, and perceived recognition [3, 6, 31, 234, 235].

Self-efficacy is a person's belief that they can succeed in a particular activity or course [27, 28]. Students' self-efficacy in academic courses may be influenced by the classroom environment [37, 161, 162, 288] and different teaching strategies [29, 34, 38]. Moreover, self-efficacy has been shown to impact students' engagement, learning, and persistence in science courses [14, 15, 31-33, 35, 36, 39, 40, 289]. Similarly, interest in a particular discipline may affect students' STEM career orientation [44, 290] and persistence in STEM courses and majors [16, 45, 46]. One study showed that changing the curriculum to stimulate the interest of the female students helped improve all of the students' understanding at the end of the year [47].

Since societal stereotypes and biases about physics can impact female students' motivational beliefs even in physics courses in which they outnumber male students, it is important to investigate how female and male students' perceptions of the inclusiveness of the learning environment in these courses influence their physics self-efficacy, interest and identity. Our theoretical framework is inspired by prior research and posits that the inclusiveness of the learning environment can influence students' motivational outcomes. Specifically, we investigated students' perception of the inclusiveness of the learning environment and how it predicts students' motivational outcomes in a physics course for bioscience majors in which women outnumber men. In our investigation, students' perception of the inclusiveness of the physics learning environment consists of their sense of belonging, their perception of the effectiveness of their interactions with their peers, and perceived recognition by their instructors/teaching assistants (TAs). Students' sense of belonging is a measure of the inclusiveness of the learning environment and has been shown to correlate with retention and self-efficacy in science [50, 63] as well as their identity [58]. Therefore, we include students' sense of belonging as an element of their perception of the inclusiveness of the learning environment and hypothesize that it will explain students' motivational beliefs at the end of a physics course. Students' positive interactions with peers have been shown to enhance their understanding and engagement in courses [253, 274, 291]. Thus, we include students' perception of the effectiveness of the interactions with peers as a measure of the inclusiveness of the learning environment and we hypothesize that the perception of positive peer interactions is likely to have a positive impact on students' motivational outcomes. Moreover, perceived recognition by instructors and TAs has been shown to play an important role in students' identity [5, 49, 59] as well as being an important factor in women's motivation [50]. Thus, we include students' perceived recognition by instructors and TAs as an element of the inclusiveness of the learning environment which can impact students' motivational outcomes.

Furthermore, our framework positing that these three factors together are good measures of student perception of the inclusiveness of the physics learning environment is informed by female students' testimonials in prior qualitative studies. In particular, during individual interviews, female students in physics courses discussed them as being important metrics for inclusiveness of the learning environment. For example, interviewed women cited how negative interactions with their peers alienated them, caused them to feel inferior, and excluded them from study groups [141]. Also, 37% of the women interviewed felt marginalized in physics courses when men were allowed to dominate the class discussion, in asking questions, and answering questions posed by the instructors [292]. Moreover, 29% of women interviewed felt that physics instructors' condescending and intimidating behavior made them feel like the learning environment was unsafe [292]. Research involving interviews also shows that instructors' disparaging comments, such as calling problems trivial or questioning students' effort, can negatively impact women's confidence [141]. Also, interviewed women discussed that their sense of belonging in the physics courses influenced their motivational beliefs.

We also note that some prior studies have investigated students' perception of the different aspects of the inclusiveness of the learning environment individually. For example, some studies investigating students' identity have investigated the role of perceived recognition from others on identity [5, 59]. Another prior study focusing on physics majors investigated how students' sense of belonging [58] predicts their physics identity. In addition, other research has investigated the relation between peer interaction and students' physics self-efficacy [164]. However, there is no research investigating the perception of the inclusiveness of the learning environment factors all together on students' motivational outcomes. Moreover, since these motivational beliefs may influence each other. Therefore it is important to investigate students' perception of the inclusiveness of the learning environment factors together to understand how they predict students' self-efficacy, interest, and identity. In addition, most of the research on students' motivational beliefs in college are either conducted in physics courses in which women are underrepresented [59, 62] or in English composition courses in which most of the students are first-year students with some in the second year [5]. However, students' physics motivational beliefs may be influenced not only in physics courses in which women are underrepresented, but also in courses where the learning environment is not equitable and inclusive due to the stereotypes and biases about who belongs in physics and can excel in it. Thus, it is crucial to investigate female and male students' perception of the inclusiveness of the learning environment and how they explain their motivational beliefs in physics courses in which women are not underrepresented. Gendered perception of the inclusiveness of the learning environment and motivational outcomes in these courses would signify inequities and the need to improve the inclusiveness of the learning environment. Therefore, we investigated how female and male students' perceptions of the inclusiveness of their learning environment predict their motivational beliefs. Since instructors have the ability to create a positive classroom culture and influence the inclusiveness of the learning environment factors, the investigation reported here can empower instructors to improve the experiences of students in their classes with a focus on marginalized students such as women.

In this study, we use structural equation modeling (SEM) [259] to examine the difference between male and female students' perceptions of the inclusiveness of the learning environment on their motivational beliefs at the end of the algebra-based introductory physics sequence for bioscience majors in which women are not underrepresented. The perception of the inclusiveness of the learning environment is shaped by experiences students have in the classroom as well as interactions outside of the classroom. These interactions include office hours, email correspondence with the instructor or TA, and students studying or doing homework together. We control for students' self-efficacy and interest at the end of physics 1 since these are students' beliefs about physics when they enter the class based on their prior experiences. The perception of the inclusiveness of the learning environment includes students' sense of belonging, perception of the effectiveness of peer interaction, and perceived recognition from instructors and TAs. Lastly, we investigated the students' motivational outcomes with regard to their physics self-efficacy, interest, and identity at the end of physics 2. An example of our final model is shown in Figure 23. All paths were considered from left to right in our model, however, only some of the paths are shown in Figure 23 for clarity. The research questions are below:

RQ1. Are there gender differences in students' motivational beliefs including physics self-efficacy, interest, and identity at the end of the course?

RQ2. How does the perception of the inclusiveness of the learning environment (including sense of belonging, perception of peer interaction, and perceived recognition) predict motivational beliefs at the end of the course?

RQ3. What is the effect of controlling for high school factors (e.g., high school GPA and SAT Math scores) on the motivational beliefs at the end of the course?

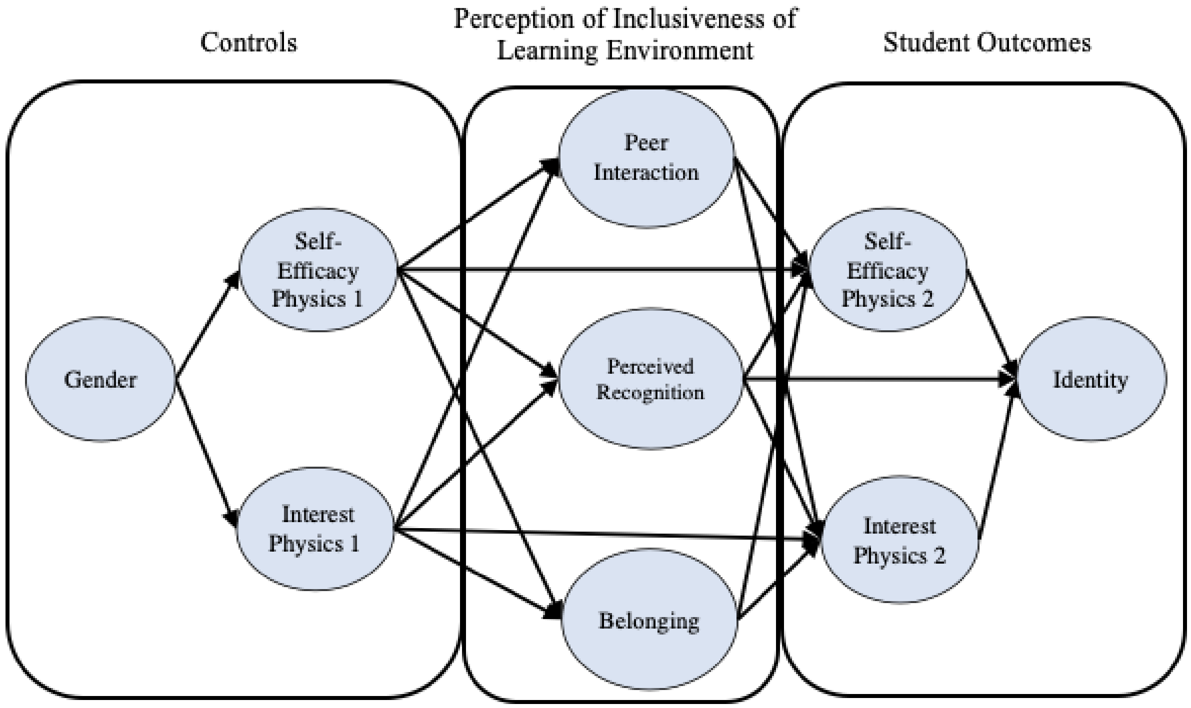


Figure 23 Schematic representation of the model based on the theoretical framework. From left to right, all possible regression paths were considered. However, only some (not all) of the regression paths are shown.

11.2 Methodology

11.2.1 Participants

In this study, we analyzed results from 873 students who completed a motivational survey at the end of the semester in introductory algebra-based physics 1 and physics 2 over two years. These courses are typically taken by students primarily on the bioscience track in their junior or senior year of undergraduate studies, with approximately 50%-70% of students expressing a desire to pursue future careers in health professions. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the university data, the participants were 38% male and 62% female students. These percentages are consistent with the typical number of women and men who receive college degrees in biology in the US [293]. The gender data provided by the university include only binary options of “male” and “female”. We recognize that gender is a socio-cultural and nonbinary construct; however, the data provided by the university only included binary options (less than 1% of the students did not provide this information and thus were not included in this study).

11.2.2 Instrument Validity

This study measured female and male students’ perception of the inclusiveness of the learning environment including their sense of belonging, perceived recognition, and interaction with their peers and how they explain students’ motivational beliefs for students enrolled in introductory algebra-based physics courses for bioscience majors. Students motivational beliefs include their physics identity, self-efficacy, interest. The survey items were constructed from items validated by others [170, 175, 236] and re-validated in our own context using one-on-one student interviews [26], exploratory factor analysis (EFA), confirmatory factor analysis (CFA) [182], analyzing the Pearson correlation between different constructs [182], and using Cronbach alpha [176]. The *physics identity* questions evaluated whether the students see themselves as a physics person [230]. The *physics self-efficacy* questions measured students' confidence in their ability to answer and understand physics problems [170-172, 230]. The *interest in physics* questions measured students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *sense of belonging* questions evaluated whether students felt like they belonged in the introductory physics classroom [50, 175]. The *perceived recognition* questions measured the extent to which the students thought that their instructors and TAs see them as a physics person [230]. Lastly, the *peer interaction* questions measured whether students thought that working with their peers was beneficial, e.g., for increasing their confidence and enthusiasm to do physics [276, 277]. The physics identity instrument only included one question, which is consistent with past studies since it has been difficult to make other questions that factor in this category in exploratory factor analysis [5, 6, 249, 250]. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) except for the sense of belonging questions which were designed on a scale of 1 to 5 to keep them consistent with the original survey [173]. A lower score was indicative of a negative endorsement of the survey construct while a higher score was related to a positive belief in the construct. Some of the items were reverse coded.

After performing an EFA to ensure that the items factored according to different constructs as envisioned, a CFA was conducted to establish a measurement model for the constructs and used in SEM. The squares of the CFA factor loadings (lambda) indicate the fraction of variance explained by the factors. The model fit indices were good and all of the factor loadings (lambda) were above 0.6, which indicates good loadings [182]. The results of the CFA model are shown in Table 28.

Table 28 Survey questions for each of the constructs along with factor loadings from the Confirmatory Factor Analysis (CFA) results for all students (N = 873). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity, peer interaction, and perceived recognition questions was strongly disagree, disagree, agree, and strongly agree. The rating scale for the physics belonging questions was not at all true, a little true, somewhat true, mostly true, and completely true. All *p*-values (of the significance test of each item loading) are $p < 0.001$.

Construct and Item	Lambda
Physics Identity	
I see myself as a physics person	1.00
Physics Self-Efficacy	
I am able to help my classmates with physics in the laboratory or recitation	0.64
I understand concepts I have studied in physics	0.71
If I study, I will do well on a physics test	0.73
If I encounter a setback in a physics exam, I can overcome it	0.66
Physics Interest	
I wonder about how physics works	0.70
In general, I find physics†	0.81
I want to know everything I can about physics	0.76
I am curious about recent discoveries in physics	0.71
Physics Perceived Recognition	
My physics instructor and/or TA sees me as a physics person	1.00
Physics Belonging	
I feel like I belong in this class	0.80
I feel like an outsider in this class	0.68
I feel comfortable in this class	0.85
I feel like I can be myself in this class	0.69
Sometimes I worry that I do not belong in this class	0.61
Physics Peer Interaction My experiences and interactions with other students in this class…	

Made me feel more relaxed about learning physics	0.75
Increased my confidence in my ability to do physics	0.95
Increased my confidence that I can succeed in physics	0.94
Increased my confidence in my ability to handle difficult physics problems	0.85

† the rating scale for this question was very boring, boring, interesting, very interesting.

Zero-order pair-wise Pearson correlations are given in Table 29. These Pearson's *r* values signify the strength of pairwise relationships between variables. The inter-correlations vary in strength, but none of the correlations are so high that the constructs cannot be separately examined. The only high inter-correlations were between post interest in physics 1 and post interest in physics 2 (0.89) and between physics identity in physics 2. Prior research has shown there is a high correlation in interest over the course of the introductory physics classes [26, 294]. The correlation is low enough that they should be considered separate constructs.

Table 29 Pearson inter-correlations are given between all the predictors and outcomes. All p-values are < 0.001.

	Pearson Correlation Coefficient							
Observed Variable	1	2	3	4	5	6	7	8
1. Post Self-Efficacy in physics 1	--	--	--	--	--	--	--	--
2. Post Interest in physics 1	0.58	--	--	--	--	--	--	--
3. Perceived Recognition in physics 2	0.30	0.32	--	--	--	--	--	--
4. Peer Interaction in physics 2	0.35	0.28	0.36	--	--	--	--	--
5. Belonging in physics 2	0.52	0.38	0.41	0.63	--	--	--	--
6. Post Self-Efficacy in physics 2	0.71	0.46	0.52	0.67	0.79	--	--	--
7. Post Interest in physics 2	0.45	0.89	0.41	0.39	0.46	0.69	--	--
8. Physics Identity in physics 2	0.46	0.52	0.57	0.37	0.45	0.58	0.56	--

11.2.3 Analysis

Initially, we compared female and male students' mean scores of the predictors and outcomes for statistical significance using t-tests and for the effect size using Cohen's d [182]. Cohen's d is $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students, and σ_{pooled} is the pooled standard deviation for all students. In general, d = 0.20 indicates a small effect size, d = 0.50 indicates a medium effect size, and d = 0.80 indicates a large effect size [182].

To quantify the statistical significance and relative strength of our framework's path links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. SEM is a statistical method consisting of two parts that are completed together; a measurement part which consists of CFA and a structural part which consists of path analysis. Path analysis can be considered an extension of multiple regression analysis, but it allows one to conduct several multiple regressions simultaneously between variables in one estimation model. SEM also allows us to predict multiple outcomes simultaneously. SEM allows us to calculate the overall goodness of fit and for all estimates to be standardized simultaneously so there can be a direct comparison between different structural components. Thus, we are able to test more complicated models than we would with multiple regression analysis. A full SEM model combines this path analysis with CFA, allowing researchers to test the validity of their constructs (using CFA) and the connections between these constructs (using path analysis) in a single model with a single set of fit indices. We report the model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [139].

Initially, we performed gender moderation analysis by conducting multi-group SEM, i.e., the model estimates were performed separately for men and women to check whether any of the relations between variables show differences across gender by using “lavaan”[138]. In particular, our moderation analysis was similar to our mediation model in Figure 23 but there was no link from gender, instead, multi-group SEM was performed separately for women and men simultaneously.

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

When the model is more complex than the preceding example of smoking and lung cancer as in our SEM model, checking to make sure there are no gender moderation effects involves checking that there are no gender moderation effects for both the measurement and structural parts. For the measurement part, to check for measurement invariance in each step of gender moderation analysis, we fixed different elements of the measurement part of the model to equality across gender. Then we compared the results to the previous step when they were allowed to vary between groups (i.e., for women and men) separately using the Likelihood Ratio Test [187]. A non-significant *p*-value at each step indicates that the fit of this model is not appreciably worse than that of the model in the previous step, so the more restrictive invariance hypothesis (when the parameters are set to the same values for women and men) is retained. Therefore, setting those different elements of the measurement part of the model to equality across gender is valid, which means that estimates are not statistically significantly different across groups (i.e., women and men).

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

Similar to “weak” and “strong” measurement invariance for the measurement part, the predictive relationship (regression path) between two variables is tested for the two groups (e.g., women and men) simultaneously. If the predictive relationship is different for the groups (i.e., the values of the regression coefficients (β) are not the same for the predictive relationship between the two constructs for women and men), then there is a gender moderation effect in the model. If moderation does not show differences by gender in any of these steps (measurement invariance holds and testing for regression coefficients shows that they can be set equal for women and men), we can utilize a gender mediation model (see Figure 23). In other words, we can interpret our model the same way for both men and women, and any gender differences can be modeled using a separate gender variable.

In our multi-group SEM model, we found a non-significant *p*-value in each step, and thus measurement invariance holds and the regression coefficients for women and men can be set equal, i.e., there are no moderation effects by gender (for men and women) in our models. Thus, we concluded that our SEM model can be interpreted similarly for men and women and we can use gender mediation analysis (instead of doing moderation by gender). Therefore, we tested the theoretical model in mediation analysis, using gender as a variable (1 for male and 0 for female) directly predicting items to examine the resulting structural paths between constructs (a schematic representation of the path analysis for the gender mediation model is shown in Figure 23). In the mediation analysis, if there are paths from gender to any of the constructs as we found in our results discussed in the next section, it implies that women and men did not have the same average value for those constructs controlling for all constructs to the left. However, it is important to note that all of the item factor loadings and regression coefficients between the constructs are the same for women and men (as found from the gender moderation analysis which preceded the mediation analysis).

11.3 Results and Discussion

11.3.1 Gender differences in predictors and outcomes

We find statistically significant differences in all predictors and outcomes disadvantaging female students (Table 30). This pattern is similar to what we find in calculus-based physics courses [59] by the end of physics 2 despite the fact that in our investigation, women are the majority in the algebra-based courses for bioscience majors (62%). Since a student's physics self-efficacy, interest, and identity can impact not only their performance in that course but also impact students' future career plans, more should be done in the physics classroom to eliminate the gender gap in these motivational factors by creating an equitable and inclusive learning environment.

Table 30 Mean predictor and outcome values by gender and effect sizes (Cohen's d) by gender. The *p*-values are indicated by no superscript for $p < 0.001$ and superscript "a" for $p = 0.001$.

Predictors and Outcomes (Score Range)	Mean		Cohen's *d*
	Male	Female	
Post Self-Efficacy in physics 1 (1-4)	2.98	2.73	0.49
Post Interest in physics 1 (1-4)	2.81	2.38	0.71
Perceived Recognition in physics 2 (1-4)	2.31	2.06	0.33
Peer Interaction in physics 2 (1-4)	2.94	2.79	0.24^{a}
Belonging in physics 2 (1-5)	3.69	3.45	0.28
Post Self-Efficacy in physics 2 (1-4)	2.94	2.73	0.40
Post Interest in physics 2 (1-4)	2.77	2.32	0.73
Physics Identity in physics 2 (1-4)	2.19	1.85	0.45

11.3.2 SEM path model

We initially conducted moderation analysis between variables using multi-group SEM between female and male students to investigate if any of the relationships between the variables were different across gender. There were no group differences between female and male students at the level of weak, and strong measurement invariance at the level of regression coefficients, so we proceeded to mediation analysis.

Then we used mediation analysis to investigate the extent to which gender differences in students' motivational outcomes at the end of the introductory physics courses (self-efficacy, interest, and physics identity) were mediated by differences in students' initial self-efficacy, interest, and perception of the inclusiveness of the learning environment of the course.

11.3.3 Model 1

In Model 1 (Figure 24), the students' perceived recognition, peer interaction, and sense of belonging were part of the inclusiveness of the learning environment in the class that mediated student motivational outcomes. The model fit indices indicate a good fit to the data: CFI = 0.931, TLI = 0.921, RMSEA = 0.053, SRMR = 0.043. In this model, we found no direct effects from gender to any of the student motivational outcomes: self-efficacy, interest, and identity in physics 2. We found that gender only had direct connections to self-efficacy ($\beta = 0.26$) and interest ($\beta = 0.37$) in physics 1. To expand further, the statistically significant path from gender to self-efficacy in physics 1 means that men are predicted to have higher mean values in their self-efficacy than women. The reason that there is no direct path from gender to students' identity in physics 2 is that the gender differences in students' physics identity (Table 30) are statistically non-significant when controlling for the other constructs in the model.

In our model, self-efficacy, interest, and perceived recognition influence physics identity at the end of physics 2 directly. In addition, the total indirect path for physics identity was found by adding all of the indirect paths together. For example, there are two indirect paths from peer interaction to identity. The first path goes from peer interaction → self-efficacy physics 2 → identity ($0.25*0.23 = 0.06$). The second path goes from peer interaction →interest physics 2 → identity ($0.11*0.30 = 0.03$). Therefore, the total indirect path from peer interaction to identity is $0.06 + 0.03 = 0.09$. Additionally, identity has a total indirect path from belonging of 0.08 and a total indirect effect from perceived recognition of 0.07.

Interest in physics 2 has the largest direct from interest in physics 1 ($\beta = 0.83$) with smaller direct effects from peer interaction ($\beta = 0.11$) and perceived recognition ($\beta = 0.10$). Although interest in physics 2 is mainly correlated with interest in physics 1, it does not mean that interest cannot be changed throughout these courses. Instructors may be able to positively influence students' peer interaction and perceived recognition, which predict students' interest in physics at the end of the course. One possibility to improve students' interest in physics is to provide more problems in class that relate to students' interests and career paths.

Self-efficacy in physics 2 has direct effects from self-efficacy in physics 1 ($\beta = 0.40$), belonging ($\beta = 0.35$), peer interaction ($\beta = 0.25$), and a small effect from perceived recognition ($\beta = 0.17$). Self-efficacy is important for students' persistence in the class and future careers. Since the learning environment can have influence over students' self-efficacy, it is important for instructors to try and improve students' sense of belonging, peer interaction, and perceived recognition.

While prior studies have investigated gender differences in students' self-efficacy, interest, and identity in introductory physics courses where women are underrepresented [5, 59], most have not taken into account all these factors in the students' perception of the inclusiveness of their learning environment. From our model, we find that students' perceived recognition directly predicts students' identity, self-efficacy, and interest at the end of the physics 2 course. Students' belonging and peer interaction directly predict students' self-efficacy and interest while indirectly predicting students' identity. Additionally, from Table 30, women may have more negative experiences with regard to perceived recognition from instructors and TAs and have lower scores in identity than men. Thus, these findings suggest that instructors have the potential to improve these perceptions of the inclusiveness of the learning environment factors in the classroom that predict student motivational outcomes.

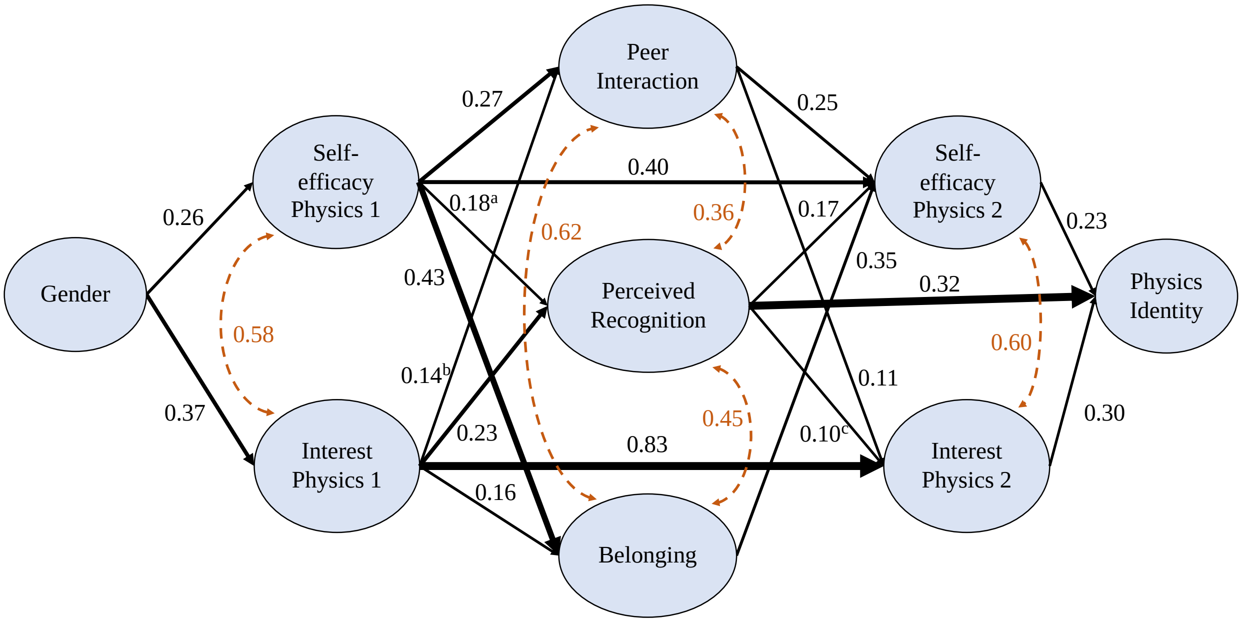


Figure 24 Result of the path analysis part of the SEM showing mediation between gender and motivational outcomes in physics through various mediating factors. The line thickness is a qualitative measure of the

relative magnitude of β values. All p-values are indicated by no superscript for *p* <0.001, superscript "a" for *p* = 0.001, "b" for *p* = 0.011, and "c" for *p* = 0.002.

11.3.3.1 Adding in High School Factors

We also analyzed a model to investigate if additional aspects of student motivational outcomes can be explained by their prior high school academic measures provided during college admissions. In this model visually represented in Figure 25, we added students' SAT math score and high school GPA as control factors. Gender has a small direct effect on both SAT Math (β = 0.14), and high school GPA (β = -0.11), which means women have a slightly higher high school GPA than men while men have a slightly higher SAT Math score than women. SAT Math only has a direct effect on self-efficacy in physics 1 (β = 0.23) whereas high school GPA only has a small direct effect on belonging in physics 2 (β = 0.11). Almost all other direct effects (and indirect effects) stayed the same from the first model with some minor changes in the value of the direct effect (for instance, the line from self-efficacy in physics 1 $\rightarrow$ perceived recognition went from 0.18 in the first model to 0.20 in this model). Thus, we conclude that these additional academic factors do not have a significant influence on student motivational outcomes. We can analyze it more clearly when we look at the variance explained in each outcome (see Table 31).

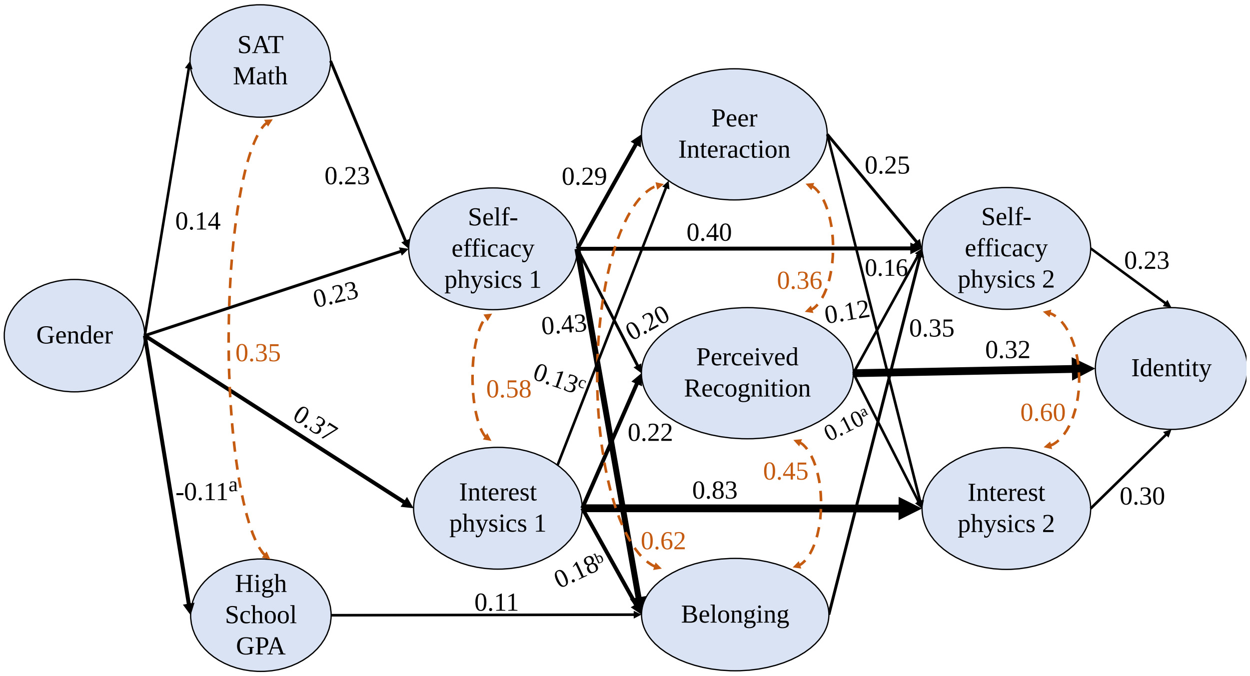


Figure 25 Result of the path analysis part of the SEM showing mediation between gender and motivational outcomes in physics through various mediating factors. The line thickness is the relative magnitude of β values. All *p*-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p = 0.002$, "b" for $p = 0.001$ and "c" for $p = 0.017$.

11.3.4 Variance explained by the models

After constructing the models, we calculated the coefficient of determination (adjusted R^2), which allows us to analyze the proportion of variance explained by each factor (Table 31). This allows us to analyze if adding additional academic outcomes explains more variance in the student outcomes for self-efficacy, interest, and identity in physics 2. We found that adding high school factors explained about the same amount of variance in post self-efficacy in physics 1 (from 0.07 to 0.12), perceived recognition (0.13 to 0.13), and belonging (0.29 to 0.31). Additionally, high school factors did not explain any more of the variance in any of the student motivational outcomes (self-efficacy, interest, or identity in physics 2). This is important since these motivational factors have the potential to not only influence students’ performance in the course but also their future career choices. Since high school academic measures don’t predict student motivational outcomes, the inclusiveness of the learning environment factors that instructors can influence have a key role to play in predicting the outcomes.

Table 31 Adjusted coefficients of determination (Adjusted R^2) for all variables in the two models pertaining to the impact of the inclusiveness of the learning environment. Model 1 is from part a above and the other model (+ H.S. Factors) is from part b above. H.S. refers to high school. All *p*-values are < 0.001

Variable	Adjusted R^2	
	Model 1	+ H.S. Factors
High School GPA	-	0.01
SAT Math	-	0.02
Post Self-Efficacy in physics 1	0.07	0.12
Post Interest in physics 1	0.13	0.14
Perceived Recognition in physics 2	0.13	0.13
Peer Interaction in physics 2	0.13	0.14
Belonging in physics 2	0.29	0.31
Post Self-Efficacy in physics 2	0.82	0.81
Post Interest in physics 2	0.83	0.83
Physics Identity in physics 2	0.49	0.49

11.4 Implications and Future Directions

In this research involving both descriptive and inferential quantitative analyses, we find gender gaps in perceptions of the inclusiveness of the physics learning environment and physics motivational beliefs disadvantaging women in introductory physics courses for bioscience majors. This is despite the fact that women are not outnumbered by men in these courses. The results are similar to what has been found earlier in introductory calculus-based courses in which women are severely underrepresented [5, 59]. Our SEM models show that perception of the inclusiveness of the physics learning environment factors (sense of belonging, perceived recognition by instructors/TAs and perception of the effectiveness of peer interaction) are important to help explain student motivational outcomes with regard to physics self-efficacy, interest, and identity at the end of physics 2. Physics instructors have the ability to influence these learning environment factors and help improve the experiences of women in their classes. These factors influence each other as well, so if an instructor can improve students' peer interaction, possibly by allowing students to work in groups equitably during class such that there is positive inter-dependence, it could positively influence students' sense of belonging as well. Thus, if instructors can provide support and improve the factors they can control in their physics classrooms, they have the potential to change student motivational outcomes for the better and make their classrooms more equitable and inclusive in the process.

The motivational belief gaps may at least partly be due to physics instructors and teaching assistants unwittingly reinforcing gender stereotypes and biases about who can excel in physics and communicating lower expectations for women. They must recognize that what is important is not what their intentions are but the impact they are having on the students. Therefore, instructors and TAs need to create a learning environment that emphasizes recognizing their students positively, allowing for positive peer interactions, and providing a space where all students can feel like they belong in physics regardless of their gender or other demographic characteristics. From our results, each perception of the inclusiveness of the learning environment factor plays an important role in predicting students' self-efficacy, interest, and identity in physics. We note that student perception of the learning environment does not only consist of what happens in the classroom. Student interactions with each other while they do homework, students' experiences in an instructor or TA’s office hours, interactions between students and the instructor over email, and other circumstances all contribute to students’ perception of the inclusiveness of the physics learning environment. All of those interactions can affect students’ identity, self-efficacy, and interest in physics at the end of the course.

There are a variety of ways that TAs/instructors can improve student interactions and the learning environment in their physics courses. Instructors can influence students’ peer interaction by providing time for the students to work together in class in an equitable and inclusive learning environment and making sure all voices are heard equally when discussing problems as a whole group. One strategy instructors can use to encourage equal contribution in group work is to assign each student a role that rotates throughout the course. These changes could improve student perception of the inclusiveness of the learning environment and increase students' motivational beliefs in the classroom if they observe that their ideas are respected by other students in the classroom. This is especially important when implementing active learning pedagogies in the classroom. If active-learning pedagogy is not implemented using teaching strategies that are equitable and inclusive, the gender gap in students’ performance may increase disadvantaging women [92]. It is also important for instructors to emphasize that struggling is normal during the learning process, and it is the stepping-stone to learning new things. Therefore, students should embrace their struggles. Additionally, short interventions, e.g., sense of belonging and mindset interventions, have been shown to eliminate gender performance gaps [151, 186] and have lasting effects beyond the class in which they are implemented [185].

The courses in this study were traditionally taught lecture-based courses. It would be important to investigate these models in evidence-based active-engagement courses (e.g., studio courses). In the future, it will also be useful to investigate, e.g., how students’ perceptions of peer interaction depend on the gender composition of the groups they were interacting with (same sex groups vs mixed groups vs working alone) and how different evidence-based active engagement classes affect these findings.

12.0 Developing an innovative sustainable science education ecosystem: Fostering equitable and inclusive learning environments

12.1 Introduction

There is an urgent need for workforce growth in science, technology, engineering, and math (STEM) fields in many countries particularly because the global economy has become increasingly dependent on rapid innovations in the STEM fields. However, individuals from certain demographic groups, e.g., women and ethnic/racial minority (ERM) students, have largely been left out from contributing to these exciting fields and continue to be severely underrepresented in STEM majors [64, 65] and careers [295]. This lack of diversity not only disadvantages these individuals from contributing to the STEM enterprise, it negatively impacts the productivity and innovation in the STEM workplace [296, 297]. Prior research suggests that gender diversity in the workforce is associated with increased revenue for organizations [297] and an organization that values the talents of diverse individuals with different backgrounds can create fertile grounds for disruptive innovation and better harness the unmet needs in under-leveraged markets [296]. Thus there has been a focus on investigating the experiences and participation of traditionally underrepresented students, e.g., women and ERM students in many STEM fields [52, 79, 115, 145, 149, 153, 260, 261] to improve equity and inclusion in these fields and increase the participation of these underrepresented groups in the STEM workforce, which is critical for innovation.

However, increasing the enrollment of traditionally underrepresented people in STEM fields is not sustainable unless the environment is equitable and inclusive and there is an established supportive culture so that all individuals regardless of their demographic group feel that they belong and can freely contribute their ideas. For example, some strategies to create an equitable and inclusive environment and increase innovation include ensuring that everyone is respected, valued, and heard, making it safe to propose novel ideas without any fear of judgment, and giving actionable feedback [296]. Creating this type of sustainable STEM ecosystem in which everyone can succeed begins with creating an innovative science education ecosystem with research-based active learning pedagogies that centers equity and inclusion [219, 298-300]. In particular, it is important to keep in mind that prior research shows that if the active learning pedagogy is not implemented using teaching strategies that are equitable and inclusive, men have been shown to not only dominate in asking and responding to questions in class but they also dominate while working in groups, which can lower women's self-efficacy [219]. In addition, in the lab context, men and women have been shown to fall into gender roles when splitting up the work which may disadvantage women [52]. Therefore, to make a sustainable science education ecosystem, the learning environment in STEM classes must be curated and implemented with equity and inclusion as a foundational and central tenet.

To develop an innovative sustainable science education ecosystem, our conceptualization of equity in STEM learning includes three pillars: equitable access and opportunity to learn, equitable and inclusive learning environment, and equitable outcomes. Thus, by equity in STEM 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 the learning outcomes are equitable. For there to be equitable learning outcomes, students from all demographic groups (e.g., regardless of their gender) who have the pre-requisites to enroll in STEM courses should have comparable learning outcomes. This conceptualization of equitable outcome is consistent with Rodrigues et al.'s equity of parity model [72]. An equitable and inclusive learning environment should be student-centered in which students are provided appropriate support and have an equal sense of belonging regardless of their prior preparation. It would also ensure that students from all demographic groups enjoy the hands-on and minds-on learning and embrace challenges as opportunities to grow their knowledge instead of being threatened by them. Equitable learning outcomes for STEM majors include the ability of STEM courses to empower all students and make them passionate about pursuing further learning and careers in related areas. We note that these three pillars 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.

Our conceptualization of equity in STEM learning is mindful of the pervasive societal stereotypes and biases about physics as well as the lack of role models that can have a detrimental psychological impact on women who are severely underrepresented. In general, when students struggle to solve challenging problems, they often respond in one of two ways. Some question whether they have what is needed to excel in STEM. Others enjoy the struggle because it means that they are tackling new concepts and learning. The negative reaction is a manifestation of a fixed mindset (believing that intelligence is immutable and struggling is a sign of a lack of intelligence), whereas the positive reaction is the sign of a growth mindset (believing in the brain's capabilities can grow with deliberate effort). In an inequitable and non-inclusive learning environment, due to societal stereotypes and lack of role models, the marginalized students are more likely than others to fall prey to the fixed mindset trap and view their struggle with challenging problems in a negative light.

We note that not only should STEM courses have learning outcomes based upon STEM related knowledge and skills we want students to learn but also those that focus on whether all students (and especially those from marginalized groups) have a high self-efficacy, interest, positive perception of recognition from others such as instructors, sense of belonging, peer interaction, and identity as people who can excel in physics. Drawing analogy with sports, we note that to help players excel in any game, such as tennis, coaches must ensure both good defense and offense. Likewise, helping students learn requires that instructors equip all students with both defensive and offensive strategies. Instructors can strengthen students' defenses by creating equitable and inclusive learning environments in which all students have high beliefs. Only if the learning environment is equitable and inclusive so that all students have strong defenses about learning can they effectively engage with the offense by tackling challenging problems and developing problem solving, reasoning, and meta-cognitive skills. In the absence of equitable and inclusive classrooms, students without strong defenses are unlikely to risk struggling with challenging problems and engage fully with hands-on and minds-on activities.

Here we discuss a research study involving physics classrooms at a large research university in the US to understand women's and men's perceptions of the inclusiveness of the learning environment, and how it predicts student outcomes (including their performance and beliefs) in introductory physics courses. Lessons learned from our investigation about equity and inclusion in physics classrooms can be invaluable for all STEM disciplines that suffer from similar issues with students from marginalized demographic groups, particularly because societal stereotypes and biases are some of the worst in physics, a field whose history is often told through the stories of brilliant men [301]. These stereotypes, biases, and lack of role models can contribute to lower student beliefs (such as self-efficacy, interest, and identity) and performance outcomes (e.g., grades) for women in physics [1-3, 53, 238] unless explicit efforts are made to make the learning environment equitable and inclusive. We emphasize that in addition to performance outcomes, students' STEM beliefs are important to investigate since students' beliefs in different STEM domains can influence their continuation in related courses, majors, and careers [43, 118, 156, 238, 240, 242, 245]. Student beliefs such as students' identity, self-efficacy, and interest in a particular STEM field are important for students' career interests [32, 54, 230], learning [40], and continuation in STEM fields [39, 159, 286, 287]. Furthermore, in physics, prior research shows that gender gaps can persist in many of these beliefs as well as in the performance outcomes at the end of the course with certain student populations [1-3, 294]. Thus, in order to create a sustainable STEM education ecosystem, equity and inclusion must be central to making sure that innovative hands-on and minds-on activities benefit all students and that all students (and particularly those from marginalized groups) can be supported equitably.

Therefore, investigating these beliefs and performance outcomes for students from different demographic groups, e.g., women and men, can provide important information on how students are persisting in these STEM courses and how educators can create an innovative sustainable ecosystem that centers on an equitable and inclusive learning environment in which students from all demographic groups have comparable outcomes. Prior work has mainly investigated students' performance outcomes as well as self-efficacy, interest, and identity and connections between these factors in physics courses in which women are underrepresented [58, 59, 97]. However, societal stereotypes and biases may impact female students' beliefs and performance even in STEM courses in which they are not numerical minorities if the learning environment is not equitable and inclusive. Therefore, research is necessary to examine how mechanisms for structuring courses in instructors' control can influence women's and men's course outcomes in contexts not frequently studied in the past, e.g., introductory physics courses for students on the bioscience track in which women are not numerical minorities. The findings of this research can provide guidelines for developing an innovative and sustainable hands-on and minds-on science education ecosystem that fosters equitable and inclusive learning environments. Since in the research presented here, we investigate students' motivational beliefs and their perception of the inclusiveness of the learning environment, we start with a brief background on each.

Self-efficacy in a particular discipline is a student's belief in their ability to solve a particular problem or goal [27, 28]. Self-efficacy has been shown to impact students' engagement, learning, and persistence in science courses [14, 15, 27, 29-42]. When tackling difficult problems, students with high self-efficacy tend to view the problems as challenges that can be overcome, whereas people with low self-efficacy tend to view them as personal threats to be avoided [27]. However, in introductory physics courses in which women are underrepresented, studies have found a gender gap in self-efficacy favoring men that sometimes widens by the end of the course even in interactive engagement courses [11, 29, 43, 91, 92, 110]. Similarly, interest in a particular discipline may affect students' perseverance, persistence, and achievement in a course [16, 38, 41, 44-46]. One study showed that changing the curriculum to stimulate the interest of the female students helped improve all of the students' understanding at the end of the year [47]. Within the expectancy-value theory, interest and competency beliefs (closely related to self-efficacy) are connected constructs that predict students' academic outcomes and career expectations [48]. Additionally, perceived recognition has been shown to play an important role in women's motivation [50]. However, studies have shown that female students do not feel recognized appropriately even before they enter college [1, 4, 24]. In a study on students' perception of support, teacher support was more strongly linked to the motivation and engagement of girls than boys [50].

A student’s identity in STEM disciplines is important to study since it plays a key role in students’ participation in classes as well as career decisions [5, 49, 54-57]. For example, “physics identity” has been studied in physics classes and has been shown to be connected with a student’s self-efficacy, interest, and perceived recognition [5, 58, 59, 134]. The science identity framework draws on the work of Gee, who defined an individual’s identity as being recognized as a certain kind of person in a given context and emphasized that identity can change over time [56]. It has been adapted in the science context to address the identities of both students and scientists [55]. Carlone and Johnson’s science identity framework includes three interrelated dimensions: competence (“I think I can”), performance (“I am able to do”), and recognition (“I am recognized by others”) [55]. Hazari et al. modified the framework by adapting it to physics specifically rather than science more generally [5]. “Competence” and “performance” were defined as students’ beliefs in their ability to understand the subject and students’ belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a good physics student. Lastly, a fourth dimension, interest, was added to the framework since students have highly varying levels of interest in physics [60, 61]. In later studies by Hazari for introductory students, performance and competence are combined into one variable [58]. In a slightly reframed version of Hazari’s physics identity framework by Kalender et al. [59], performance/competence was framed as self-efficacy (closely related to competency belief). Additionally, recognition was framed explicitly as “perceived recognition” by students for clarity.

Students’ identity in STEM disciplines plays an important role in students’ participation in classes as well as career decisions [5, 49, 54-57]. However, prior studies have shown that it can be more difficult for women to form a physics identity than men [1-3, 6]. This could be due to gender stereotypes and biases about who can excel in physics courses. In general, the image of a physicist is portrayed as male, which can make women feel less welcome and accepted in the physics community. In addition, the innate abilities of genius and brilliance are often seen as important factors to succeed in physics [23]. However, genius is often associated with boys [169], and girls from a young age tend to shy away from fields associated with innate brilliance or genius. Archer et al. [1] investigated the impact of physics-related cultural attributions on girls' or women's decisions to pursue physics and reported that science-keen girls or young women who name physics as their favorite subject slowly lose their interest due to alienation, discrimination, and gender-biased beliefs about physics. All of the stereotypes and biases can influence women's perception of their ability to do physics before they enter the classroom. In addition, faculty members' unconscious gendered beliefs regarding the students' ability can be one source of the threat and alienation that women in STEM experience [26]. One study showed that science faculty members in biological and physical sciences exhibit biases against female students by rating men significantly more competent when the curriculum vitae are identical except for the name of the student being a male or female name [26]. These highly problematic stereotypes and biases are founded in the historical marginalization of certain groups, e.g., women in physics, and continue to manifest today in many ways, including gendered beliefs and barriers to women excelling in physics when there is no explicit focus on making the learning environment equitable and inclusive.

In addition to motivational factors that predict physics identity, other motivational factors that contribute to the student perception of the inclusiveness of the learning environment can also influence how women perform in their STEM classes and beyond [62]. For example, students' interaction with peers has been shown to enhance understanding and engagement in courses. In addition, students' sense of belonging in physics has been shown to correlate with retention and self-efficacy [50, 63], so it is important to understand how it predicts both the performance and motivational outcomes of women and men at the end of the course.

Inequitable outcomes in students' beliefs and performance may be a result of inequitable access to resources, inadequate support, and inequitable learning environments. Thus, it is important to investigate student perception of the inclusiveness of the learning environment in STEM courses to foster an inclusive education ecosystem in which all students regardless of their demographic group affiliations can succeed. The study reported here used structural equation modeling (SEM) to investigate factors in physics 1 that are part of the inclusiveness of the learning environment that can influence student motivational and academic outcomes. We analyzed factors that instructors have control over as part of the inclusiveness of the learning environment in their courses, specifically students' interaction with their peers, students' sense of belonging in the course, and their perceived recognition by others (including instructors and teaching assistants or TAs).

While many studies have investigated gender differences in beliefs in introductory physics courses, most have not considered the inclusiveness of the students' learning environment. The inclusiveness of the learning environment includes experiences students have in the classroom as well as interactions outside of the classroom, such as students' experiences during office hours or via email correspondences with the instructor or TA and students studying or doing homework together. We control for students' high school factors that include their high school grade point average (GPA) and standardized math scores (SAT math score) as well as their self-efficacy and interest at the beginning of physics 1 since these are students' beliefs about physics when they enter the course based on prior experiences. The perception of the inclusiveness of the learning environment consists of the student's perception of the effectiveness of peer interaction, their sense of belonging, and perceived recognition (from instructors, TAs, friends, and family). We discuss an investigation of the students' outcomes pertaining to their physics performance (as measured by the end of the semester grade) as well as their physics self-efficacy, interest, and identity at the end of physics 1 to answer the following research questions. Our final statistical model that shows the path analysis is shown in Figure 26.

RQ1. Are there gendered differences in students' beliefs at the end of the physics 1 course and do they change from the beginning to the end of the course?

RQ2. Do academic measures (e.g., high school grade point average and standardized math scores) predict students' motivational beliefs and performance at the end of the physics 1 course?

RQ3. How does the students' perception of the inclusiveness of the learning environment predict their motivational and performance outcomes at the end of the physics 1 course?

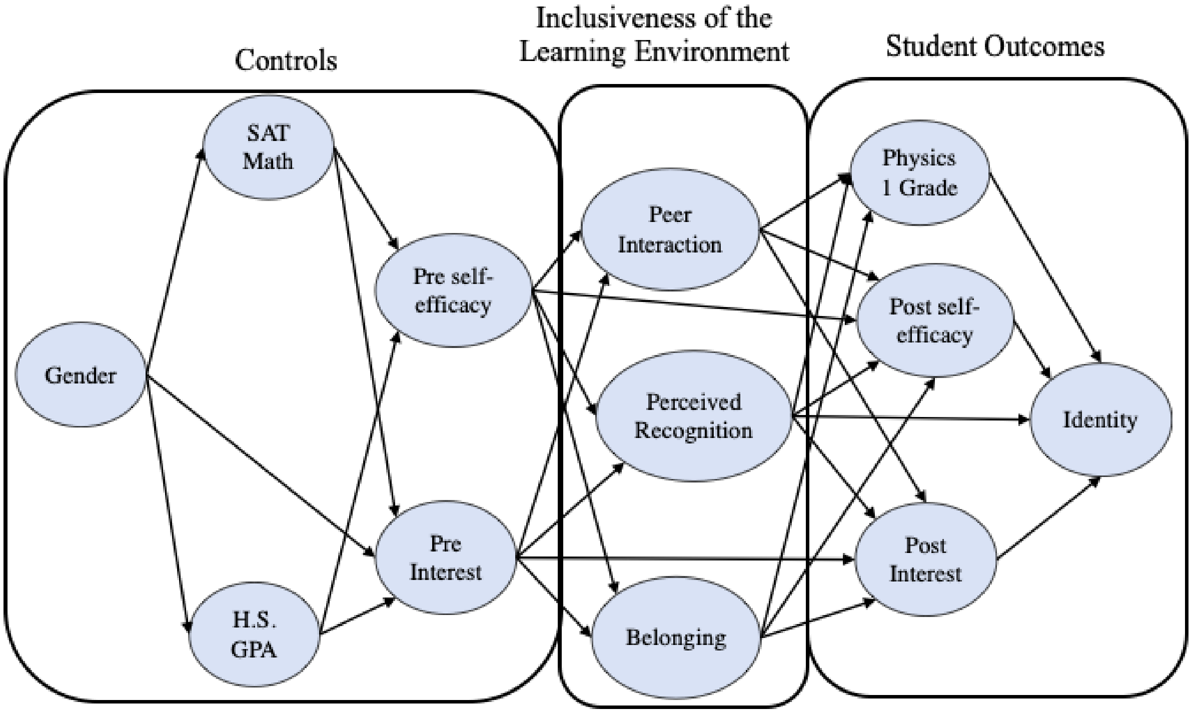


Figure 26 Schematic representation of the path analysis part of the statistical model. From left to right, all possible paths were considered. Some, but not all, of the regression paths are shown.

12.2 Materials and Methods

12.2.1 Participants

In this study, a motivational survey covering the components in our model was administered to students at the beginning and end of one semester of a traditionally taught introductory algebra-based physics 1 course at a large research university in the US. The course is typically taken by students on the bioscience track in their junior or senior year of undergraduate studies. The course had many sections and had three primarily traditionally taught lecture classes and one recitation per week. There was a similar grading policy across sections such that in addition to some points for homework, students' grades were mainly based on 2-3 midterm exams and 1 final exam. In general, there were little to no evidence-based active learning strategies implemented in the courses and there was no intentional effort to make the course equitable and inclusive. We analyzed the data for more than five hundred students who completed the survey. The university provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the university data, the participants were 35% male and 65% female students. We recognize that gender is not a binary construct. However, the data provided by the university only included binary options of male or female (less than 1% of the students did not provide this information and thus were not included in the analysis).

12.2.2 Motivational Survey Instrument Validity

The validated motivational survey instrument used in the study measured students' physics identity, self-efficacy, and interest as well as their perception of the inclusiveness of the learning environment as measured by their sense of belonging, perceived recognition, and interaction with their peers. Thus, the survey questions asked about different aspects of the students' beliefs at two points in time (beginning and end of the course) and student perception of the inclusiveness of the learning environment at the end of the course. Students were asked to answer all of the survey questions with regard to the physics course they were enrolled in. The *physics identity* questions evaluated whether the students saw themselves as a physics person, i.e., a person who can excel in physics [230]. The *physics self-efficacy* questions measured students' confidence in their ability to answer and understand physics problems [170-172, 230, 236]. The *interest in physics* questions measured students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *sense of belonging* questions evaluated whether students felt like they belonged in the introductory physics classroom [50, 175]. The *perceived recognition* questions measured the extent the student thought that other people see them as a physics person [230]. Lastly, the effectiveness of *peer interaction* questions measured whether students thought that working with their peers was beneficial to their confidence to do physics [276, 302]. The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) except for the sense of belonging questions which were designed on a scale of 1 to 5 [173]. A lower score was indicative of a negative endorsement of the survey construct while a higher score was related to a positive belief in the construct.

The survey items were adapted from previous research [170, 174, 175, 236] and revalidated in our own context using one-on-one student interviews, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA) [26], Pearson Correlations, and Cronbach alpha. The survey questions for each construct and factor loadings for each question are given in Table 32.

Table 32 Survey questions for each of the motivational constructs along with factor loadings (lambdas, λ) from the Confirmatory Factor Analysis (CFA) results for all students (N = 501). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity, peer interaction, and perceived recognition questions was strongly disagree, disagree, agree, strongly agree. The rating scale for the physics belonging questions was not at all true, a little true, somewhat true, mostly true, and completely true. All *p*-values (of the significance test of each item loading) are $p < 0.001$.

Construct and Item	λ
Physics Identity	
I see myself as a physics person	1.00
Physics Self-Efficacy	

I am able to help my classmates with physics in the laboratory or recitation	0.60
I understand concepts I have studied in physics	0.74
If I study, I will do well on a physics test	0.76
If I encounter a setback in a physics exam, I can overcome it	0.72
Physics Interest	
I wonder about how physics works†	0.57
In general, I find physics††	0.75
I want to know everything I can about physics	0.74
I am curious about recent discoveries in physics	0.64
Physics Perceived Recognition	
My family sees me as a physics person	0.89
My friends see me as a physics person	0.92
My physics instructor and/or TA sees me as a physics person	0.71
Physics Belonging	
I feel like I belong in this class	0.81
I feel like an outsider in this class	0.74
I feel comfortable in this class	0.82
I feel like I can be myself in this class	0.62
Sometimes I worry that I do not belong in this class	0.69
Physics Peer Interaction	
My experiences and interactions with other students in this class…	
Made me feel more relaxed about learning physics	0.75
Increased my confidence in my ability to do physics	0.93

Increased my confidence that I can succeed in physics	0.96
Increased my confidence in my ability to handle difficult physics problems	0.90

Additionally, Table 33 shows the Pearson's correlation *r* values which signify the strength of the relationship between constructs. The table shows that the correlations vary in strength, but none of the correlations in our survey are so high that the constructs cannot be separately examined. The only high inter-correlations were between pre interest and post interest (0.86). Past work has shown that there is a high correlation in interest throughout the introductory courses in which no explicit hands-on and minds-on efforts are made to make students more interested [26, 294]. All correlations in Table 33 are low enough that they can be considered separate constructs.

Table 33 Pearson correlations are given between all the predictors and outcomes. All *p*-values are < 0.001

	Pearson Correlation Coefficient									
Observed Variable	1	2	3	4	5	6	7	8	9	10
1. SAT Math	--	--	--	--	--	--	--	--	--	--
2. H.S. GPA	n.s.	--	--	--	--	--	--	--	--	--
3. Pre Self-Efficacy	n.s.	n.s.	--	--	--	--	--	--	--	--
4. Pre Interest	n.s.	n.s.	0.65*	--	--	--	--	--	--	--
5. Perceived Recognition	0.15**	n.s.	0.30	0.33	--	--	--	--	--	--
6. Peer Interaction	n.s.	n.s.	0.21**	0.25	0.46	--	--	--	--	--
7. Belonging	0.26	n.s.	0.36	0.25	0.59	0.62	--	--	--	--
8. Post Self-Efficacy	0.30	n.s.	0.41	0.29	0.62	0.62	0.81	--	--	--
9. Post Interest	n.s.	0.132*	0.19*	0.66	0.61	0.47	0.51	0.60	--	--
10. Grade in physics 1	0.48	n.s.	0.13	n.s.	0.37	0.23	0.44	0.49	0.20	--
11. Physics Identity	0.14**	n.s.	0.31	0.38	0.79	0.43	0.58	0.60	0.62	0.29

* = p < 0.05, ** = p < 0.01, no * = p < 0.001, n.s. = non significant (p > 0.05)

12.2.3 Analysis.

Initially, we compared female and male students' mean scores for the predictors and outcomes in our model (see Figure 26) for statistical significance using t-tests and for the effect size using Cohen's d [182]. Cohen's d is defined as $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students and σ_{pooled} is the pooled standard deviation (or weighted standard deviation for men and women) for all students. σ_{pooled} is defined as $\sigma_{pooled} = (\sigma_m{}^2 + \sigma_f{}^2)/2$, where σ_m is the standard deviation for men and σ_f is the standard deviation for women.

To quantify the statistical significance and relative strength of our framework's path links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. We report the model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for the goodness of fit are as follows: CFI and TLI > 0.90, and RMSEA and SRMR < 0.08 [139]. Additional details about SEM can be found in the following papers [62, 303, 304].

12.3 Results

12.3.1 Gendered differences in predictors and outcomes

Statistically significant differences were found in the majority of the predictors and outcomes in favor of male students, except for the high school grade point average which favors women (see Table 34). Additionally, there was no statistically significant difference in the standardized math scores between men and women. The societal stereotypes, biases, and previous experiences women have before they come into the class could be one reason women have lower scores on these beliefs than men even at the beginning of the course (pre). Moreover, student perceptions of the inclusiveness of the learning environment at the end of the course are lower for women than men (see perceived recognition, perceived effectiveness of peer interaction, and sense of belonging in Table 34). Since students' beliefs have important implications for their course engagement and learning outcomes, it is critical for instructors to create a sustainable educational ecosystem in which hands-on and minds-on activities center equity and inclusion to ensure that all students regardless of their demographic group have similar outcomes (both in terms of their performance and their beliefs). In Appendix I we provide the percentage of men and women who selected each response to the questions. This provides a sense of how students answered each question.

Table 34 Mean predictor and outcome values as well as effect sizes (Cohen's d) by gender (N = 292 for women and N = 144 for men for the pre values and N= 327 for women and N = 174 for the post values and high school factors since we only required students to have the post values). Statistical significance (p-values) is given by superscript a for $p = 0.002$, superscript b for $p = 0.171$, and no superscript for $p < 0.001$.

Predictors and Outcomes (Score Range)	Mean		Cohen's *d*
	Male	Female	
High School Grade Point Average (0-5)	4.02	4.17	-0.32^{a}
Standardized Math Score (200-800)	680	670	0.16^{b}
Pre Self-Efficacy (1-4)	3.07	2.83	0.58
Pre Interest (1-4)	2.74	2.45	0.58
Post Self-Efficacy (1-4)	2.92	2.55	0.66
Post Interest (1-4)	2.66	2.27	0.69
Physics 1 Grade (0-4)	3.17	2.89	0.36
Perceived Recognition (1-4)	2.22	1.92	0.43
Peer Interaction (1-4)	2.90	2.58	0.43

Belonging (1-5)	3.58	3.01	0.60
Physics Identity (1-4)	2.04	1.64	0.59

12.3.2 SEM path model

We used SEM to investigate the relationships between the constructs and to unpack each construct's contribution to explaining the self-efficacy, interest, physics identity, and grades of women and men at the end of physics 1. We initially tested moderation analysis of variables using multi-group SEM between female and male students to see if any of the relationships differed across gender. There were no group differences at the level of weak or strong measurement invariance and regression coefficients, so we proceeded to mediation analysis. We used mediation analysis to understand the extent gendered differences in students' outcomes at the end of physics 1 (self-efficacy, interest, grade, and physics identity) were mediated by differences in student's initial self-efficacy, interest, pre-college academic measures (high school grade point average and standardized math scores), and the perception of the inclusiveness of the learning environment in the physics class.

In our model, the inclusiveness of the learning environment (students' sense of belonging, perception of the effectiveness of peer interaction, and perceived recognition) mediated the outcomes. The result of the SEM is presented in Figure 27. The model fit indices indicate a good fit to the data: CFI = 0.924 (>0.90), TLI = 0.914 (>0.90), RMSEA = 0.051 (<0.08), and SRMR = 0.080 (<0.08). Self-efficacy, interest, and perceived recognition had a direct effect on physics identity, and there were no direct effects from gender, grade in physics 1, or any of the pre college and pre test factors. Students' perceived recognition had the largest correlation ($\beta = 0.59$) while self-efficacy (β = 0.13) and interest (β = 0.18) had similar, but smaller direct effects on physics identity.

Additionally, self-efficacy at the end of physics 1 is largely correlated by student’s sense of belonging (β = 0.44) with additional effects from students’ pre self-efficacy (β = 0.26), peer interaction (β = 0.26), standardized math scores (β = 0.14), and perceived recognition (β = 0.12). Similarly, interest at the end of physics 1 had the largest correlation from pre interest (β = 0.71) with smaller effects from peer interaction (β = 0.21), perceived recognition (β = 0.11), and high school grade point average (β = 0.12). Lastly, student’s grade in physics 1was correlated with their standardized math score (β = 0.40), student’s sense of belonging (β = 0.25), and perceived recognition (β = 0.15).

Interestingly, we found that gender only directly correlated with students pre self-efficacy (β = 0.32) and pre interest (β = 0.33) as well as smaller correlations with peer interaction (β = 0.14) and belonging (β = 0.13). The gendered differences in the pre-test values may be partially caused by past experiences and societal stereotypes about physics women have before they enter the physics classroom. However, women have additional negative experiences in the classroom from the perceptions of the inclusiveness of the learning environment factors despite women making up the majority of students in the class.

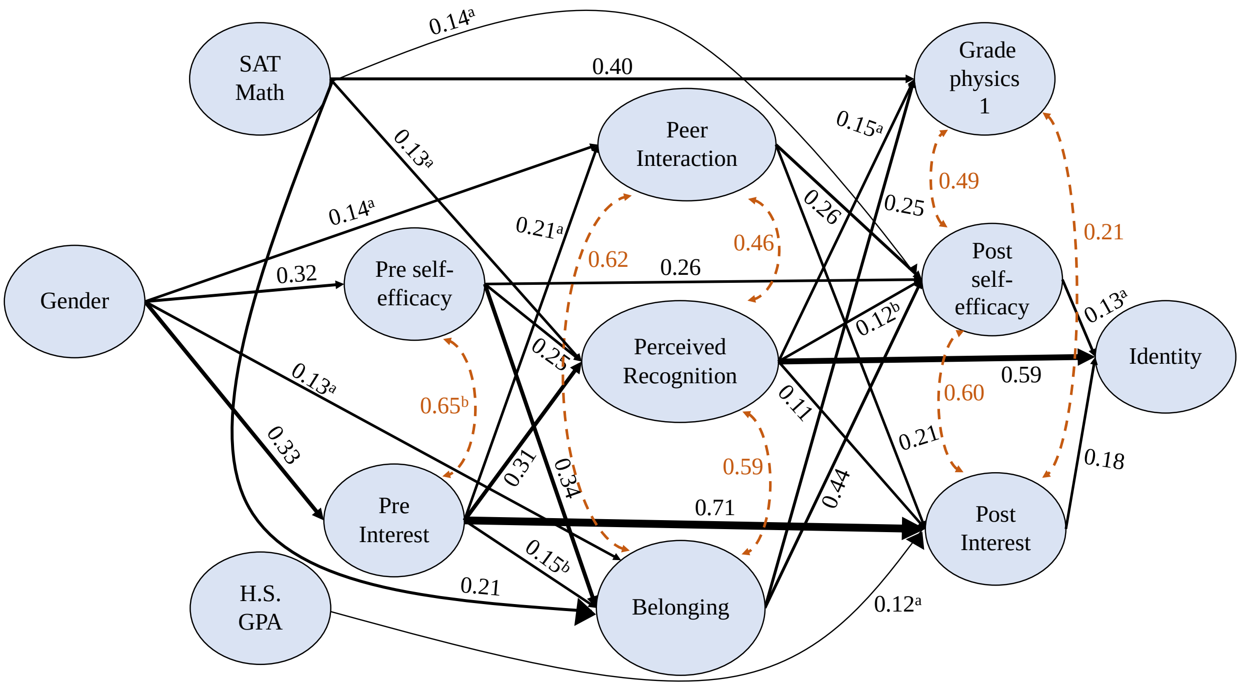


Figure 27 Result of the SEM between gender and outcomes in physics through various mediating factors. Perceived recognition, peer interaction, and belonging are factors in the inclusiveness of the learning environment. The line thickness corresponds to the relative magnitude of β values. All *p*-values are indicated by no superscript for $p < 0.001$, superscript “a” for $p < 0.005$ and superscript “b” for $p < 0.05$ values.

12.4 Discussion and Implications

In accordance with our framework, all students, especially those from marginalized demographic groups such as women, must be given adequate support to excel in their coursework as well as to develop high STEM related motivational beliefs in order to foster a sustainable science education ecosystem and a sustainable STEM workforce. If the STEM learning and work environments do not provide equitable opportunities for individuals from all demographic groups to contribute their talents, quality and innovation will be compromised. In other words, in order to create a sustainable STEM ecosystem, equity and inclusion must be centered so that all students can be supported in order for them to realize their potentials. However, using an example from an introductory physics 1 course for students on the bioscience track at a large university in the US in which women are not underrepresented, we find evidence of an inequitable and non-inclusive learning environment. In particular, we find that women have lower grades and motivational beliefs than men at the end of physics 1. This trend is comparable to the calculus-based physics courses [4, 59] in which women are underrepresented showing these inequities cannot simply be explained by the numerical representation of women and have their roots in stereotypes and biases related to who belongs in these STEM disciplines and who can excel in them as well as the dominant culture in these disciplines that perpetuates the inequities.

In response to **RQ1,** we find that women had lower motivational beliefs and perceptions of the inclusiveness of the learning environment in physics 1. The course learning environment was not equitable and inclusive to eliminate these gender differences. Our finding shows that the percentage of women in a physics course is not enough to create a sustainable educational ecosystem, instead, the learning environment must be equitable and inclusive in order for women (and other marginalized students) to excel. Since students' performance and beliefs in physics courses can impact their future career choices, it is important to create a sustainable educational ecosystem and make physics learning environments equitable and inclusive so that the gaps between marginalized and dominant groups can be eliminated.

In response to **RQ2**, our SEM model indicates that the high school factors (high school grade point average and standardized math scores) predicted students' interest, self-efficacy, perceived recognition, and grade at the end of the course. However, the perception of the inclusiveness of the learning environment factors had larger correlations with students' beliefs at the end of the course.

In response to **RQ3,** the inclusiveness of the learning environment factors were important for explaining students' physics self-efficacy, interest, grade, and identity at the end of the course. Perceived recognition predicted all of the motivational and academic outcomes at the end of the course while the effectiveness of peer interaction predicted students' self-efficacy and interest and sense of belonging predicted students' grades and self-efficacy. Although interest at the end of the course was primarily predicted by interest at the beginning of the course, intstructors have the potential to improve students interest in physics if they explicitly focus on it as a goal. One possible way to influence students' interest is to engage them with problems that relate to their prospective majors and occupations or that are of interest to them in general.

We note that the physics course in this study was a traditionally taught lecture-based course in which student grades heavily depended on two or three midterm exams and a final exam. The courses consisted of 3 hours of lecture per week taught by the instructor and 1 hour of recitation per week taught by a TA. It is important to recognize that even in this traditionally taught lecture-based course, students often received feedback from their instructors in multiple ways, including receiving praise for asking or answering a question in class (which often advantages male students since they dominate these situations) and their interactions with students during office hours or over email. In addition, students interact with their TAs in recitation classes by asking questions about the homework or class material at the start of recitation, when completing group work during recitation, and during the TA's office hours.

Instructors have the ability to positively transform the inclusiveness of the learning environment (which would lead to an increased sense of belonging, effectiveness of peer interaction, and perceived recognition) and make their classes more equitable and inclusive. These beliefs can influence each other as well. For example, if an instructor can improve the students' perception of the effectiveness of peer interaction, by allowing students to work in groups during class and ensuring that all students feel safe, valued, and respected participating in the discussions without the fear of being wrong, they could influence the students' sense of belonging as well. In other words, if instructors can provide support for one of the factors they can most readily control (e.g., effectiveness of peer interaction) and make their classes more equitable and inclusive, they are likely to improve student outcomes in the process. In the hands-on lab, it would be beneficial to have students contribute equally to each task as opposed to splitting the group work so that all students have the opportunity to engage in all the types of work that make up science (as opposed to women becoming secretaries and managers and men doing the tinkering with lab equipment) [112]. Moreover, brief social-psychological classroom interventions have also been shown to eliminate or reduce the gender gap in performance [151, 152, 185, 186].

Other researchers have pointed to the structural changes to implement at the institution and classroom level to make the learning environment more equitable and inclusive and make the science education ecosystem sustainable [47, 253, 305, 306]. Structural changes at the institution-level require centering disadvantaged students in the design of curricula and pedagogies. Instructors can make their courses student-centered, e.g., by adopting pedagogy that focuses on societal implications of physics [47] in addition to providing mentoring/support for students who are marginalized [253]. These students must be provided appropriate mentoring, guidance, scaffolding, and support in college so that the structural hurdles they encounter in STEM fields can be dismantled and they are not put at a disadvantage relative to their privileged peers [306].

In summary, instructors and teaching assistants need to provide an inclusive learning environment that emphasizes recognizing their students, allows for positive peer interactions, and provides a space where all students can feel that they belong. From our analysis presented here, it is clear that student perception of the inclusiveness of the learning environment factors play a central role in predicting not only students' grades but also their self-efficacy, interest, and identity at the end of the course. We emphasize again that it is important to note that student perception of the inclusiveness of the learning environment is not shaped only by what happens in the classroom. Student interactions with each other while they do their homework, students' experiences during an instructor's or TA's office hours, interactions between students and the instructor over email, and other circumstances all contribute to the students' perceptions of the inclusiveness of the learning environment.

We hope that this research conducted in traditionally taught physics classes in the US can serve as an example of how the current science education ecosystem is not sustainable because students from marginalized groups, e.g., women, are continuing to have concerns about the inclusiveness of the learning environment and this perception predicts their performance as well as beliefs at the end of the course. In accordance with our framework, it is important to make intentional efforts to create an innovative equitable and inclusive learning environment to help create a sustainable science education ecosystem so that all students can benefit from the hands-on and minds-on learning regardless of their demographics.

13.0 How perception of learning environment predicts male and female students’ grades and motivational outcomes in algebra-based introductory physics courses

13.1 Introduction and Theoretical Framework

While prior studies have investigated issues pertaining to women’s representation in science, technology, engineering, and math (STEM) fields [19, 77, 114, 115, 145-153, 163], most studies in college physics courses have focused on courses in which women are outnumbered by men. In addition, several studies also reported gender disparity in students’ performance in some STEM disciplines [76, 90, 307]. In these studies, several factors have been studied that negatively impact women and could lead to their underrepresentation and underperformance. Some of the factors include societal stereotypes and biases about who belongs in these fields as well as the potentially intertwined issue of motivational beliefs, such as self-efficacy, interest, and identity pertaining to STEM disciplines [29, 30, 34, 117, 156, 240, 241, 243, 244].

Self-efficacy in a discipline refers to students’ belief in their ability to accomplish tasks or solve problems [27, 28]. Self-efficacy has been shown to impact students’ engagement, learning, and persistence in science courses as well as contribute to the students’ science identity [27, 29, 31, 33, 35, 38, 39, 42]. For example, when tackling difficult problems, students with high self-efficacy tend to view the problems as challenges that can be overcome whereas people with low self-efficacy tend to view them as threats to be avoided [27]. Similarly, interest in a particular discipline may affect students’ perseverance, persistence, and achievement in STEM courses [16, 38, 41, 44-46, 286]. One study showed that changing the curriculum to stimulate the interest of the female students helped improve all of the students’ understanding at the end of the year [47]. In

Eccles’ expectancy-value theory, interest and self-efficacy can influence each other as well as a student’s performance in a class [48, 308]. This is important in physics courses since several studies have shown performance gaps between men's and women’s grades and scores on conceptual tests in calculus-based and algebra-based physics courses [8-13]. Some hypothesize that causes of this gender gap include career goals, prior preparation, and motivational beliefs [12-22].

In addition, self-efficacy and interest have been shown to influence students’ identity in STEM courses. Science identity is defined as identifying with academic domains in science, i.e., whether students see themselves as science people or people who can excel in science disciplines [2, 5, 266]. Students’ identity in STEM disciplines has been shown to play an important role in their in-class participation in related courses and choices of majors and careers [5, 49, 54-56, 267]. The science identity framework by Johnson et al. [55] includes three dimensions: competence (“I think I can”), performance (“I am able to do”), and recognition (“I am recognized by others”). Hazari et al. modified the framework specifically for physics. “Competence” and “performance” were defined as students’ beliefs in their ability to understand the subject and students’ belief in their ability to perform physics tasks. Additionally, recognition was framed as recognition by others as being a good physics student. Lastly, a fourth dimension, interest, was added to the framework by Hazari et al. since students have highly varying levels of interest in physics [60, 61]. Here we use a slightly reframed version of the physics identity framework by Kalender et al. [59]. In particular, performance and competence were combined into a single latent variable, self-efficacy (closely related to competency belief). Studies have shown that it can be more difficult for women to form a physics identity than men [1-3, 6]. However, most of the studies in the college context concerning physics identity and factors that influence it are conducted in classes in which women are underrepresented [4, 5]. According to Gee's identity framework [56], students' identity in a domain can be context dependent. Therefore, it is important to examine the physics identity of students in physics courses in which women outnumber men, e.g., in introductory physics courses for bioscience majors in which women are not underrepresented.

Students in these courses may be influenced by societal stereotypes and biases even before they enter the classroom. One common stereotype is the idea of genius and brilliance are important factors to succeed in physics [23]. However, genius is often associated with boys [169], and girls tend to shy away from fields associated with innate brilliance or genius from a young age [24]. Studies have found that by the age of six, girls are less likely than boys to believe they are "really really smart" and less likely to choose activities that are designed for "brilliant people" [24]. As these students get older, often the norms in the science curriculum tend not to represent the interests and values of girls [1, 309]. All of these stereotypes and factors can influence female students' perception of their ability to do physics before they enter the classroom. So, it is possible that although women are the majority in algebra-based physics courses, these societal stereotypes can still influence their outcomes in those classes if the learning environment is not equitable and inclusive.

Although other studies have investigated gender differences in motivational beliefs and correlations between students' attitudes and performance in introductory physics courses, most studies have not considered students' perception of the learning environment in influencing their outcomes particularly in courses in which women are not underrepresented. In our study, within the perception of the learning environment, we include students' self-reported sense of belonging, perceived recognition, e.g., by instructors and teaching assistants, and interaction with peers in the class. Students' sense of belonging in physics has been shown to correlate with their retention and self-efficacy [50, 63] and students' interaction with peers has been shown to enhance their understanding and engagement in courses [274]. A student's perceived recognition by others has been shown to play an important role in a student's identity [49] and is especially related to women's other motivation beliefs [50]. Studies have shown that female students do not feel recognized appropriately even before they enter college [1, 4, 24]. Our prior individual interviews suggest that students' perceived recognition by instructors and teaching assistants (TAs) impacts their self-efficacy and interest in physics (e.g., see [52, 53]). In addition, other learning environment factors, such as students' sense of belonging and interaction with their peers, can play a significant role in their learning and motivational outcomes.

This study uses Structural Equation Modeling (SEM) to examine the difference between male and female students on the role of the learning environment in influencing motivational and performance outcomes at the end of a mandatory two-semester algebra-based introductory physics course sequence for bioscience majors, controlling for their high school GPA, SAT math scores, grade, self-efficacy and interest in college physics 1. We controlled for students' past performance in high school and in the physics 1 course as well as their motivational beliefs when they started the course to try and isolate the effect of the learning environment on student outcomes. In our algebra-based introductory physics courses, 62% of the enrolled students are women. Even though women are not underrepresented in the algebra-based courses, it is important to investigate their motivational beliefs and performance outcomes since stereotypes about who can excel at physics could still have consequences in these courses.

Combining these important motivational beliefs and outcomes in our study, a visual of our model that focuses on the role of the learning environment is shown in Figure 28. All regression paths were considered from left to right in our model. However, only some of the paths are shown for clarity. High school factors, SAT math scores, and GPA, as well as outcomes from physics 1 (self-efficacy, interest, and grade), are controlled for at the beginning of the model. We then investigate how the perception of the learning environment (perceived recognition, peer interaction, and belonging) predicts student outcomes at the end of the two-semester physics course sequence (grade, self-efficacy, interest, and physics identity). Our research questions are as follows:

RQ1. Are there gender differences in students' grades and physics motivational beliefs at the end of the mandatory two semester course sequence?

RQ2. Is there moderation (interactions) by gender for any of the regression paths (predictive relationships)?

RQ3. If moderation does not affect any path, does gender mediate

  3.a. the factors that were controlled for?

  3.b. the perception of the physics learning environment controlling for the controlled factors?

  3.c. the student outcomes after controlling for everything else in the model?

RQ4. Which factors, including those in the physics learning environment, post interest, post self-efficacy, and grade at the end of the course, predict physics identity?

RQ5. How does the perception of the physics learning environment predict student outcomes in physics 2, controlling for factors at the beginning of the course?

RQ6. Which aspects of the perception of the learning environment explain a major portion of the variance in each outcome?

RQ7. What unique role did each of the factors in the learning environment play in mediating student outcomes?

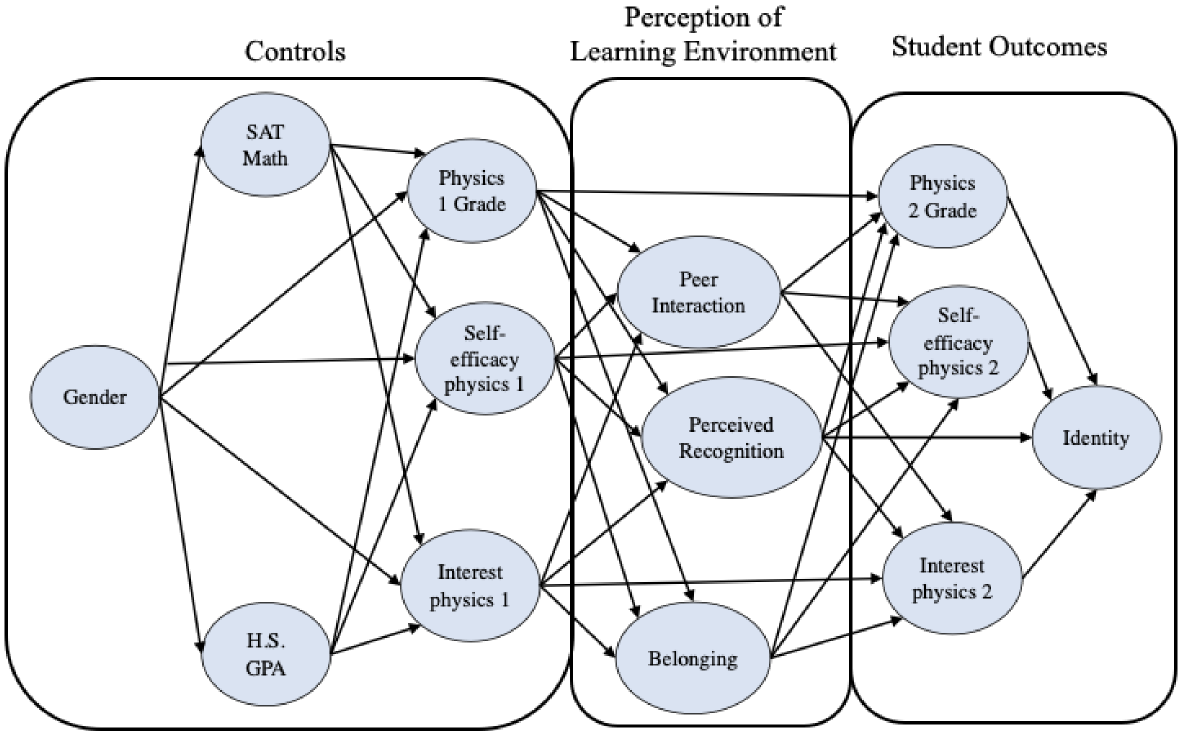


Figure 28 Schematic representation of the model based on the theoretical framework. From left to right, all possible paths were considered. Some, but not all, of the regression paths are shown (since others were not statistically significant).

13.2 Methodology

In this study, a validated survey covering different motivational constructs in our theoretical framework was administered to students at a large public research university in the U.S. The survey was given at the end of the semester in two semesters of traditionally taught lecture-based introductory algebra-based physics courses over the course of two years. The courses each had one recitation per week and were similar in the grading policy where students' grades mainly consisted of 2-3 midterms and 1 final exam. In general, there were little to no evidence-based active learning strategies implemented in the courses. This course is primarily taken by junior or senior bioscience majors for whom this two-semester physics course sequence is mandatory. We analyzed the results for 854 students who completed the survey in introductory physics 1 and physics 2 classes. The University provided demographic information such as age, gender, and ethnic/racial information using an honest broker process by which the research team received the information without knowledge of the identities of the participants. From the university data, the participants were 38% male and 62% female students. We recognize that gender is fluid and not a binary construct; however, the data collected by the institution is in binary terms, so we use that in this study. Less than 1% of the students did not choose male or female and thus was not included in the study.

13.2.1 Instrument Validity

The survey used in this study measured students' physics identity, self-efficacy, interest, sense of belonging, perceived recognition, and perception of interaction with peers for students enrolled in introductory algebra-based physics courses. The survey was adapted from previous research [170, 175] and revalidation of the motivational beliefs at our institution involved conducting one-on-one student interviews [26], Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and calculation of the Pearson Correlations. The survey questions for each construct and factor loadings for each question are given in Table 35.

Table 35 Survey questions for each of the motivational constructs along with factor loadings (Lambda) from the Confirmatory Factor Analysis (CFA) for all students (N = 854). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity, peer interaction, and perceived recognition questions was strongly disagree, disagree, agree, strongly agree. The

rating scale for the physics belonging questions was not at all true, a little true, somewhat true, mostly true, and completely true All *p*-values are < 0.001.

Construct and Item	Lambda
Physics Identity	
I see myself as a physics person.	1.000
Physics Self-Efficacy	
I am able to help my classmates with physics in the laboratory or recitation.	0.641
I understand concepts I have studied in physics.	0.710
If I study, I will do well on a physics test.	0.734
If I encounter a setback in a physics exam, I can overcome it.	0.664
Physics Interest	
I wonder about how physics works.[†]	0.696
In general, I find physics [††]	0.810
I want to know everything I can about physics.	0.759
I am curious about recent discoveries in physics.	0.705
Physics Perceived Recognition	
My family sees me as a physics person.	0.914
My friends see me as a physics person.	0.914
My physics instructor and/or TA sees me as a physics person.	0.678
Physics Belonging	
I feel like I belong in this class.	0.795
I feel like an outsider in this class.	0.676
I feel comfortable in this class.	0.845

I feel like I can be myself in this class.	0.688
Sometimes I worry that I do not belong in this class.	0.614
Physics Peer Interaction	
My experiences and interactions with other students in this class…	
Made me feel more relaxed about learning physics.	0.748
Increased my confidence in my ability to do physics.	0.946
Increased my confidence that I can succeed in physics.	0.942
Increased my confidence in my ability to handle difficult physics problems.	0.846

† The rating scale for this question was never, once a month, once a week, every day.

†† the rating scale for this question was very boring, boring, interesting, very interesting.

Table 36 Pearson inter-correlations are given between all the predictors and outcomes.

Pearson Correlation Coefficient

Observed Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. High School GPA	--	--	--	--	--	--	--	--	--	--	--	--
2. SAT Math	0.34	--	--	--	--	--	--	--	--	--	--	--
3. Post Self-Efficacy in Physics 1	0.11*	0.26	--	--	--	--	--	--	--	--	--	--
4. Post Interest in Physics 1	ns	ns	0.58	--	--	--	--	--	--	--	--	--
5. Physics 1 Grade	0.42	0.49	0.34	.093*	--	--	--	--	--	--	--	--
6. Perceived Recognition in Physics 2	ns	0.12**	0.40	0.52	0.19	--	--	--	--	--	--	--
7. Peer Interaction in Physics 2	ns	0.10*	0.36	0.28	0.13**	0.36	--	--	--	--	--	--
8. Belonging in Physics 2	0.15	0.18	0.53	0.38	0.26	0.45	0.62	--	--	--	--	--
9. Post Self-Efficacy in Physics 2	0.10*	0.22	0.72	0.46	0.22	0.58	0.67	0.79	--	--	--	--
10. Post Interest in Physics 2	0.10**	ns	0.45	0.89	ns	0.57	0.39	0.47	0.60	--	--	--
11. Physics 2 Grade	0.36	0.46	0.25	ns	0.68	0.22	0.21	0.35	0.32	0.10**	--	--
12. Physics Identity in Physics 2	ns	0.13	0.46	0.52	0.17	0.82	0.37	0.45	0.20	0.57	0.20	--

* = $p < 0.05$, ** = $p < 0.01$, no superscript * = $p < 0.001$, ns = non-significant

The pair-wise Pearson correlations are given in Table 36. Pearson's *r* values signify the strength of the pairwise correlations between constructs. The inter-correlations vary in strength, but none of the correlations are so high that the constructs cannot be separately examined. The only high inter-correlations were between post interest in physics 1 and post interest in physics 2 (0.89) and between physics identity in physics 2 and perceived recognition in physics 2 (0.82). Prior study shows that there is a high correlation in interest from the beginning to the end of the physics courses and from physics 1 and physics 2 in calculus-based introductory physics courses [26]. Additionally, perceived recognition is associated with external identity whereas physics identity asks about internal identity so there tends to be a high correlation between the constructs. Both correlations are low enough that they can be considered separate constructs.

The *physics identity* question evaluated whether the students see themselves as a physics person [230]. The *physics self-efficacy* questions measured students' confidence in their ability to understand physics and solve problems [170-172, 230]. The *interest in physics* questions measured students' enthusiasm and curiosity to learn physics and ideas related to physics [171]. The *sense of belonging* questions on the survey measured whether students felt like they belonged in the introductory physics classroom or not [50, 175]. The *perceived recognition* questions measured the extent to which students thought that other people see them as a physics person [230]. Lastly, the perception of *peer interaction* questions (which we will refer to as peer interaction for brevity) measured whether students thought that working with their peers is beneficial for their confidence to do physics [276, 277].

The questions in the study were designed on a Likert scale of 1 (low endorsement) to 4 (high endorsement) except for the sense of belonging questions which were designed on a scale of 1 to 5 [173]. A lower score was indicative of a negative endorsement of the survey construct while a higher score was related to a positive belief of the construct. Some of the questions were reverse coded (e.g., I feel like an outsider in this class).

13.2.2 Analysis

Initially, we analyzed the descriptive statistics and compared female and male students' mean scores on the predictors and outcomes for statistical significance using t-tests and to investigate the effect size using Cohen's d [182]. Cohen's d is $d = (\mu_m - \mu_f)/\sigma_{pooled}$, where μ_m is the average score of male students, μ_f is the average score of female students and σ_{pooled} is the pooled standard deviation (or weighted standard deviation for men and women) for all students. To quantify the statistical significance and relative strength of our frameworks' path links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. SEM is an extension of multiple regression and allows one to conduct several multiple regressions simultaneously between variables in one estimation model. This is an improvement over multiple regression since it allows us to calculate the overall goodness of fit and allows for all estimates to be standardized simultaneously so there can be direct comparisons between different structural components. SEM also has an option to handle missing data using the full estimation maximum likelihood "ML" estimation feature, which improves power since it imputes missing data so students only missing some data are not dropped. We report model fit for SEM by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for goodness of fit are as follows: CFI and TLI > 0.90, and SRMR and RMSEA < 0.08 [139].

The model estimates were performed using gender moderation analysis to check whether any of the relations between variables show differences across gender by using “lavaan” to conduct multi-group SEM [187]. In moderation analysis, the predictive relationship (regression path) between two variables is tested for two or more different groups (e.g., men and women) simultaneously. If the predictive relationship is different for the groups (i.e., the values of the regression coefficients are not the same for the correlation between the two constructs for different groups), then there is a moderation effect in the model. For example, in a study focusing on how smoking predicts lung cancer, if there was a moderation by gender, the predictive relation (regression coefficient) between smoking and lung cancer would be different for women and men. However, if the regression coefficients for how smoking predicts lung cancer were exactly the same for women and men, then there is no moderation by gender and one can just focus on mediation analysis by gender. 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 [187]. Since we did not find any statistically significant moderation by gender, we tested the theoretical model in a mediation analysis, using gender as a variable directly predicting items to examine the resulting structural paths between constructs (in the mediation analysis, if there are paths from gender to any of the constructs as we found in our results discussed in the next section, it implies that women and men did not have the same average value for those constructs even though the regression coefficients between the constructs are the same for women and men). We made various models with different learning environment factors to investigate the effects of the various factors on the student outcomes as shown in Figure 28.

13.3 Results and Discussion

13.3.1 Gender differences in predictors and outcomes

Women had statistically significantly lower mean values in the majority of the predictors and outcomes than men, except for High School GPA, which favors women (see Table 37). There was also no statistically significant difference in the mean of the physics 2 grade between men (3.14) and women (3.05). The gender differences in motivational outcomes in physics 2 despite women having no statistically significant grade difference is a cause for concern since student motivational outcomes can impact their future career trajectories. Therefore, it is important to improve the physics learning environments to make them equitable and inclusive so that these gaps in motivational outcomes can be eliminated.

Table 37 Mean predictor and outcome values by gender as well as statistical significance (*p*-values) and effect sizes (Cohen's *d*) by gender.

Predictors and Outcomes (Score Range)	Mean		*p*-value	Cohen's *d*
	Male	Female		
High School GPA (0-5)	4.05	4.14	0.003	-0.22
SAT Math (200-800)	682	660	<0.001	0.32
Post Self-efficacy in physics 1 (1-4)	2.98	2.73	<0.001	0.49
Post Interest in physics 1 (1-4)	2.81	2.38	<0.001	0.71
Physics 1 Grade (0-4)	3.39	3.23	0.003	0.40
Perceived Recognition in physics 2 (1-4)	2.24	1.98	<0.001	0.39
Peer Interaction in physics 2 (1-4)	2.94	2.79	0.001	0.24
Belonging in physics 2 (1-5)	3.69	3.45	<0.001	0.28
Post Self-efficacy in physics 2 (1-4)	2.94	2.73	<0.001	0.40
Post Interest in physics 2 (1-4)	2.77	2.32	<0.001	0.73
Physics 2 Grade (0-4)	3.14	3.05	0.104	0.12
Physics Identity in physics 2 (1-4)	2.19	1.85	<0.001	0.45

13.3.2 SEM Path Model

We used SEM to investigate the relationships between the constructs and to unpack each construct's contribution to explaining the self-efficacy, interest, and physics identity of women and men at the end of physics 2. We initially conducted gender moderation analysis between variables using multi-group SEM to investigate if any of the predictive relationships between the motivational constructs were different across gender. There were no group differences at the levels of weak and strong measurement invariance and at the level of regression coefficients. Thus, there is no moderation effect by gender. Therefore, we proceeded to gender mediation analysis. In particular, we proceeded to gender mediation analysis to understand the extent to which gender differences in students' outcomes at the end of the introductory physics sequence (self-efficacy, interest, grade, and physics identity) were mediated by differences in students' initial self-efficacy, interest, prior knowledge in physics, pre-college academic measures (High School GPA and SAT Math) and the perception of the physics learning environment.

13.3.2.1 Model 1: Perceived Recognition

In our first model, the only learning environment factor included was perceived recognition (we keep this factor because it has been previously found to be a predictor of physics identity in the theoretical framework we adapted). The result of the SEM path model is presented visually in Figure 29. The model fit indices indicate a good fit to the data: CFI = 0.931 (> 0.90), TLI = 0.919 (>0.90), RMSEA = 0.052 (<0.08), and SRMR = 0.044 (<0.08). Interestingly, we found that the only direct connections to gender involved the high school factors, SAT Math ($\beta = 0.15$) and H.S. GPA ($\beta = -0.11$), as well as self-efficacy in physics 1 ($\beta = 0.22$) and interest (0.37) at the end of physics 1. For example, gender directly mediates SAT Math and self-efficacy in physics 1 favoring men but the regression coefficient showing the predictive relation between SAT Math and self-efficacy in physics 1 is 0.24 for *both* women and men (if this regression coefficient was not the same for women and men, we would not have been able to do mediation analysis by gender since in that case, gender would have moderated the relation between these two constructs). We note that gender has indirect effects on physics 1 grade, perceived recognition, and the outcome variables (physics 2 grade, self-efficacy, interest, and identity).

Thus, in this model shown in Figure 29, self-efficacy, interest and perceived recognition in physics 2 had a direct effect on physics identity, and there were no direct effects from gender, grade in physics 2, or any of the pre-college and physics 1 variables. Students' perceived recognition in physics 2 had the largest direct effect (regression coefficient $\beta = 0.69$) while self-efficacy ($\beta = 0.13$) and interest ($\beta = 0.10$) had similar, smaller direct effects on the physics identity.

Self-efficacy in physics 2 is largely predicted by self-efficacy in physics 1 ($\beta = 0.60$) and perceived recognition ($\beta = 0.32$). Similarly, interest in physics 2, has direct effects from interest in physics 1 and perceived recognition with interest in physics 1 has the largest direct effect ($\beta = 0.82$), and perceived recognition has a smaller direct effect ($\beta = 0.14$). The students' grade in physics 2 has the largest direct effect from their grade in physics 1 ($\beta = 0.59$), and small direct effects from their SAT math scores ($\beta = 0.17$) and perceived recognition ($\beta = 0.08$). Perceived recognition plays an important role in predicting student motivational outcomes and therefore instructors must find different ways to positively recognize students in their classroom.

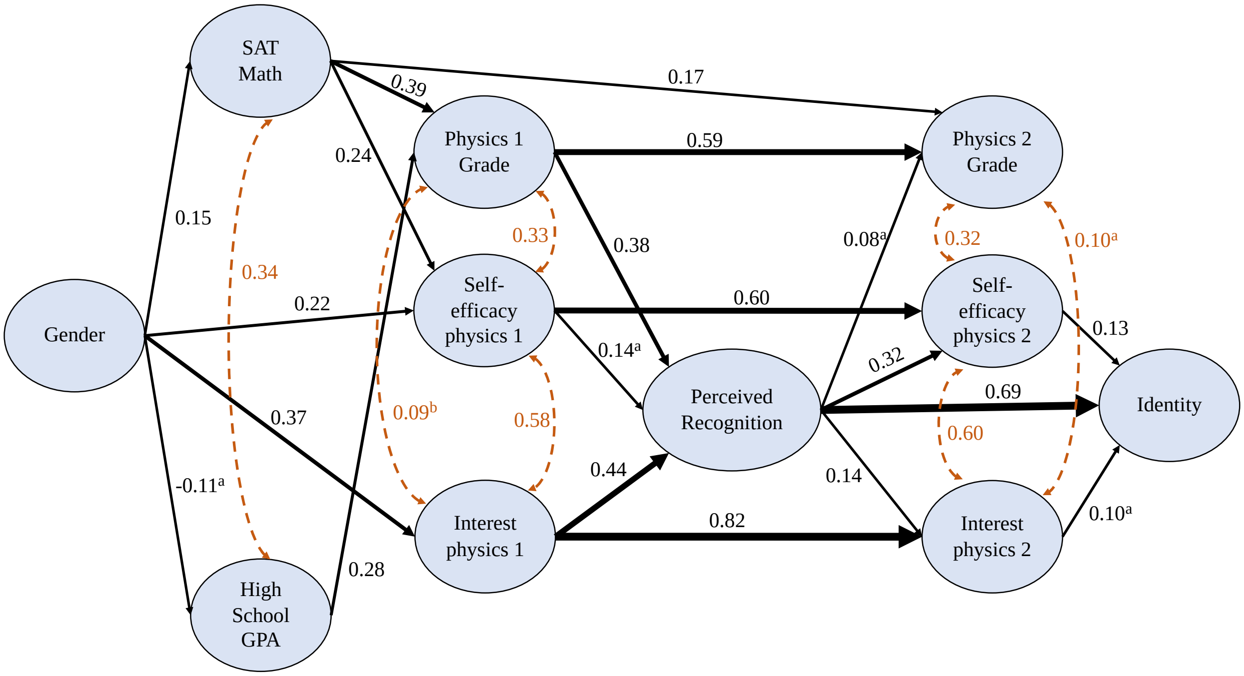


Figure 29 Result of the path analysis part of the SEM showing mediation between gender and outcomes in physics through various mediating factors for Model 1. Perceived recognition is the only motivational construct in the learning environment. The line thickness is the relative magnitude of the regression coefficients β. The dashed lines indicate covariances between constructs. All *p*-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p < 0.01$, and superscript "b" for $p < 0.05$.

13.3.2.2 Model 2: Perceived Recognition and Peer Interaction

In our second model, we added perception of peer interaction into the learning environment of physics 2. The result of the SEM is presented visually in Figure 30. The model fit indices indicate an even better fit to the data than the first model: CFI = 0.943 (>0.90), TLI = 0.934 (>0.90), RMSEA = 0.048 (<0.08), and SRMR = 0.045 (< 0.08). There are minimal differences in the direct effect of gender to the other factors in the model. Additionally, identity has the same direct paths and β values as in Model 1. However, there are differences in the outcomes of grade, self-efficacy, and interest in physics 2. The direct path from perceived recognition to grade is not present in this model and instead is replaced with a direct path from peer interaction (β = 0.11). The direct paths from perceived recognition and interest in physics 1 to interest in physics 2 are slightly smaller compared to Model 1and there is a new direct path from peer interaction (β = 0.16). The largest change is in the direct paths to self-efficacy in physics 2. The direct paths from self-efficacy in physics 1 and perceived recognition decrease by approximately a β value of 0.10 (from 0.60 to 0.50 and from 0.32 to 0.21, respectively). Additionally, a direct path from peer interaction to self-efficacy in physics 2 is significant in this model (β = 0.40). Therefore, students' perception of their peer interaction plays a significant role in predicting their learning and motivational outcomes. Thus, it may be beneficial for instructors to influence peer interaction by introducing more active learning pedagogy in their classroom in an equitable and inclusive learning environment (e.g., ensuring that men do not dominate and ignore or marginalize women during peer interaction).

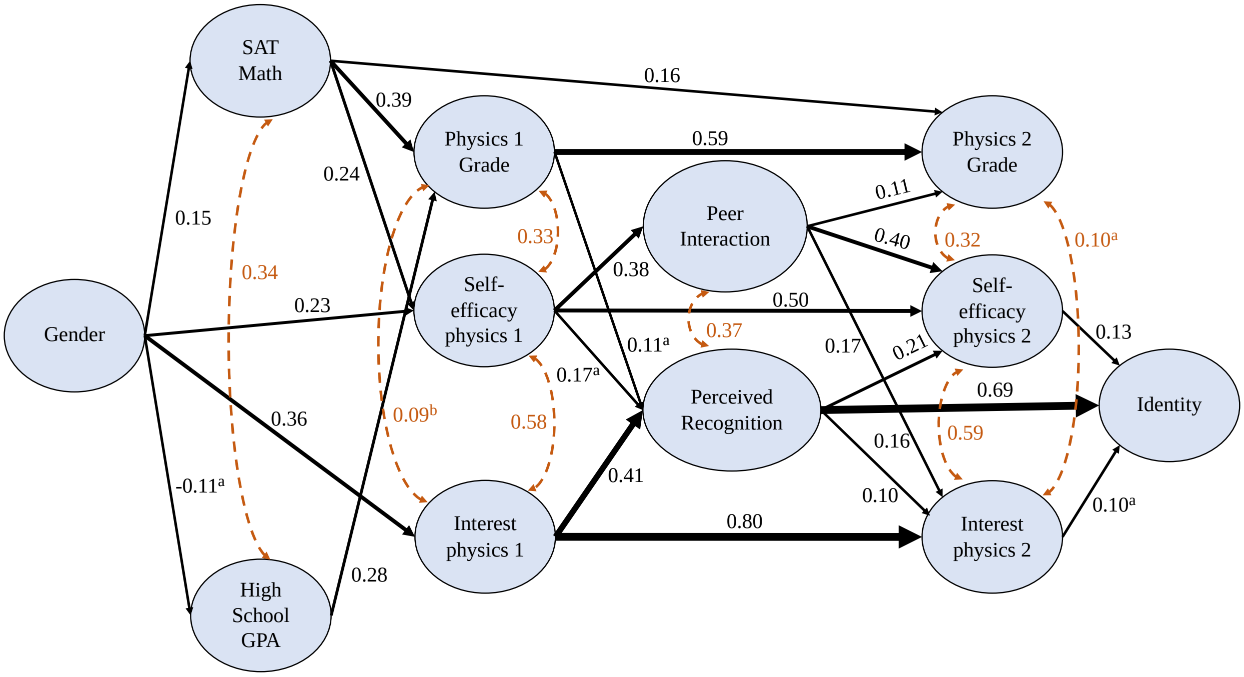


Figure 30 Result of the path analysis of the SEM between gender and outcomes in physics through various mediating factors for Model 2. Perceived recognition and peer interaction are the two factors included in the learning environment. The line thickness is the relative magnitude of β values. The dashed lines indicate

covariances between constructs. All *p*-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p < 0.01$, and superscript "b" for $p < 0.05$.

13.3.2.3 Model 3: Perceived Recognition and Belonging

In our third model, the learning environment consisted of perceived recognition and belonging in physics 2. The result of the SEM is presented visually in Figure 31. The model fit indices indicate a comparable fit to the data as the first model: CFI = 0.931 (> 0.90), TLI = 0.922 (0.90), RMSEA = 0.049 (<0.08), and SRMR = 0.045 (<0.08). There are minimal differences in the direct effect of gender to various factors. The same direct paths and β values to identity from the first two models are present. However, there are some differences in the outcomes of grade, self-efficacy, and interest in physics 2. The only direct path to physics 2 grade from the learning environment factors is from belonging (β = 0.17). Similar to the model with peer interaction and perceived recognition, the path between perceived recognition and self-efficacy was reduced and a direct path from belonging to self-efficacy is present in this model (β = 0.50). Since belonging is important for students' outcomes in the course, it is crucial that instructors increase students' sense of belonging. For example, instructors could ensure that students' ideas are respected by others in the class and students recognize that everyone struggles while learning physics and struggling is the stepping stone to learning physics.

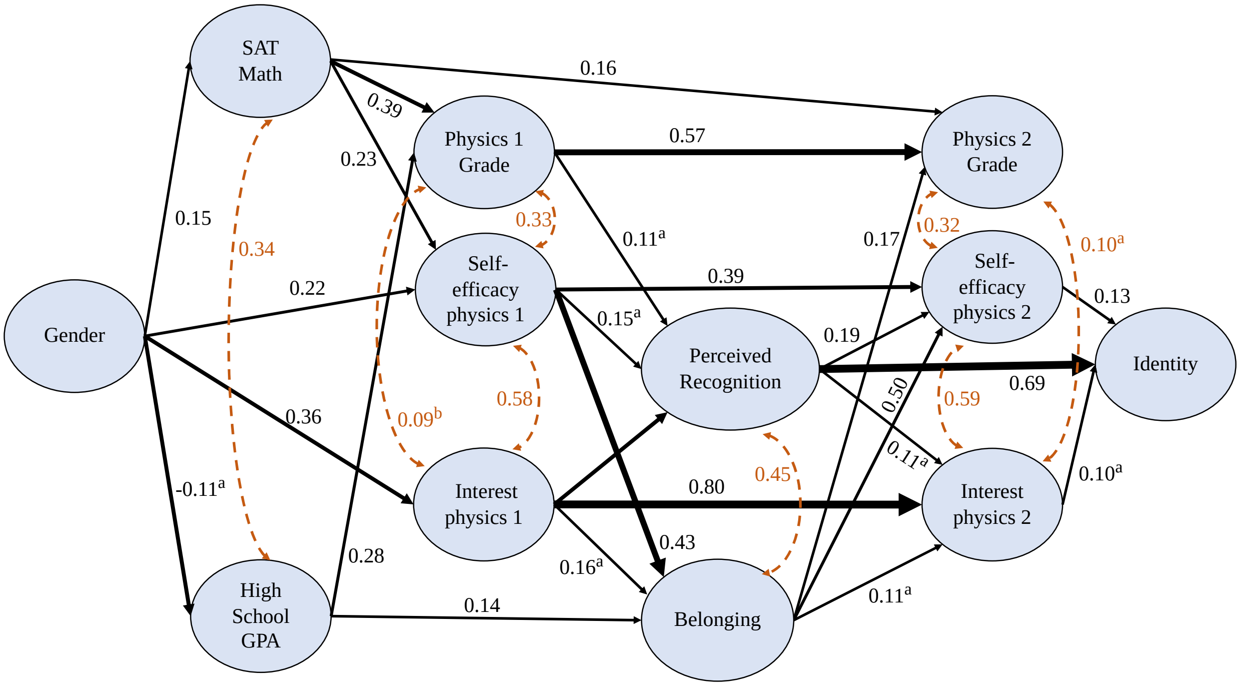


Figure 31 Result of the path analysis part of SEM between gender and outcomes in physics through various mediating factors for Model 3. Perceived recognition and belonging are factors in the learning environment. The line thickness is the relative magnitude of β values. The dashed lines indicate covariances between constructs. All p-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p < 0.01$, and superscript "b" for $p < 0.05$.

13.3.2.4 Model 4: Perceived Recognition, Peer Interaction, and Belonging

Our fourth and final model included all three perceptions of the physics learning environment factors; belonging, peer interaction, and perceived recognition in physics 2. The result of the SEM is presented visually in Figure 32. The model fit indices indicate a good fit to the data: CFI = 0.936 (> 0.90), TLI = 0.927 (> 0.90), RMSEA = 0.047 (< 0.08), and SRMR = 0.046 (< 0.08). One difference in this model from the second model is that the direct path from peer interaction to physics 2 grade does not appear and instead there is a direct path from belonging to physics 2 grade ($\beta = 0.17$). Self-efficacy has a direct path from belonging ($\beta = 0.34$). The direct paths to self-efficacy from peer interaction and perceived recognition are reduced (from 0.40 to 0.26 and from 0.21 to 0.17, respectively). The direct paths to identity from perceived recognition, self-efficacy, and interest in physics 2 remain unchanged in all four models. Thus, all three learning environment factors (perceived recognition, peer interaction, and belonging) predict student outcomes in the physics 2 course.

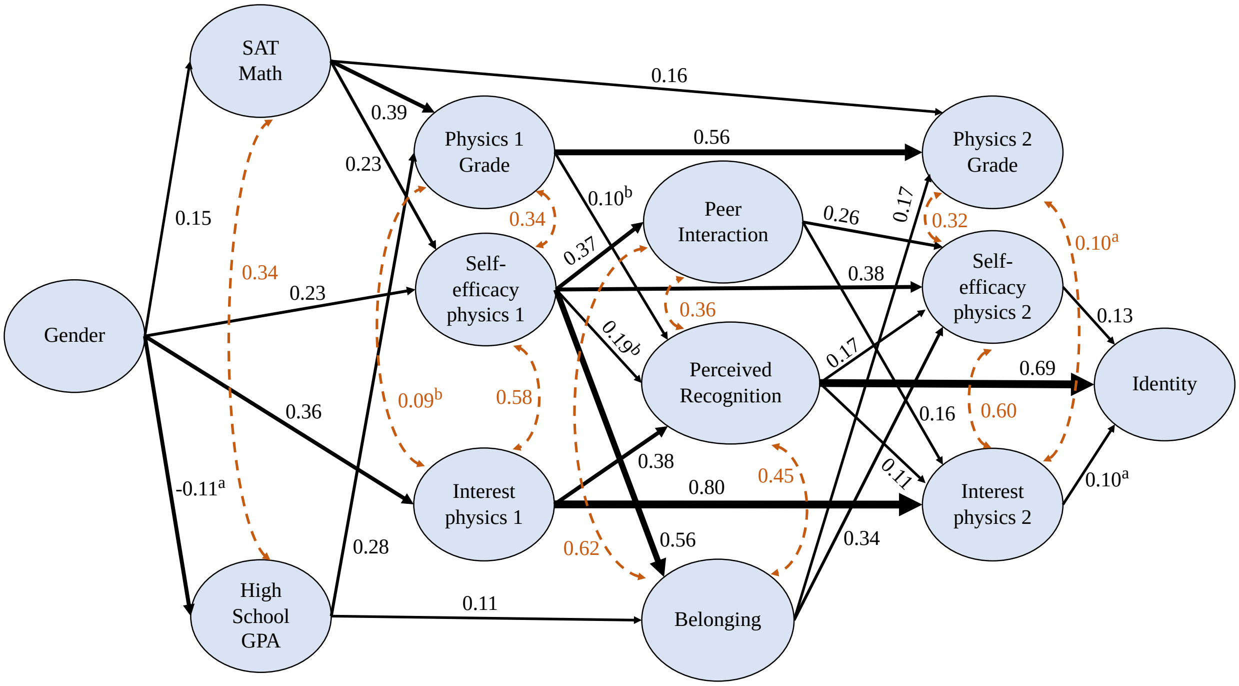


Figure 32 Result of the path analysis part of the SEM between gender and outcomes in physics through various mediating factors for Model 4. Perceived recognition, peer interaction, and belonging are the motivational constructs included in the learning environment. The line thickness indicates the relative magnitude of β values. The dashed lines indicate covariances between constructs. All *p*-values are indicated by no superscript for $p < 0.001$, superscript "a" for $p < 0.01$, and superscript "b" for $p < 0.05$

13.3.3 Direct and Indirect Paths

Since not all of the factors directly relate to student outcomes, we found the direct and indirect paths of all the control variables and learning environment factors to the student outcomes in Model 4. An indirect path is measured by finding the product of each path (β) from a given variable to an outcome variable. The total indirect path for a given variable and outcome was found by adding all of the indirect paths together. For example, there are two indirect paths from peer interaction to identity. The first path goes from peer interaction → S.E. physics 2 → identity (0.26*0.13 = .03). The second path goes from peer interaction → interest physics 2 → identity (0.16*0.10 = .02). So, the total indirect path from peer interaction to identity is 0.03 + 0.02 = 0.05. The total path was calculated by adding the direct path and total indirect path together. A summary of all direct and indirect paths can be found in Table 38.

Table 38 Regression coefficients for direct paths and indirect paths from the controlled variables and learning environment factors to every student outcome in Model 4.

Outcome	Predictor	Direct Path	Total Indirect Path	Total Path
Grade physics 2	SAT Math	0.16	0.24	0.40
	High School GPA	0.00	0.18	0.18
	Grade physics 1	0.56	0.00	0.56
	Self-efficacy physics 1	0.00	0.10	0.10
	Interest physics 1	0.00	0.00	0.00
	Peer Interaction	0.00	0.00	0.00
	Perceived Recognition	0.00	0.00	0.00
	Belonging	0.17	0.00	0.17
Self-efficacy physics 2	SAT Math	0.00	0.17	0.17
	High School GPA	0.00	0.04	0.04
	Grade physics 1	0.00	0.02	0.02
	Self-efficacy physics 1	0.38	0.32	0.70
	Interest physics 1	0.00	0.06	0.06
	Peer Interaction	0.26	0.00	0.26
	Perceived Recognition	0.17	0.00	0.17
	Belonging	0.34	0.00	0.34
Interest physics 2	SAT Math	0.00	0.02	0.02
	High School GPA	0.00	0.00	0.00
	Grade physics 1	0.00	0.01	0.01
	Self-efficacy physics 1	0.00	0.08	0.08
	Interest physics 1	0.80	0.04	0.84

	Peer Interaction	0.16	0.00	0.16
	Perceived Recognition	0.11	0.00	0.11
	Belonging	0.00	0.00	0.00
Identity	SAT Math	0.00	0.07	0.07
	High School GPA	0.00	0.03	0.03
	Grade physics 1	0.00	0.07	0.07
	Self-efficacy physics 1	0.00	0.23	0.23
	Interest physics 1	0.00	0.35	0.35
	Peer Interaction	0.00	0.05	0.05
	Perceived Recognition	0.69	0.03	0.72
	Belonging	0.00	0.04	0.04

13.3.4 Variance explained by each model

After constructing the models with each additional learning environment factor, we calculated the coefficient of determination (Adjusted R^2) of each construct in every model to investigate the proportion of variance explained by each model (Table 39). We found that each model explained the same amount of variance for physics interest, grade, and identity in physics 2. This means that adding in peer interaction and belonging did not explain any more variance in those three outcome variables. However, the variance in self-efficacy for physics 2 increased for each consecutive model. Model 1 explained 61% of the variance in self-efficacy, model 2 explained 75% of the variance, model 3 explained 78%, and model 4 explained 81% of the variance. Therefore, all three of the learning environment factors play a crucial role in predicting student outcomes (and particularly self-efficacy) in the course. Thus, it is important that instructors consider all three learning environment factors to create a more equitable and inclusive environment so all students can succeed.

Table 39 Adjusted coefficients of determination (Adjusted R^2) for all variables in the four different models on the impact of the learning environment. P.R. is perceived recognition, Peer is peer interaction, and Bel is belonging.

	Adjusted R^2			
Variable	Model 1 P.R.	Model 2 P.R. + Peer	Model 3 P.R. + Bel	Model 4 P.R. + Peer + Bel
High School GPA	0.00	0.00	0.00	0.00
SAT Math	0.01	0.01	0.01	0.01
Post Self-Efficacy in physics 1	0.11	0.11	0.11	0.11
Post Interest in physics 1	0.12	0.12	0.12	0.12
Physics 1 Grade	0.30	0.30	0.30	0.30
Perceived Recognition in Physics 2	0.29	0.29	0.29	0.28
Peer Interaction in physics 2	--	0.13	--	0.13
Belonging in physics 2	--	--	0.30	0.32
Post Self-Efficacy in physics 2	0.61	0.75	0.78	0.81
Post Interest in physics 2	0.82	0.82	0.82	0.82
Physics 2 Grade	0.49	0.49	0.50	0.50
Physics Identity in physics 2	0.69	0.69	0.69	0.69

*all *p*-values are < 0.001

13.4 Summary and Implications

Women taking traditionally taught algebra-based physics courses have lower motivational beliefs than men at the end of physics 2, a pattern similar to what we observed in the calculus-based physics courses [4, 59], despite the higher representation of women in the former—62% of students versus 30%, respectively. In response to **RQ1,** we find that at the end of physics 2, women had lower self-efficacy, interest, perceived recognition, peer interaction, belonging, and physics identity (Table 37). Women and men did not have statistically significant differences in course grade at the end of physics 2, but women had statistically significantly lower grades than men in physics 1 (Table 37). One hypothesis for gender difference in grades in physics 1 (but not in physics 2) is that the high school physics course that many students take overlaps with the content of physics 1 (mechanics) and male students might benefit more from the way those high school courses are taught and their greater familiarity with this content may provide them an advantage in college physics 1 particularly if college instructors do not make an effort to create an equitable and inclusive learning environment and level the playing field. However, what is also a cause for concern is that women had lower motivational beliefs in both courses than men. Also, women had lower self-efficacy coming into the physics 2 course, which we hypothesize to be due to societal stereotypes about who can excel in physics and previous gendered experiences women had before they came into the class. The learning environment in physics 2 did not alleviate these gender differences. Since students' self-efficacy has important implications for student performance in the course and can also impact their career trajectories, it is important to make physics learning environments equitable and inclusive so that the gap can be eliminated. This entails creating a learning environment in which women feel recognized by their instructors and have a high sense of belonging in the course since these beliefs are important for excelling in the course and play a key role in predicting physics identity, interest, and self-efficacy.

In response to **RQ2,** we did not find moderation by gender for any regression path in our models. Therefore, SEM mediation models were implemented to help answer **RQ3**. We find that gender mediates certain factors we controlled for, such as the high school GPA, SAT math scores, interest, and self-efficacy at the end of physics 1. However, gender did not directly predict any learning environment or student outcome factors. This was true for all of our models and further supports the finding that women begin and end physics 2 with lower motivational beliefs than men. This is potentially due to societal stereotypes and biases women experience about who belongs in physics and can excel in physics throughout their life and persist if instructors do not create an equitable and inclusive learning environment. Thus, instructors must be provided incentives and support to make physics learning environments equitable and inclusive in order to eliminate these differences.

To answer question **RQ4**, in all four models perceived recognition, interest, and self-efficacy at the end of physics 2 directly influenced identity. Perceived recognition had the most influence on identity ($\beta = 0.69$) followed by self-efficacy ($\beta = 0.13$) and interest ($\beta = 0.10$). This is consistent with past studies [59]. Indirect effects on identity include belonging and peer interaction, as seen in Table 38. However, there was no direct path from students' grades to their physics identity so that how well students did in the course did not predict whether they felt like a physics person. Since students' physics identity may have important implications for their future career choices, it is important to improve all students' physics identity in these introductory physics courses.

Turning to the learning environment factors, we focus on responses to RQ5 and RQ6. The factors that predicted self-efficacy in both physics 1 and 2 were perceived recognition (β = 0.17), peer interaction (β = 0.26), and belonging (β = 0.34). Interest in physics 2 was predicted by the learning environment factors of perceived recognition (β = 0.11) and peer interaction (β = 0.16) as well as interest in physics 1 (β = 0.80). Unlike self-efficacy, which was predicted by all learning environment factors, interest at the end of the physics 2 course was primarily predicted by interest at the end of physics 1. Despite this finding in traditionally taught courses, interest in physics has the potential to show desired change at the end of the course if instructors explicitly focus on this goal. One possible way to influence students' interest is to engage them with problems that relate to their prospective majors or occupations or that are of interest to them in general. The only learning environment factor that predicted physics 2 grade in the fourth and final model was belonging (β = 0.17). In addition, physics 1 grade and SAT math also affected this outcome. While the only direct path to grade in the final model was students' sense of belonging, perceived recognition and peer interaction may still predict physics 2 grade through some of the shared variances in these factors. As we observed in models 1 and 2, peer interaction and perceived recognition had direct paths to physics 2 grade when belonging was not included as a learning environment factor. From these results, it is clear that the learning environment is important when predicting student outcomes. Therefore, physics instructors should embrace their role in improving student outcomes in their courses and create an equitable and inclusive learning environment in which all students thrive.

In response to **RQ7**, Table 39 shows the unique role each of the learning environment factors plays in mediating student outcomes. The variance explained in interest, grade, and identity in physics 2 do not change significantly when more learning environment factors are added. For interest, this is because most of the variance in all models is explained by interest in physics 1 instead of the learning environment factors. For grade in physics 2, this is because most of the learning environment influence comes from the shared variance between the learning environment factors. When only perceived recognition is included in the model, it directly influences physics 2 grade; when perceived recognition and peer interaction are included in the model, peer interaction directly influences physics 2 grade; and in the final model, only belonging influences physics 2 grade. It is important to note that only about 50% of the variance is explained for the physics 2 grade, so there are still other factors that predict grade that are not included in the model. Perceived recognition plays a large role in predicting identity and was included in each of our models. Unlike other outcomes, self-efficacy had more variance explained when more learning environment factors were added. In particular, peer interaction played an important role in predicting self-efficacy and it explained 14% more of the variance when included in the model. When belonging is also added to the model, it explains an additional 6% of the variance. Thus, it is important to include all three learning environment factors in the model to help explain the student outcomes. While each factor plays a unique role in explaining student outcomes, sense of belonging, peer interaction, and perceived recognition can influence each other to create better outcomes in equitable and inclusive learning environments [310]. Not only can the learning environment factors influence each other, but as seen from Table 37, self-efficacy and grade, self-efficacy and interest, as well as self-efficacy and perceived recognition are all correlated with each other.

Instructors could potentially change student outcomes for the better if they can provide support for factors they can control (i.e., learning environment factors). It is important to note that the physics 1 and physics 2 courses in this study were traditionally taught lecture-based physics courses where student grades heavily depended on two or three midterms and a final exam. There were little to no research-based active engagement strategies used in the classroom. Instructors can influence students' peer interaction with each other by providing time for students to work in groups during class or recitations and making sure that all student voices are heard equally while discussing problems. This could in turn help increase students' sense of belonging in the classroom if they observe that their ideas are respected by other students in the classroom. In addition, by not letting men dominate the conversations in class and explicitly praising effort and affirming women when they do well or make progress in the class, the gender gap in students' perceived recognition could potentially be decreased. This is especially important when implementing active learning pedagogy in the classroom. If not implemented using teaching strategies that are equitable and inclusive, men have been shown to not only dominate responding to questions in class but also while working in groups which can lower women's self-efficacy [219]. Moreover, instructors should be careful not to say that problems are "trivial", "easy" or "obvious" when students ask them for help after trying their best because otherwise female students are more likely to feel disparaged (or negatively recognized). What instructors need to realize is that what is important is the impact they are having on the students and not their intentions. Brief social-psychological classroom interventions, or mindset or sense of belonging interventions, have also been shown to eliminate or reduce the gender gap in performance [151, 152, 185, 186]. It would also help if instructors make it clear to students that they have high expectations of them, but they know that they have what it takes to meet those expectations and excelling in physics entails working hard and working smart and using effective approaches to learning and getting help from instructors and peers frequently as needed.

In summary, instructors and TAs need to create a learning environment that emphasizes recognizing their students, allowing for positive peer interactions, and providing a space where all students can feel like they belong in physics. From our analysis, these factors play important roles in predicting students’ grade, self-efficacy, interest, and identity in physics. We hypothesize that many of these differences are due to the stereotypes and past experiences women have that perpetuate through to the end of these introductory physics courses even though women are not underrepresented in them. It is important to note that the learning environment includes not only what happens in the classroom. Student interactions with each other while they are doing homework, students’ experiences in an instructor or TA’s office hours, interactions between students and the instructor over email, and other circumstances all contribute to the students’ learning environment. All of those interactions can affect students' identity, self-efficacy, and interest in physics by the end of the semester.

14.0 Future Directions

The research presented here gives insight into gender differences in physics motivational beliefs in the introductory physics courses and how the inclusiveness of the physics learning environment predicts students' motivational beliefs and performance for students on the bioscience track. These findings can be very useful for instructors interested in making these courses more equitable and inclusive. However, more research is needed.

The research presented here focuses on gender differences in these courses. The data obtained from the university only included the binary options of male and female; however, gender is a more complex, multidimensional construct. Therefore, in future research, it would be important to provide students with multiple options to investigate a more nuanced gender construct. Moreover, qualitative research would also be useful to gain further insights into the quantitative findings discussed here.

In addition, there is a need to investigate and address the impact of other facets of students' identity that can lead to inequitable outcomes in these physics courses. For example, other factors such as students' race/ethnicity, income status, first-generation college status, disability, and sexual orientation may be important to consider in future analyses.

This research was carried out at a large research institution and may not be generalizable across institutions of different types. Thus, studies that investigate student motivational beliefs for institutions with different student populations should be conducted such as students at two-year colleges, all-women's colleges, and racially diverse colleges, e.g., minority serving institutions (MSI).

While there is more research needed, investigating students' motivational beliefs can provide insight into the reasons for inequitable outcomes in physics courses in which women are not numerical minorities. This research and future research on students' motivational beliefs can be useful to improve equity and inclusion in introductory physics courses in which women are not numerical minorities.

Appendix A Full SEM Models

In the main text, we included four SEM models, here we include more detail about each model. We have included data of results from a motivational survey given to introductory physics 2 course for bioscience majors. We used the physics identity framework (as explained in section IV) as an example to show how the proposed framework is used to select a good model from several statistically equivalent ones. The models predicted students' physics identity through self-efficacy, interest, and perceived recognition. According to the Lee–Hershberger replacing rules [136], the best known rules in the SEM literature for generating equivalent structural models, there are 27 statistically equivalent models with different predictive relations between the three mediating constructs (self-efficacy, interest, and perceived recognition). Here we discuss four of the models to show how our framework could guide us to select a good model. While each model is statistically equivalent, the instructional implications of each model are different.

Initially, we compared female and male students' mean scores of the predictors and outcomes for statistical significance using *t*-tests and for the effect size using Cohen's *d* [182]. Table 40 shows that were statistically significant differences in all of the predictors and outcomes in favor of male students despite the fact that women are the majority in the algebra-based courses (60%).

Table 40 Mean pre predictor and outcome values in physics 1 by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender. All p-values < 0.001.

Predictors and Outcomes (Score Range)	Mean		Cohen's *d*
	Male	Female	
Perceived Recognition (1-4)	2.24	1.98	0.39
Self-Efficacy (1-4)	2.94	2.73	0.40
Interest (1-4)	2.77	2.32	0.73
Physics Identity (1-4)	2.19	1.85	0.45

To quantify the significance and relative strength of our frameworks links, we used Structural Equation Modeling (SEM) as a statistical tool by using R (lavaan package) with a maximum likelihood estimation method [138]. The level of SEM model fit can be represented by using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residuals (SRMR). Commonly used thresholds for the goodness of fit are as follows: CFI and TLI > 0.90, and SRMR and RMSEA < 0.08 [139]. We initially tested gender moderation between different constructs using multi-group SEM (between male and female students) to see if the relationship between the motivational variables were different across gender. There were no group differences at the level of weak and strong measurement invariance including at the level of regression coefficients. Therefore, we proceeded to gender mediation analysis to understand how gender mediates physics identity at the end of the introductory physics 2 course for bioscience majors. 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.

First, we consider a model in which there is no predictive relationship among self-efficacy, interest, and perceived recognition. Instead, there are covariances between each motivational factor. Figure 33 shows the path analysis results of this SEM model. The model fit indices indicate a good fit to the data: CFI = 0.966, TLI = 0.954, RMSEA = 0.061, SRMR = 0.038. As shown in Figure 33, 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 40 showing that women have significantly lower scores than men in all of these three motivational constructs. However, the direct effect of gender on physics identity is statistically insignificant even though women's identity is significantly lower than that of men, as shown in Table 40. This result indicates that self-efficacy, interest, and perceived recognition mediate the effect of gender on physics identity. In addition, perceived recognition has the largest direct effect on identity (β = 0.70) while self-efficacy and interest play smaller roles (β = 0.12 and β = 0.10 respectively).

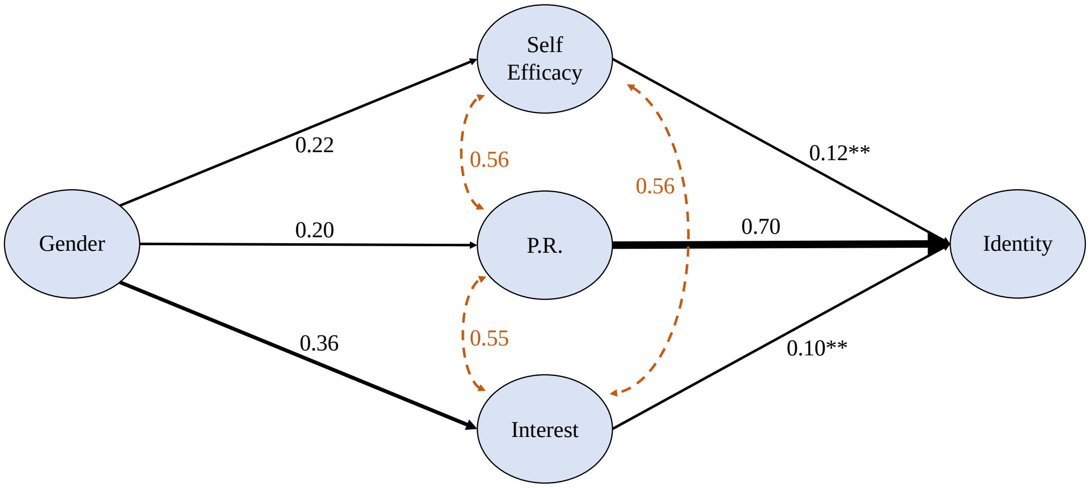
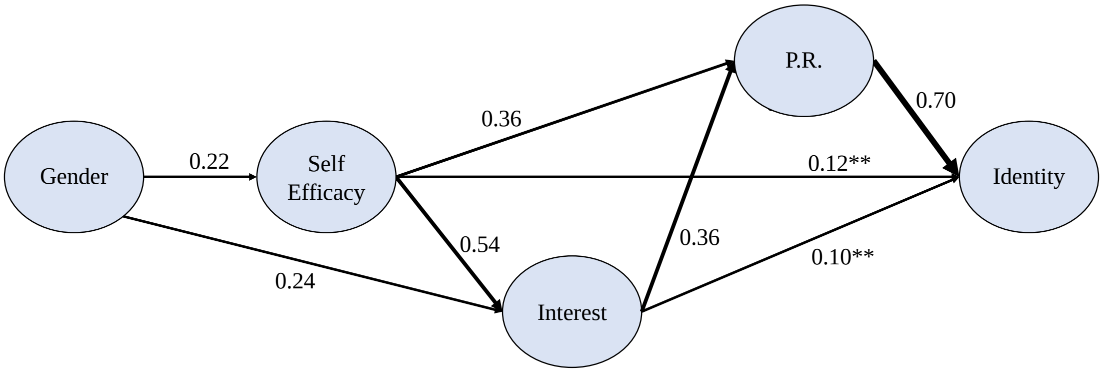


Figure 33 Results of the path analysis part of SEM Model 1, in which there are only covariances between each pair of constructs: self-efficacy, perceived recognition (R.R.), and interest. The dashed lines represent residual covariances between constructs. The solid lines represent regression paths, and the numbers on the lines are standardized regression coefficients (β values), which represent the strength of the regression

relations. Each regression line thickness qualitatively corresponds to the magnitude of the β value. Regression coefficients with $p < 0.010$ are indicated by superscript ** and with $p \leq 0.001$ are indicated by no superscript. 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 34). 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. Each of the model indices indicate a good fit to the data with Model 2 indices: CFI = 0.966, TLI = 0.955, RMSEA = 0.060, SRMR = 0.038 and Model 3 indices: CFI = 0.966, TLI = 0.956, RMSEA = 0.059, SRMR = 0.038. From Figure 34, gender directly predicts self-efficacy and interest in Model 2 while gender directly predicts interest in Model 3. In addition, in both models, physics identity is predicted by self-efficacy, perceived recognition, and interest with similar β values to Model 1.

(a) Model 2



(b) Model 3

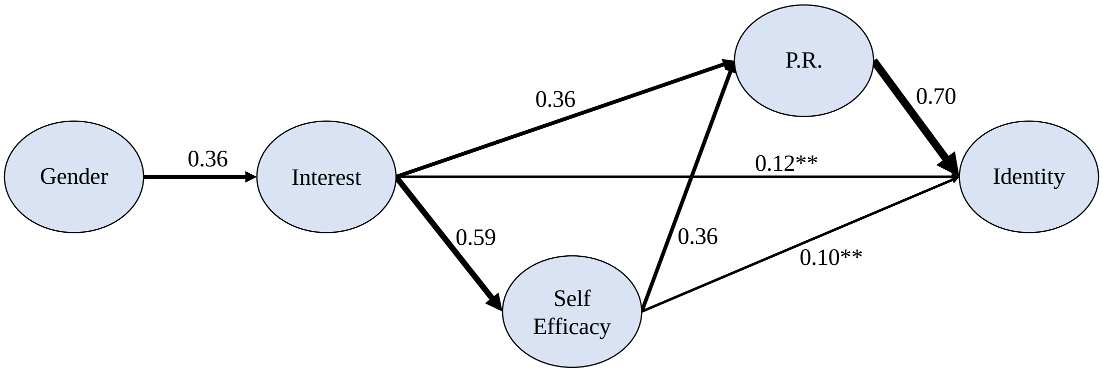


Figure 34 Results of the path analysis part of the SEM models that show how the relationship between gender and physics identity is mediated through self-efficacy, interest, and perceived recognition (PR). (a) In Model 2, self-efficacy predicts interest and perceived recognition, and interest predicts perceived recognition. (b) In Model 3, interest predicts self-efficacy and perceived recognition, and self-efficacy predicts perceived recognition. Regression coefficients with $p < 0.010$ are indicated by superscript ** and with $p \leq 0.001$ are indicated by no superscript. For clarity, we have removed the statistically insignificant regression paths

Finally, we consider a model in which perceived recognition predicts self-efficacy and interest, and self-efficacy predicts interest (see Figure 35). The model fit indices indicate a good fit to the data: CFI = 0.966, TLI = 0.954, RMSEA = 0.061, SRMR = 0.038. In Model 4, gender directly predicts perceived recognition, self-efficacy, and interest. Similar to Models 1-3, there is no statistically significant direct regression line from gender to physics identity in Model 4. In addition, self-efficacy, perceived recognition, and identity directly predict physics identity with similar β values to Models 1-3.

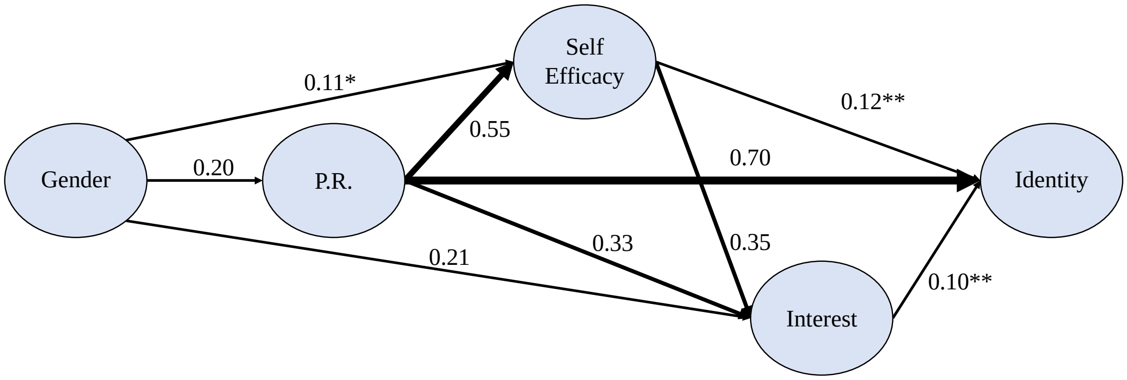


Figure 35 Results of the path analysis part of SEM Model 4, in which perceived recognition predicts interest and self-efficacy, and self-efficacy predicts interest. Each regression line thickness qualitatively corresponds to the magnitude of β values. Regression coefficients with $p < 0.010$ are indicated by superscript ** and with $p \leq 0.001$ are indicated by no superscript. For clarity, we have removed the statistically insignificant regression path from gender to identity.

To compare the equivalence of different models, we compared the Akaike information criterion (AIC) and Bayesian information criterion (BIC) of each model. They are used to test non-nested models. A lower AIC and BIC means that the model is considered to more likely be the true model. However, the AIC and BIC can also be affected by the number of regression lines. Model 1 has an AIC = 17988 and BIC = 18197; Model 2 has an AIC = 17986 and a BIC = 18190; Model 3 has an AIC = 17984 and BIC = 18183; Model 4 has AIC = AIC = 17988 and BIC = 18197. The AIC had a maximum difference of 4, whereas the BIC had a maximum difference of 14. For this case, these differences can be considered negligible since the differences in AIC and BIC come from cutting nonsignificant paths from gender as the models changed.

In addition, we also calculated the coefficient of determination R^2 of physics identity for each statistically equivalent SEM model, which represents the proportion of the variance in physics identity explained by each model. The results show that R^2 of identity is 0.69 in all of the models, which means that each equivalent model explains 69% of the variance in physics identity, i.e., these models explain students' physics identity equally well. Thus, these statistically equivalent SEM models are all robust from a statistical point of view.

Since all 4 models we investigated are statistically equivalent, according to the theoretical framework, we now must consider the instructional implications of each model. In Model 1, self-efficacy and interest are only predicted by gender. However, self-efficacy and interest may be considered fixed by others. Model 1 does not provide suggestions that instructors can use to improve those motivational factors. In addition in Models 2 and 3, gender only predicts self-efficacy or interest respectively which could be interpreted as a deficit model. In particular, these equivalent models can be interpreted to imply, that women are not feeling positively recognized by their instructors and teaching assistants (TAs) as much as men because they have lower interest and self-efficacy than men. While statistically equivalent to Models 1-3, Model 4 with perceived recognition predicting self-efficacy and interest is more likely to give them the message that students' interest and self-efficacy in physics can be influenced by the recognition they receive from instructors. Thus, Model 4 is also more likely to inspire instructors to create a more inclusive and equitable learning environment in which all students including those from the underrepresented groups feel more positively recognized and affirmed.

Next, we must assess whether the instructionally beneficial models are also supported by additional evidence. Model 4 also reflects findings from prior interviews. The interviews show that recognition by others and instructors or TAs is critical in shaping students' self-efficacy and interest [52, 53, 140, 141]. Thus, 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 focusing on the model's instructional implications and supporting evidence from individual interviews with students.

Appendix B Percentages of students who answered each self-efficacy question

In the main text, we investigated how students' self-efficacy changes from physics 1 to physics 2 by comparing their average scores on the self-efficacy constructs in the two courses. Below, we provide the percentages of women (Table 41) and men (Table 42) who selected each answer choice from a 4-point Likert scale for each self-efficacy question in physics 1 and physics 2.

Table 41 Percentages of women who answered each self-efficacy question by the options they selected with 1 being the low value (NO!) and 4 being the high values (YES!). The rating scale for the self-efficacy questions was: NO! no yes YES!.

Women's post self-efficacy								
	physics 1				physics 2			
Question	1	2	3	4	1	2	3	4
9	6%	24%	60%	10%	5%	28%	61%	6%
10	4%	20%	69%	7%	5%	27%	63%	5%
11	6%	29%	51%	14%	6%	23%	59%	12%
12	5%	30%	58%	7%	4%	28%	59%	9%

Table 42 Percentages of men who answered each self-efficacy question by the options they selected with 1 being the low value (NO!) and 4 being the high values (YES!). The rating scale for the self-efficacy questions was: NO! no yes YES!.

Men's post self-efficacy								
	physics 1				physics 2			
Question	1	2	3	4	1	2	3	4
9	1%	20%	67%	12%	6%	15%	70%	9%
10	2%	15%	68%	15%	3%	14%	73%	10%
11	2%	10%	63%	25%	2%	11%	66%	21%
12	1%	16%	69%	14%	4%	14%	66%	16%

Appendix C SEM for instructors and TA

The path analysis part of the SEM for how perceived recognition consisting only of the question “My physics instructor and/or TA see me as a physics person”, predicts students’ grade in physics 1 is in Figure 36. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): CFI = 0.991 (> 0.90), TLI = 0.965 (> 0.90), RMSEA = 0.039 (<0.08), and SRMR = 0.026 (< 0.08). Similar to the model in the main paper, students’ post perceived recognition directly predicts students’ grades in physics 1 with a similar β value.

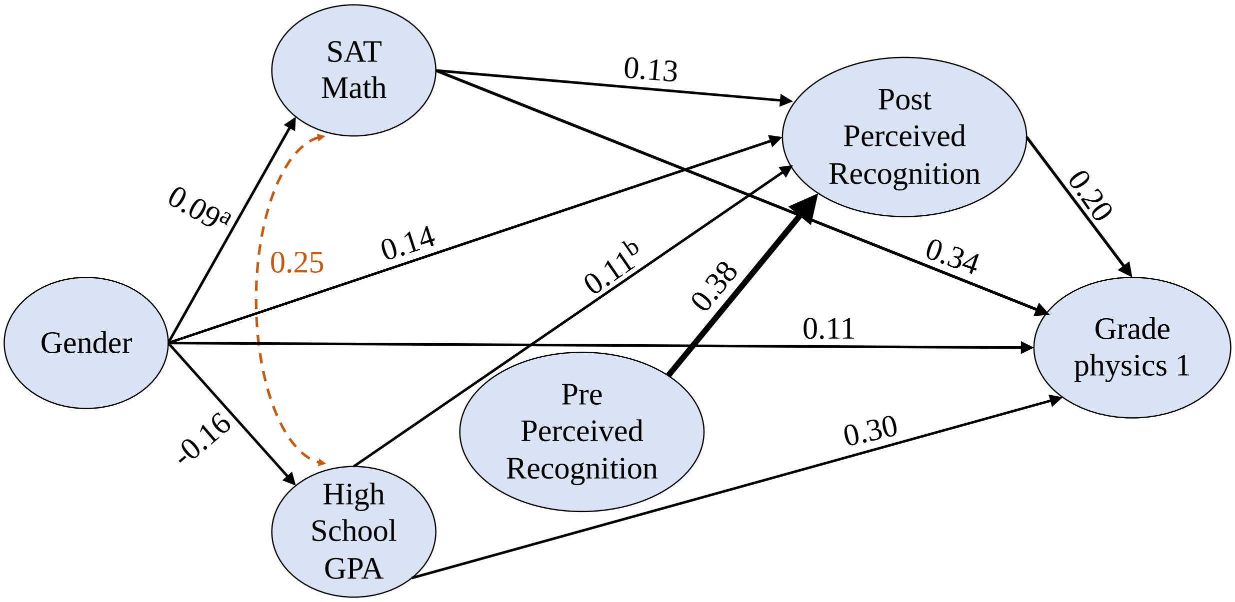


Figure 36 Result of the path analysis part of the SEM between gender and physics 1 grade through students’ perceived recognition as a physics person. The line thickness indicates the relative magnitude of β values. The dashed line indicates covariances between constructs. All *p*-values are indicated by no superscript for $p <$ 0.001, superscript “a” for $p = 0.025$, and superscript “b” for $p = 0.001$.

Appendix D Percentages of Male and Female Students Who Selected Each Choice for Each Survey Item

Below, we provide the percentages of men and women who selected each answer choice for each sense of belonging item in the pre and post survey in physics 1. This distribution provides a sense of how students shifted their answers from pre to post survey. From Table 43, we can see that in general, the percentage of women who selected 1 or 5 increased from pre to post, and the percentage of men who selected 5 increased from pre to post while the percentage of men who selected 3 decreased. In addition, more women than men selected the answer choices 1 (not at all true) or 2 (a little true), while men were more likely to select the answer choices 4 (mostly true) and 5 (completely true).

Table 43 Percentages of 521 women and 293 in physics 1 who responded to each sense of belonging item by the options they selected with 1 being the low value (not at all true) and 5 being the high value (completely true). The rating scale for the sense of belonging items was: not at all true, a little true, somewhat true, mostly true, and completely true.

Women's sense of belonging distribution across choices										
	pre test					post test				
Question	1	2	3	4	5	1	2	3	4	5
1	5%	18%	45%	25%	7%	11%	24%	33%	23%	9%
2	3%	12%	34%	33%	18%	4%	9%	18%	34%	35%
3	3%	18%	44%	31%	4%	13%	24%	31%	23%	9%
4	6%	13%	32%	33%	16%	6%	12%	24%	28%	30%
Men's sense of belonging distribution across choices										
1	3%	15%	41%	28%	13%	3%	12%	33%	36%	16%
2	3%	12%	31%	27%	27%	2%	4%	18%	29%	47%
3	1%	9%	48%	33%	9%	5%	16%	31%	35%	13%
4	2%	10%	26%	39%	23%	3%	6%	18%	31%	42%

Appendix E Percentages of men and women who answered each question

In Table 44 and Table 45, we provide the percentages of men and women who selected each answer choice for each question. This distribution provides a sense of how students shifted their answers from the pre-test to the post-test.

Table 44 Percentages of women who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

Women								
	Pre-Test				Post-Test			
Question	1	2	3	4	1	2	3	4
Physics Identity								
1	27%	59%	13%	1%	45%	45%	9%	1%
Physics Perceived Recognition								
2	31%	53%	15%	1%	40%	46%	13%	1%
3	30%	55%	13%	2%	38%	46%	13%	3%
4	16%	52%	29%	3%	26%	38%	32%	4%
Physics Self-efficacy								
5	14%	45%	39%	2%	11%	29%	54%	6%
6	4%	19%	68%	9%	6%	29%	59%	6%
7	1%	6%	66%	27%	11%	39%	41%	9%
8	1%	14%	66%	19%	7%	39%	48%	6%
Physics Interest								
9	24%	53%	20%	3%	23%	35%	33%	9%
10	5%	32%	58%	5%	13%	38%	46%	3%
11	5%	44%	45%	6%	18%	54%	26%	2%
12	5%	33%	54%	8%	15%	40%	41%	4%

Table 45 Percentages of men who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

Men								
	Pre-Test				Post-Test			
Question	1	2	3	4	1	2	3	4
Physics Identity								
1	12%	62%	24%	2%	21%	56%	21%	2%
Physics Perceived Recognition								
2	17%	64%	18%	1%	25%	49%	23%	3%
3	15%	63%	21%	1%	21%	48%	28%	3%
4	8%	52%	35%	5%	12%	41%	39%	8%
Physics Self-efficacy								
5	3%	38%	53%	6%	5%	21%	63%	11%
6	1%	9%	80%	10%	2%	19%	65%	14%
7	0%	2%	54%	44%	5%	14%	57%	24%
8	0%	5%	65%	31%	3%	22%	60%	15%
Physics Interest								
9	13%	42%	37%	8%	10%	26%	51%	13%
10	2%	17%	72%	10%	3%	28%	59%	10%
11	1%	30%	58%	11%	3%	49%	41%	7%
12	3%	17%	68%	12%	5%	31%	56%	8%

Appendix F SEM model with science identity

We provide path analysis part of the SEM for how the factors that predict physics identity also predict students' science identity. The science identity question on our survey asked students whether they "strongly disagree, disagree, agree, strongly agree" with "I see myself as a scientist". The survey item was adapted from the survey by Godwin et. al. [6] and re-validated in our own context. The results of the SEM are presented visually in Figure 37. The model fit indices indicate a good fit to the data (acceptable fit thresholds in parentheses): CFI = 0.978 (> 0.90), TLI = 0.971 (> 0.90), RMSEA = 0.043 (<0.08), and SRMR = 0.033 (< 0.08). All three of the mediating constructs (physics perceived recognition, self-efficacy, and interest) predict physics identity at the end of the physics course similar to Figure 15. The mediating constructs of physics self-efficacy also predict science identity at the end of the physics course.

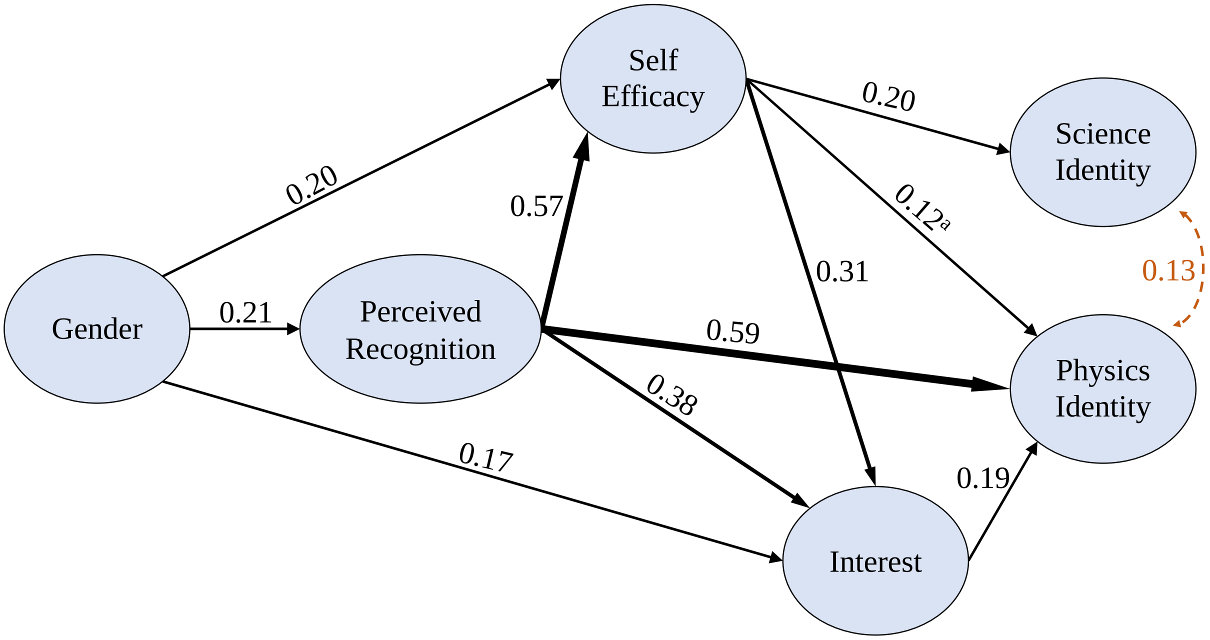


Figure 37 Result of the path analysis part of the SEM with mediation between gender and science/physics identity through physics perceived recognition, self-efficacy, and interest. Each line thickness qualitatively denotes the relative magnitude of the standardized regression coefficients β shown. The dashed lines indicate covariance. All *p*-values for β are indicated by no superscript for $p < 0.001$, "a" for $p = 0.015$, "b".

Appendix G Students' Pre and Post Beliefs

The descriptive statistics of the students' pre and post beliefs in each course are shown in Table 46 and Table 47. Since the perceived recognition and identity constructs were not included in our survey at the beginning of physics 1 in the first year of study, we only include data for matched students (students who took both the pre and post survey) in the second year. However, the students are not matched from physics 1 to physics 2. Therefore, the sample size is smaller than that in the main text; however, the results from one year shown below are qualitatively similar to the results for two years (discussed in the main text).

Table 46 Mean physics 1 pre and post predictor and outcome values by gender as well as statistical significance (p-values) and effect sizes (Cohen's d) by gender for 260 women and 126 men. All p-values < 0.001.

Predictors and	Pre Mean		Cohen's		Post Mean		Cohen's	
Outcomes	Male	Female	*d*	*p*-value	Male	Female	*d*	*p*-value
Perceived Recognition	2.11	1.98	0.20	0.065	2.16	1.95	0.35	0.001
Self-Efficacy	3.04	2.85	0.48	<0.001	2.90	2.70	0.38	<0.001
Interest	2.88	2.50	0.68	<0.001	2.80	2.35	0.73	<0.001
Physics Identity	2.26	1.94	0.44	<0.001	2.08	1.81	0.37	0.001

Table 47 Mean physics 2 pre and post predictor and outcome values by gender as well as statistical significance (p-values) and effect sizes (Cohen's *d*) by gender for 274 women and 148 men. All *p*-values < 0.001.

Predictors and Outcomes	Pre Mean		Cohen's *d*	*p*-value	Post Mean		Cohen's *d*	*p*-value
	Male	Female			Male	Female		
Perceived Recognition	2.21	2.03	0.28	0.006	2.26	2.05	0.31	0.003
Self-Efficacy	3.02	2.81	0.53	<0.001	2.94	2.74	0.38	<0.001
Interest	2.80	2.31	0.91	<0.001	2.79	2.26	0.87	<0.001
Physics Identity	2.16	1.91	0.37	<0.001	2.19	1.88	0.43	<0.001

Appendix H Percentages of Male and Female Students Who Selected Each Choice for Each Survey Item

Below, we provide the percentages of men and women who selected each answer choice for each question in the pre and post survey in physics 1 and physics 2. This distribution provides a sense of how students shifted their answers from pre to post survey. Table 48 and Table 49 are for women and men, respectively, in physics 1 while Table 50 and Table 51 are for women and men, respectively, in physics 2.

Table 48 Percentages of 260 women in physics 1 who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was: strongly disagree, disagree, agree, strongly agree.

Women								
	Pre survey				Post survey			
Question	1	2	3	4	1	2	3	4
Physics Identity								
1	27%	59%	13%	1%	45%	45%	9%	1%
Physics Perceived Recognition								
2	31%	53%	15%	1%	40%	46%	13%	1%
3	30%	55%	13%	2%	38%	46%	13%	3%
4	16%	52%	29%	3%	26%	38%	32%	4%
Physics Self-efficacy								
5	14%	45%	39%	2%	11%	29%	54%	6%
6	4%	19%	68%	9%	6%	29%	59%	6%
7	1%	6%	66%	27%	11%	39%	41%	9%
8	1%	14%	66%	19%	7%	39%	48%	6%
Physics Interest								
9	24%	53%	20%	3%	23%	35%	33%	9%
10	5%	32%	58%	5%	13%	38%	46%	3%
11	5%	44%	45%	6%	18%	54%	26%	2%
12	5%	33%	54%	8%	15%	40%	41%	4%

Table 49 Percentages of 126 men in physics 1 who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was: strongly disagree, disagree, agree, strongly agree.

Men								
	Pre survey				Post survey			
Question	1	2	3	4	1	2	3	4
Physics Identity								
1	9%	59%	28%	4%	19%	58%	19%	4%
Physics Perceived Recognition								
2	20%	60%	17%	3%	17%	59%	23%	1%
3	21%	59%	18%	2%	18%	62%	18%	2%
4	12%	53%	30%	5%	8%	54%	32%	6%
Physics Self-efficacy								
5	5%	33%	59%	3%	4%	26%	60%	10%
6	1%	10%	77%	12%	5%	14%	66%	15%
7	0%	3%	56%	41%	2%	13%	60%	25%
8	0%	5%	69%	26%	2%	22%	64%	12%
Physics Interest								
9	8%	37%	36%	19%	3%	26%	42%	29%
10	1%	16%	66%	17%	5%	20%	64%	11%
11	0%	19%	67%	14%	7%	30%	54%	9%
12	3%	20%	56%	21%	4%	31%	52%	13%

Table 50 Percentages of 274 women in physics 2 who answered each question in physics 2 by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree

Women								
	Pre survey				Post survey			
Question	1	2	3	4	1	2	3	4
Physics Identity								
1	28%	54%	17%	1%	30%	53%	15%	2%
Physics Perceived Recognition								
2	26%	55%	18%	1%	26%	53%	19%	2%
3	25%	55%	18%	2%	26%	49%	23%	2%
4	17%	49%	31%	3%	19%	47%	30%	4%
Physics Self-efficacy								
5	5%	30%	64%	1%	7%	30%	57%	6%
6	2%	19%	75%	4%	5%	27%	64%	4%
7	2%	15%	70%	13%	4%	20%	62%	14%
8	1%	18%	73%	8%	2%	26%	61%	11%
Physics Interest								
9	22%	48%	25%	5%	23%	45%	26%	6%
10	10%	40%	48%	2%	14%	40%	42%	4%
11	9%	59%	30%	2%	15%	59%	24%	2%
12	11%	37%	48%	4%	12%	42%	40%	6%

Table 51 Percentages of 148 men in physics 2 who answered each question in physics 2 by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high values (YES! and strongly agree). The rating scale for the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

Men								
	Pre survey				Post survey			
Question	1	2	3	4	1	2	3	4
Physics Identity								
1	13%	61%	23%	3%	16%	53%	26%	5%
Physics Perceived Recognition								
2	16%	57%	23%	4%	18%	53%	25%	4%
3	16%	57%	23%	4%	18%	48%	26%	8%
4	10%	52%	35%	3%	12%	43%	40%	5%
Physics Self-efficacy								
5	2%	21%	68%	9%	7%	20%	63%	10%
6	1%	7%	82%	10%	3%	12%	72%	13%
7	1%	6%	66%	27%	1%	11%	65%	23%
8	1%	11%	71%	17%	3%	16%	67%	14%
Physics Interest								
9	7%	30%	48%	15%	6%	26%	42%	26%
10	3%	19%	61%	17%	3%	23%	56%	18%
11	2%	30%	59%	9%	5%	41%	45%	9%
12	4%	22%	62%	12%	3%	30%	53%	14%

Appendix I Students Beliefs

The descriptive statistics of the students' beliefs in each course are shown for women in Table 52 and Table 53 and for men in Table 54 and Table 55.

Table 52 Percentages of 292 women in the pre test and 327 women in the post test in physics 1 who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high value (YES! and strongly agree). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity, effectiveness of peer interaction, and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

Motivational Belief	Question	1	2	3	4
Pre Self-efficacy	1	14%	45%	39%	2%
	2	4%	19%	68%	9%
	3	1%	6%	66%	27%
	4	1%	14%	66%	19%
Pre Interest	1	24%	53%	20%	3%
	2	5%	32%	58%	5%
	3	5%	44%	45%	6%
	4	5%	33%	54%	8%
Post Self-efficacy	1	11%	29%	54%	6%
	2	6%	29%	59%	6%

	3	11%	39%	41%	9%
	4	7%	39%	48%	6%
Post Interest	1	23%	35%	33%	9%
	2	13%	38%	46%	3%
	3	18%	54%	26%	2%
	4	15%	40%	41%	4%
Post Identity	1	45%	45%	9%	1%
Post Perceived Recognition	1	40%	46%	12%	2%
	2	38%	46%	13%	3%
	3	28%	40%	29%	3%
Post Peer Interaction	1	9%	23%	49%	19%
	2	11%	35%	43%	11%
	3	13%	34%	43%	10%
	4	12%	37%	41%	10%

Table 53 Percentages of 327 women in physics 1 who answered each belonging question by the options they selected with 1 being the low value (not at all true) and 5 being the high value (completely true). The rating

scale for the physics belonging questions was not at all true, a little true, somewhat true, mostly true, and completely true.

Motivational Belief	Question	1	2	3	4	5
Post Belonging	1	23%	26%	30%	14%	7%
	2	9%	17%	17%	31%	26%
	3	21%	27%	30%	16%	6%
	4	13%	19%	31%	26%	11%
	5	14%	15%	23%	21%	27%

Table 54 Percentages of 144 men in the pre test and 174 men in the post test in physics 1 who answered each question by the options they selected with 1 being the low value (NO! and strongly disagree) and 4 being the high value (YES! and strongly agree). The rating scale for most of the self-efficacy and interest questions was NO! no yes YES! while the rating scale for the physics identity, peer interaction, and perceived recognition questions was strongly disagree, disagree, agree, strongly agree.

Motivational Belief	Question	1	2	3	4
Pre Self-efficacy	1	3%	38%	53%	6%
	2	1%	9%	80%	10%
	3	0%	2%	54%	44%
	4	0%	5%	64%	31%
Pre Interest	1	13%	42%	37%	8%
	2	2%	17%	71%	10%

	3	1%	30%	58%	11%
	4	3%	17%	68%	12%
Post Self-efficacy	1	5%	21%	63%	11%
	2	2%	18%	65%	15%
	3	5%	14%	57%	24%
	4	4%	22%	60%	14%
Post Interest	1	10%	26%	51%	13%
	2	3%	28%	59%	10%
	3	3%	48%	41%	8%
	4	5%	31%	56%	8%
Post Identity	1	21%	56%	21%	2%
Post Perceived Recognition	1	25%	49%	22%	4%
	2	21%	48%	28%	3%
	3	16%	43%	36%	5%
Post Peer Interaction	1	1%	25%	54%	20%
	2	3%	22%	58%	17%
	3	3%	25%	53%	19%
	4	3%	27%	54%	16%

Table 55 Percentages of 174 men in physics 1 who answered each belonging question by the options they selected with 1 being the low value (not at all true) and 5 being the high value (completely true). The rating scale for the physics belonging questions was not at all true, a little true, somewhat true, mostly true, and completely true.

Motivational Belief	Question	1	2	3	4	5
Post Belonging	1	10%	20%	32%	26%	12%
	2	2%	8%	16%	34%	40%
	3	3%	24%	27%	31%	15%
	4	4%	17%	31%	31%	17%
	5	2%	9%	16%	34%	39%

Bibliography

1. Archer, L., et al., *The "exceptional" physics girl: A sociological analysis of multimethod data from young women aged 10–16 to explore gendered patterns of post-16 participation.* American Educational Research Journal, 2017. **54**(1): p. 88-126.

2. Monsalve, C., et al. *Examining the relationship between career outcome expectations and physics identity*. in *Proceedings of the Physics Education Research Conference*. 2016. Sacramento, CA.

3. Lock, R.M., Z. Hazari, and G. Potvin, *Physics career intentions: The effect of physics identity, math identity, and gender.* AIP Conference Proceedings, 2013. **1513**(1): p. 262-265.

4. Kalender, Z.Y., et al., *Gendered patterns in the construction of physics identity from motivational factors.* Physical Review Physics Education Research, 2019. **15**(2): p. 020119.

5. Hazari, Z., et al., *Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study.* Journal of Research in Science Teaching, 2010. **47**(8): p. 978-1003.

6. Godwin, A., et al., *Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice.* Journal of Engineering Education, 2016. **105**(2): p. 312-340.

7. Patrick, A., C. Seepersad, and M. Borrego. *A combined model for predicting engineering identity in undergraduate students*. in *2018 ASEE Annual Conference & Exposition*. 2018. Salt Lake City, UT.

8. Tai, R.H. and P.M. Sadler, *Gender differences in introductory undergraduate physics performance: University physics versus college physics in the USA.* International Journal of Science Education, 2001. **23**(10): p. 1017-1037.

9. Henderson, R., et al., *Exploring the gender gap in the Conceptual Survey of Electricity and Magnetism.* Physical Review Physics Education Research, 2017. **13**(2): p. 020114.

10. Wee, A.T., B.E. Baaquie, and A.C. Huan, *Gender differences in undergraduate physics examination performance and learning strategies in Singapore.* Physics Education, 1993. **28**(3): p. 158.

11. Karim, N.I., A. Maries, and C. Singh, *Do evidence-based active-engagement courses reduce the gender gap in introductory physics?* European Journal of Physics, 2018. **39**(2): p. 025701.

12. Hazari, Z., R.H. Tai, and P.M. Sadler, *Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors.* Science Education, 2007. **91**(6): p. 847-876.

13. Kost, L.E., S.J. Pollock, and N.D. Finkelstein, *Characterizing the gender gap in introductory physics.* Physical Review Special Topics - Physics Education Research, 2009. **5**(1): p. 010101.

14. Cavallo, A.M.L., W.H. Potter, and M. Rozman, *Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors.* School Science and Mathematics, 2004. **104**(6): p. 288-300.

15. Felder, R.M., et al., *A longitudinal study of engineering student performance and retention. III. Gender differences in student performance and attitudes.* Journal of Engineering Education, 1995. **84**(2): p. 151-163.

16. Strenta, A.C., et al., *Choosing and leaving science in highly selective institutions.* Research in Higher Education, 1994. **35**(5): p. 513-547.

17. Sadler, P.M. and R.H. Tai, *Success in introductory college physics: The role of high school preparation.* Science Education, 2001. **85**(2): p. 111-136.

18. Madsen, A., 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?* Physical Review Special Topics - Physics Education Research, 2013. **9**(2): p. 020121.

19. Lorenzo, M., C.H. Crouch, and E. Mazur, *Reducing the gender gap in the physics classroom.* American Journal of Physics, 2006. **74**(2): p. 118-122.

20. Hart, G.E. and P.D. Cottle, *Academic backgrounds and achievement in college physics.* The Physics Teacher, 1993. **31**(8): p. 470-475.

21. Jones, M.G., A. Howe, and M.J. Rua, *Gender differences in students' experiences, interests, and attitudes toward science and scientists.* Science Education, 2000. **84**(2): p. 180-192.

22. Eddy, S.L. and S.E. Brownell, *Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines.* Physical Review Physics Education Research, 2016. **12**(2): p. 020106.

23. Leslie, S.-J., et al., *Expectations of brilliance underlie gender distributions across academic disciplines.* Science, 2015. **347**(6219): p. 262-265.

24. Bian, L., S.-J. Leslie, and A. Cimpian, *Gender stereotypes about intellectual ability emerge early and influence children's interests.* Science, 2017. **355**(6323): p. 389-391.

25. Blickenstaff, J.C., *Women and science careers: Leaky pipeline or gender filter?* Gender and Education, 2005. **17**(4): p. 369-386.

26. Marshman, E., et al., *A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences.* Canadian Journal of Physics, 2018. **96**(4): p. 391-405.

27. Bandura, A., *Self-efficacy*, in *Encyclopedia of Psychology*, R.J. Corsini, Editor. 1994, Wiley: New York. p. 368-369.

28. Bandura, A., *Self-efficacy: Toward a unifying theory of behavioral change.* Psychological Review, 1977. **84**(2): p. 191.

29. Nissen, J.M. and J.T. Shemwell, *Gender, experience, and self-efficacy in introductory physics.* Physical Review Physics Education Research, 2016. **12**(2): p. 020105.

30. Lindstrøm, C. 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), 2011. **19**(2): p. 1-19.

31. Sawtelle, V., E. Brewe, and L.H. Kramer, *Exploring the relationship between self-efficacy and retention in introductory physics.* Journal of Research in Science Teaching, 2012. **49**(9): p. 1096-1121.

32. Correll, S.J., *Constraints into preferences: Gender, status, and emerging career aspirations.* American Sociological Review, 2004. **69**(1): p. 93-113.

33. Zimmerman, B.J., *Self-efficacy: An essential motive to learn.* Contemporary Educational Psychology, 2000. **25**(1): p. 82-91.

34. Fencl, H. and K. Scheel, *Research and Teaching: Engaging students - An examination of the effects of teaching strategies on self-efficacy and course climate in a nonmajors physics course.* Journal of College Science Teaching, 2005. **35**(1): p. 20.

35. Schunk, D.H. and F. Pajares, *The development of academic self-efficacy*, in *Development of achievement motivation*. 2002, Elsevier. p. 15-31.

36. Bouffard-Bouchard, T., S. Parent, and S. Larivee, *Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students.* International Journal of Behavioral Development, 1991. **14**(2): p. 153-164.

37. Britner, S.L. and F. Pajares, *Sources of science self-efficacy beliefs of middle school students.* Journal of Research in Science Teaching, 2006. **43**(5): p. 485-499.

38. Bailey, J.M., et al., *Meeting students halfway: Increasing self-efficacy and promoting knowledge change in astronomy.* Physical Review Physics Education Research, 2017. **13**(2): p. 020140.

39. Britner, S.L., *Motivation in high school science students: A comparison of gender differences in life, physical, and earth science classes.* Journal of Research in Science Teaching, 2008. **45**(8): p. 955-970.

40. Vincent-Ruz, P. and C.D. Schunn, *The increasingly important role of science competency beliefs for science learning in girls.* Journal of Research in Science Teaching, 2017. **54**(6): p. 790-822.

41. Wang, M.-T. and J. Degol, *Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields.* Developmental Review, 2013. **33**(4): p. 304-340.

42. Cheryan, S., et al., *Why are some STEM fields more gender balanced than others?* Psychological Bulletin, 2017. **143**(1): p. 1.

43. Marshman, E., et al., *Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?* Physical Review Physics Education Research, 2018. **14**(2): p. 020123.

44. Lichtenberger, E. and C. George-Jackson, *Predicting high school students' interest in majoring in a STEM field: Insight into high school students' postsecondary plans.* Journal of Career and Technical Education, 2013. **28**(1): p. 19-38.

45. Harackiewicz, J.M., et al., *Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation.* Journal of Educational Psychology, 2002. **94**(3): p. 562-575.

46. Hidi, S., *Interest: A unique motivational variable.* Educational Research Review, 2006. **1**(2): p. 69-82.

47. Häussler, P. and L. Hoffmann, *An intervention study to enhance girls' interest, self-concept, and achievement in physics classes.* Journal of Research in Science Teaching, 2002. **39**(9): p. 870-888.

48. Wigfield, A. and J.S. Eccles, *The development of achievement task values: A theoretical analysis.* Developmental Review, 1992. **12**(3): p. 265-310.

49. Vincent-Ruz, P. and C.D. Schunn, *The nature of science identity and its role as the driver of student choices.* International Journal of STEM Education, 2018. **5**(1): p. 48.

50. Goodenow, C., *Classroom belonging among early adolescent students: Relationships to motivation and achievement.* The Journal of Early Adolescence, 1993. **13**(1): p. 21-43.

51. Moss-Racusin, C.A., et al., *Science faculty's subtle gender biases favor male students*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(41): p. 16474.

52. Doucette, D., R. Clark, and C. Singh, *Hermione and the Secretary: How gendered task division in introductory physics labs can disrupt equitable learning.* European Journal of Physics, 2020. **41**(3): p. 035702.

53. Doucette, D. and C. Singh, *Why are there so few women in physics? Reflections on the experiences of two women.* The Physics Teacher, 2020. **58**(5): p. 297-300.

54. Stets, J.E., et al., *The science identity and entering a science occupation.* Social Science Research, 2017. **64**: p. 1-14.

55. Carlone, H.B. and A. Johnson, *Understanding the science experiences of successful women of color: Science identity as an analytic lens.* Journal of Research in Science Teaching, 2007. **44**(8): p. 1187-1218.

56. Gee, J.P., *Chapter 3: Identity as an analytic lens for research in education.* Review of Research in Education, 2000. **25**(1): p. 99-125.

57. Tonso, K.L., *Student engineers and engineer identity: Campus engineer identities as figured world.* Cultural Studies of Science Education, 2006. **1**(2): p. 273-307.

58. Hazari, Z., et al., *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.* Journal of Research in Science Teaching, 2020. **57**(10): p. 1583-1607.

59. Kalender, Z.Y., et al., *Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way.* Physical Review Physics Education Research, 2019. **15**(2): p. 020148.

60. Hazari, Z. and C. Cass, *Towards meaningful physics recognition: What does this recognition actually look like?* The Physics Teacher, 2018. **56**(7): p. 442-446.

61. Hazari, Z., et al., *The importance of high school physics teachers for female students' physics identity and persistence.* The Physics Teacher, 2017. **55**(2): p. 96-99.

62. Li, Y. and C. Singh, *Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses.* Physical Review Physics Education Research, 2021. **17**(1): p. 010143.

63. Masika, R. and J. Jones, *Building student belonging and engagement: Insights into higher education students' experiences of participating and learning together.* Teaching in Higher Education, 2016. **21**(2): p. 138-150.

64. *TEAM-UP Report*. 2020, American Institute of Physics.

65. *AIP Statistics*. Available from: https://www.aip.org/statistics/data-graphics/percent-physics-bachelors-and-phds-earned-women-classes-1975-through-2016.

66. Cho, S., K.W. Crenshaw, and L. McCall, *Toward a field of intersectionality studies: Theory, applications, and praxis.* Signs: Journal of women in culture and society, 2013. **38**(4): p. 785-810.

67. Mitchell, J.D., C.Y. Simmons, and L.A. Greyerbiehl, *Intersectionality & Higher Education*. 2014, New York, NY: Peter Lang

68. Morton, T.R. and E.C. Parsons, *# BlackGirlMagic: The identity conceptualization of Black women in undergraduate STEM education.* Science Education, 2018. **102**(6): p. 1363-1393.

69. Crenshaw, K., *Mapping the margins: Intersectionality, identity politics, and violence against women of color.* Stan. L. Rev., 1990. **43**: p. 1241.

70. Charleston, L., et al., *Intersectionality and STEM: The role of race and gender in the academic pursuits of African American women in STEM.* Journal of Progressive Policy & Practice, 2014. **2**(3): p. 273-293.

71. Rodriguez, M., R. Barthelemy, and M. McCormick, *Critical race and feminist standpoint theories in physics education research: A historical review and potential applications.* Physical Review Physics Education Research, 2022. **18**(1): p. 013101.

72. Rodriguez, I., et al., *Impact of equity models and statistical measures on interpretations of educational reform.* Physical Review Special Topics-Physics Education Research, 2012. **8**(2): p. 020103.

73. Dasgupta, N., *Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model.* Psychological Inquiry, 2011. **22**(4): p. 231-246.

74. Binning, K.R., et al., *Changing social contexts to foster equity in college science courses: An ecological-belonging intervention.* Psychological Science, 2020. **31**(9): p. 1059-1070.

75. Aguilar, L., G. Walton, and C. Wieman, *Psychological insights for improved physics teaching.* Physics Today, 2014. **67**(5): p. 43-49.

76. Brewe, E., et al., *Toward equity through participation in modeling instruction in introductory university physics.* Physical Review Special Topics - Physics Education Research, 2010. **6**(1): p. 010106.

77. Miyake, A., et al., *Reducing the gender achievement gap in college science: A classroom study of values affirmation.* Science, 2010. **330**(6008): p. 1234-1237.

78. Maries, A., K. Whitcomb, and C. Singh, *Gender inequities throughout STEM.* Journal of College Science Teaching, 2022. **51**(3): p. 27-36.

79. Whitcomb, K.M. and C. Singh, *Underrepresented minority students receive lower grades and have higher rates of attrition across STEM disciplines: A sign of inequity?* International Journal of Science Education, 2021. **43**(7): p. 1054-1089.

80. Whitcomb, K.M. and C. Singh, *For physics majors, gender differences in introductory physics do not inform future physics performance.* European Journal of Physics, 2020. **41**(6): p. 065701.

81. Rodriguez, I., 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.* Physical Review Physics Education Research, 2016. **12**(2): p. 020118.

82. Whitcomb, K., S. Cwik, and C. Singh, *Not all disadvantages are created equal: Racial/Ethnic minority students have largest disadvantage among demographic groups in both STEM and non-STEM GPA.* AERA Open, 2021. **7**(1): p. 1-16.

83. Coletta, V.P., J.A. Phillips, and J. Steinert. *FCI normalized gain, scientific reasoning ability, thinking in physics, and gender effects*. in *AIP conference proceedings*. 2012. American Institute of Physics.

84. Traxler, A., et al., *Gender fairness within the force concept inventory.* Physical Review Physics Education Research, 2018. **14**(1): p. 010103.

85. Blue, J. and P. Heller. *Using matched samples to look for sex differences*. in *AIP Conference Proceedings*. 2004. American Institute of Physics.

86. McCullough, L., *Gender, context, and physics assessment.* Journal of International Women's Studies, 2004. **5**(4): p. 20-30.

87. Mears, M., *Gender differences in the Force Concept Inventory for different educational levels in the United Kingdom.* Physical Review Physics Education Research, 2019. **15**(2): p. 020135.

88. Henderson, R., et al., *Item-level gender fairness in the Force and Motion Conceptual Evaluation and the Conceptual Survey of Electricity and Magnetism.* Physical Review Physics Education Research, 2018. **14**(2): p. 020103.

89. Dew, M., et al., *Gendered performance differences in introductory physics: A study from a large land-grant university.* Physical Review Physics Education Research, 2021. **17**(1): p. 010106.

90. Pollock, S.J., N.D. Finkelstein, and L.E. Kost, *Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?* Physical Review Special Topics-Physics Education Research, 2007. **3**(1): p. 010107.

91. Good, M., 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.* Physical Review Physics Education Research, 2019. **15**(2): p. 020129.

92. Maries, A., 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 Physics Teacher, 2020. **58**: p. 430-433.

93. Watkins, J.E., *Examining Issues of Underrepresented Minority Students in Introductory Physics*. 2010: Harvard University Cambridge, MA.

94. Van Dusen, B. and J. Nissen, *Equity in college physics student learning: A critical quantitative intersectionality investigation.* Journal of Research in Science Teaching, 2020. **57**(1): p. 33-57.

95. Salehi, S., et al., *Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics.* Physical Review Physics Education Research, 2019. **15**(2): p. 020114.

96. Stewart, J., et al., *Mediational effect of prior preparation on performance differences of students underrepresented in physics.* Physical Review Physics Education Research, 2021. **17**(1): p. 010107.

97. Kalender, Z.Y., et al., *Damage caused by women's lower self-efficacy on physics learning.* Physical Review Physics Education Research, 2020. **16**(1): p. 010118.

98. Malespina, A. and C. Singh, *Gender differences in test anxiety and self-efficacy: Why instructors should emphasize low-stakes formative assessments in physics courses.* European Journal of Physics, 2022. **43**(3): p. 035701.

99. Espinosa, T., et al., *Reducing the gender gap in students' physics self-efficacy in a team- and project-based introductory physics class.* Physical Review Physics Education Research, 2019. **15**(1): p. 010132.

100. Whitcomb, K.M., et al., *Comparison of self-efficacy and performance of engineering undergraduate women and men.* International Journal of Engineering Education, 2020. **36**(6): p. 1996-2014.

101. Barthelemy, R.S. and A.V. Knaub, *Gendered motivations and aspirations of university physics students in Finland.* Physical Review Physics Education Research, 2020. **16**(1): p. 010133.

102. Hyater-Adams, S., et al., *Critical look at physics identity: An operationalized framework for examining race and physics identity.* Physical Review Physics Education Research, 2018. **14**(1): p. 010132.

103. Wang, J. and Z. Hazari, *Promoting high school students' physics identity through explicit and implicit recognition.* Physical Review Physics Education Research, 2018. **14**(2): p. 020111.

104. Cwik, S. and C. Singh, *Students' sense of belonging in introductory physics course for bioscience majors predict their grade.* Physical Review Physics Education Research, accepted: 2022.

105. Ladewig, A., M. Keller, and U. Klusmann, *Sense of belonging as an important factor in the pursuit of physics: Does it also matter for female participants of the german physics olympiad?* Frontiers in Psychology, 2020. **11**: p. 11:548781.

106. Lewis, K.L., et al., *Fitting in or opting out: A review of key social-psychological factors influencing a sense of belonging for women in physics.* Physical Review Physics Education Research, 2016. **12**(2): p. 020110.

107. Dweck, C.S., *Mindset: The New Psychology of Success*. 2008: Random House Digital, Inc.

108. Kalender, Z.Y., et al., *Framework for unpacking students' mindsets in physics by gender.* Physical Review Physics Education Research, 2022. **18**: p. 010116.

109. Malespina, A., C.D. Schunn, and C. Singh, *Whose ability and growth matter? Gender, mindset and performance in physics.* International Journal of STEM Education, 2022. **9**(1): p. 28.

110. Cwik, S. and C. Singh, *Damage caused by societal stereotypes: Women have lower physics self-efficacy controlling for grade even in courses in which they outnumber men.* Physical Review Physics Education Research, 2021. **17**(2): p. 020138.

111. Cwik, S., 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*. 2020. Virtual Conference.

112. Doucette, D. and C. Singh, *Share it, don't split it: Can equitable group work improve student outcomes?* The Physics Teacher, 2022. **60**(3): p. 166-168.

113. Rosen, D.J. and A.M. Kelly, *Epistemology, socialization, help seeking, and gender-based views in in-person and online, hands-on undergraduate physics laboratories.* Physical Review Physics Education Research, 2020. **16**(2): p. 020116.

114. Kost-Smith, L.E., S.J. Pollock, and N.D. Finkelstein, *Gender disparities in second-semester college physics: The incremental effects of a ``smog of bias".* Physical Review Special Topics - Physics Education Research, 2010. **6**(2): p. 020112.

115. Blue, J., A.L. Traxler, and X.C. Cid, *Gender matters*. Physics Today, 2018. **71**(3): p. 40.

116. Steele, C.M. and J. Aronson, *Stereotype threat and the intellectual test performance of African Americans.* Journal of Personality and Social Psychology, 1995. **69**(5): p. 797-811.

117. Marchand, G.C. and G. Taasoobshirazi, *Stereotype threat and women's performance in physics.* International Journal of Science Education, 2013. **35**(18): p. 3050-3061.

118. Maries, A., N.I. Karim, and C. Singh, *Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics?* Physical Review Physics Education Research, 2018. **14**(2): p. 020119.

119. Gonsalves, A.J., A. Danielsson, and H. Pettersson, *Masculinities and experimental practices in physics: The view from three case studies.* Physical Review Physics Education Research, 2016. **12**(2): p. 020120.

120. Little, A.J., et al., *Exploring mindset's applicability to students' experiences with challenge in transformed college physics courses.* Physical Review Physics Education Research, 2019. **15**(1): p. 010127.

121. Danielsson, A.T., *Exploring woman university physics students 'doing gender' and 'doing physics'.* Gender and Education, 2012. **24**(1): p. 25-39.

122. Quinn, K.N., et al., *Group roles in unstructured labs show inequitable gender divide.* Physical Review Physics Education Research, 2020. **16**(1): p. 010129.

123. Quichocho, X.R., et al., *Who does physics? Understanding the composition of physicists through the lens of Women of Color and LGBTQ+ women physicists.* 2019 PERC Proceedings, 2019: p. 24-25.

124. Ong, M., *Body projects of young women of color in physics: Intersections of gender, race, and science.* Social Problems, 2005. **52**(4): p. 593-617.

125. Rosa, K. and F.M. Mensah, *Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life.* Physical Review Physics Education Research, 2016. **12**(2): p. 020113.

126. Cochran, G., T. Hodapp, and E. Brown. *Identifying barriers to ethnic/racial minority students' participation in graduate physics*. in *Physics Education Research Conference*. 2018. Cincinnati, OH.

127. Boveda, M. and A.E. Weinberg, *Facilitating intersectionally conscious collaborations in physics education.* The Physics Teacher, 2020. **58**: p. 480.

128. Johnson, A., *An intersectional physics identity framework for studying physics settings*, in *Physics Education and Gender: Identity as an Analytic Lens for Research*, A.J. Gonsalves and A.T. Danielsson, Editors. 2020, Springer International Publishing: Cham. p. 53-80.

129. Kelly, A.M., *Social cognitive perspective of gender disparities in undergraduate physics.* Physical Review Physics Education Research, 2016. **12**(2): p. 020116.

130. Canning, E.A., et al., *STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes.* Science Advances, 2019. **5**(2): p. eaau4734.

131. Johnson, A., et al., *Common challenges faced by women of color in physics, and actions faculty can take to minimize those challenges.* The Physics Teacher, 2017. **55**(6): p. 356-360.

132. Traxler, A.L., et al., *Enriching gender in physics education research: A binary past and a complex future.* Physical Review Physics Education Research, 2016. **12**(2): p. 020114.

133. Crowley, K., et al., *Parents explain more often to boys than to girls during shared scientific thinking.* Psychological Science, 2001. **12**(3): p. 258-261.

134. Cwik, S. and C. Singh, *How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses.* Physical Review Physics Education Research, 2021. **17**(2): p. 020143.

135. Kanim, S. and X.C. Cid, *Demographics of physics education research.* Physical Review Physics Education Research, 2020. **16**: p. 020106.

136. Lee, S. and S. Hershberger, *A simple rule for generating equivalent models in covariance structure modeling.* Multivariate Behavioral Research, 1990. **25**(3): p. 313-334.

137. Kline, R.B., *Principles and Practice of Structural Equation Modeling*. 2015: Guilford publications.

138. Team, R.C., *R: A language and environment for statistical computing*. 2013.

139. MacCallum, R.C., M.W. Browne, and H.M. Sugawara, *Power analysis and determination of sample size for covariance structure modeling.* Psychological Methods, 1996. **1**(2): p. 130.

140. Li, Y., 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 Physics Teacher, 2020. **58**(7): p. 484-487.

141. Santana, L.M. and C. Singh. *Negative impacts of an unwelcoming physics environment on undergraduate women*. in *Physics Education Research Conference 2021*. 2021. Virtual Conference.

142. Kishimoto, K., *Anti-racist pedagogy: From faculty's self-reflection to organizing within and beyond the classroom.* Race Ethnicity and Education, 2018. **21**(4): p. 540-554.

143. Cochran, G. and M. Boveda. *A framework for improving diversity work in physics*. in *PERC Proceedings*. 2020. Virtual Conference.

144. Cochran, G., M. Boveda, and C. Prescod-Weinstein, *Intersectionality in STEM Education Research*, in *Handbook of Research in STEM Education*, Routledge, Editor. 2020. p. 257-266.

145. Center, N.R. *Science & Engineering Degree Attainment: 2004-2014*. 2015; From 2004 to 2014, science and engineering degrees increased in prevalence for both genders. The trend was driven by growth in the so-called "hard sciences."]. Available from: https://nscresearchcenter.org/snapshotreport-degreeattainment15/.

146. *Percent of Physics Bachelors and PhDs Earned by Women, Classes of 1975 through 2016. American Institute of Physics*. 2020; Available from: https://www.aip.org/statistics/data-graphics/percent-physics-bachelors-and-phds-earned-women-classes-1975-through-2016.

147. Seymour, E., *Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology.* Science Education, 2002. **86**(1): p. 79-105.

148. Brainard, S.G. and L. Carlin, *A six-year longitudinal study of undergraduate women in engineering and science.* Journal of Engineering Education, 1998. **87**(4): p. 369-375.

149. Whitten, B.L., S. Foster, and M. Duncombe, *What works for women in undergraduate physics?* Physics Today, 2003. **56**(9): p. 46.

150. Tsui, L., *Effective Strategies to Increase Diversity in STEM Fields: A Review of the Research Literature.* The Journal of Negro Education, 2007. **76**(4): p. 555-581.

151. Walton, G.M., et al., *Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering.* Journal of Educational Psychology, 2015. **107**(2): p. 468-485.

152. Harackiewicz, J.M., et al., *Closing achievement gaps with a utility-value intervention: Disentangling race and social class.* Journal of Personality and Social Psychology, 2016. **111**(5): p. 745.

153. Seymour, E., N.M. Hewitt, and C.M. Friend, *Talking About Leaving: Why Undergraduates Leave the Sciences*. Vol. 12. 1997, Boulder, CO: Westview Press.

154. Tobias, S., *They're Not Dumb, They're Different: Stalking the Second Tier*. 1990, Tucson, AZ: Research Corporation.

155. Correll, S.J., *Gender and the career choice process: The role of biased self-assessments.* American Journal of Sociology, 2001. **106**(6): p. 1691-1730.

156. Eccles, J.S., *Understanding women's educational and occupational choices: Applying the Eccles et al. model of achievement-related choices.* Psychology of Women Quarterly, 1994. **18**(4): p. 585-609.

157. Benbow, C.P. and L.L. Minor, *Mathematically talented males and females and achievement in the high school sciences.* American Educational Research Journal, 1986. **23**(3): p. 425-436.

158. Chemers, M.M., et al., *The role of efficacy and identity in science career commitment among underrepresented minority students.* Journal of Social Issues, 2011. **67**(3): p. 469-491.

159. Robinson, K.A., et al., *Science identity development trajectories in a gateway college chemistry course: Predictors and relations to achievement and STEM pursuit.* Contemporary educational psychology, 2019. **56**: p. 180-192.

160. Stewart, J., et al., *Using the Social Cognitive Theory Framework to Chart Gender Differences in the Developmental Trajectory of STEM Self-Efficacy in Science and Engineering Students.* Journal of Science Education and Technology, 2020. **29**(6): p. 758-773.

161. Schunk, D.H. and F. Pajares, *The Development of Academic Self-Efficacy*, in *Development of Achievement Motivation: A Volume in the Educational Psychology Series*, A. Wigfield and J.S. Eccles, Editors. 2002, Academic Press: San Diego. p. 15-31.

162. Dou, R., et al., *Understanding the development of interest and self-efficacy in active-learning undergraduate physics courses.* International Journal of Science Education, 2018. **40**(13): p. 1587-1605.

163. Vincent-Ruz, P., et al., *The effect of math SAT on women's chemistry competency beliefs.* Chemistry Education Research and Practice, 2018. **19**(1): p. 342-351.

164. Miller, K., et al., *Response switching and self-efficacy in Peer Instruction classrooms.* Physical Review Special Topics-Physics Education Research, 2015. **11**(1): p. 010104.

165. Dou, R., H. Cian, and V. Espinosa-Suarez, *Undergraduate STEM majors on and off the pre-med/health track: A STEM identity perspective.* CBE—Life Sciences Education, 2021. **20**(2): p. 1-12.

166. Maloney, E.A., J.R. Sattizahn, and S.L. Beilock, *Anxiety and cognition.* Wiley Interdisciplinary Reviews: Cognitive Science, 2014. **5**(4): p. 403-411.

167. Raelin, J.A., et al., *The gendered effect of cooperative education, contextual support, and self‐efficacy on undergraduate retention.* Journal of Engineering Education, 2014. **103**(4): p. 599-624.

168. Henderson, R., V. Sawtelle, and J.M. Nissen, *Gender & self-efficacy: A call to physics educators.* The Physics Teacher, 2020. **58**(5): p. 345-348.

169. Upson, S. and L.F. Friedman, *Where are all the female geniuses?* Scientific American Mind, 2012. **23**(5): p. 63-65.

170. Glynn, S.M., et al., *Science motivation questionnaire II: Validation with science majors and nonscience majors.* Journal of Research in Science Teaching, 2011. **48**(10): p. 1159-1176.

171. Learning Activation Lab. *Activation lab tools: Measures and data collection instruments*. 2017; Available from: http://www.activationlab.org/tools/.

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

173. Likert, R., *A technique for the measurement of attitudes.* Archives of Psychology, 1932. **22**(140): p. 55.

174. Redish, E.F., J.M. Saul, and R.N. Steinberg, *Student expectations in introductory physics.* American Journal of Physics, 1998. **66**(3): p. 212-224.

175. *PERTS Academic Mindsets Assessment*. 2020; Available from: https://www.perts.net/orientation/ascend.

176. Cronbach, L.J., *Coefficient alpha and the internal structure of tests.* Psychometrika, 1951. **16**(3): p. 297-334.

177. Embretson, S.E. and S.P. Reise, *Item Response Theory for Psychologists*. 2000, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. xi, 371-xi, 371.

178. ; Available from: https://www.stata.com/meeting/australia15/abstracts/materials/oceania15_rosier.pdf.

179. Norman, G., *Likert scales, levels of measurement and the "laws" of statistics.* Advances in health sciences education, 2010. **15**(5): p. 625-632.

180. De Winter, J. and D. Dodou, *Five-point likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012).* Practical Assessment, Research, and Evaluation, 2010. **15**(1): p. 11.

181. Samejima, F., *Estimation of latent ability using a response pattern of graded scores*, in *Psychometrika monograph supplement*. 1969. p. 17.

182. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*. 2013: Routledge.

183. Cohen, J., et al., *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 2013: Routledge.

184. Welsch, D.M. and M. Winden, *Student gender, counselor gender, and college advice.* Education economics, 2019. **27**(2): p. 112-131.

185. Yeager, D.S. and G.M. Walton, *Social-psychological interventions in education: They're not magic.* Review of Educational Research, 2011. **81**(2): p. 267-301.

186. Binning, K., et al., *Changing social norms to foster the benefits of collaboration in diverse workgroups.* Psychological Science, 2020. **31**(9): p. 1059-1070.

187. Tomarken, A.J. and N.G. Waller, *Structural equation modeling: Strengths, limitations, and misconceptions.* Annual Review of Clinical Psychology, 2005. **1**(1): p. 31-65.

188. Dixon-Román, E., H. Everson, and J. McArdle, *Race, poverty and SAT scores: Modeling the influences of family income on black and white high school students' SAT performance.* Teachers College Record, 2013. **115**(4): p. 1-33.

189. Reeves, R. and D. Halikias, *Race gaps in SAT scores highlight inequality and hinder upward mobility.* Washington, DC: Brookings Institute, 2017.

190. Santelices, M.V. and M. Wilson, *Unfair treatment? The case of Freedle, the SAT, and the standardization approach to differential item functioning.* Harvard Educational Review, 2010. **80**(1): p. 106-134.

191. Geiser, S., *The growing correlation between race and SAT scores: New findings from California*. 2015.

192. Walton, G.M. and G.L. Cohen, *A question of belonging: Race, social fit, and achievement.* Journal of personality and social psychology, 2007. **92**(1): p. 82.

193. Rodriguez, S., K. Cunningham, and A. Jordan, *STEM identity development for Latinas: The role of self-and outside recognition.* Journal of Hispanic Higher Education, 2019. **18**(3): p. 254-272.

194. Aschbacher, P.R., E. Li, and E.J. Roth, *Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine.* Journal of Research in Science Teaching, 2010. **47**(5): p. 564-582.

195. Starr, C.R., et al., *Engaging in science practices in classrooms predicts increases in undergraduates' STEM motivation, identity, and achievement: A short-term longitudinal study.* Journal of Research in Science Teaching, 2020. **57**(7): p. 1093-1118.

196. Grunspan, D.Z., et al., *Males under-estimate academic performance of their female peers in undergraduate biology classrooms.* PloS one, 2016. **11**(2): p. e0148405.

197. Knobloch-Westerwick, S., C.J. Glynn, and M. Huge, *The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest.* Science Communication, 2013. **35**(5): p. 603-625.

198. Lincoln, A.E., et al., *The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s.* Social Studies of Science, 2012. **42**(2): p. 307-320.

199. Lock, R.M., Z. Hazari, and G. Potvin, *Impact of out-of-class science and engineering activities on physics identity and career intentions.* Physical Review Physics Education Research, 2019. **15**(2): p. 020137.

200. Hazari, Z., C. Cass, and C. Beattie, *Obscuring power structures in the physics classroom: Linking teacher positioning, student engagement, and physics identity development.* Journal of Research in Science Teaching, 2015. **52**(6): p. 735-762.

201. Mueller, C.M. and C.S. Dweck, *Praise for intelligence can undermine children's motivation and performance.* Journal of personality and social psychology, 1998. **75**(1): p. 33.

202. Pollack, E., *The Only Woman in the Room: Why Science is Still a Boys' Club*. 2015: Beacon Press.

203. Xie, Y. and K.A. Shauman, *Women in Science*. 2003: Harvard University Press.

204. Harber, K.D., et al., *Students' race and teachers' social support affect the positive feedback bias in public schools.* Journal of Educational Psychology, 2012. **104**(4): p. 1149.

205. Godwin, A., et al. *Understanding engineering identity through structural equation modeling*. in *2013 IEEE Frontiers in Education Conference (FIE)*. 2013. IEEE.

206. Hyater-Adams, S., et al. *Understanding connections between physics and racial identities through recognition and relational resources*. in *2016 Physics Education Research Conference Proceedings*. 2016.

207. Roeser, R.W., C. Midgley, and T.C. Urdan, *Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging.* Journal of educational psychology, 1996. **88**(3): p. 408.

208. Strayhorn, T.L., *College Students' Sense of Belonging: A Key to Educational Success for All Students*. 2012: Routledge.

209. Hausmann, L.R., 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.* Research in higher education, 2007. **48**(7): p. 803-839.

210. Pittman, L.D. and A. Richmond, *Academic and psychological functioning in late adolescence: The importance of school belonging.* The Journal of Experimental Education, 2007. **75**(4): p. 270-290.

211. Wilson, D., et al., *Belonging and academic engagement among undergraduate STEM students: A multi-institutional study.* Research in Higher Education, 2015. **56**(7): p. 750-776.

212. Trujillo, G. and K.D. Tanner, *Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity.* CBE—Life Sciences Education, 2014. **13**(1): p. 6-15.

213. Ong, M., J.M. Smith, and L.T. Ko, *Counterspaces for women of color in STEM higher education: Marginal and central spaces for persistence and success.* Journal of Research in Science Teaching, 2018. **55**(2): p. 206-245.

214. Lewis, K.L., et al., *Fitting in to move forward: Belonging, gender, and persistence in the physical sciences, technology, engineering, and mathematics (pSTEM).* Psychology of Women Quarterly, 2017. **41**(4): p. 420-436.

215. London, B., et al., *The influences of perceived identity compatibility and social support on women in nontraditional fields during the college transition.* Basic and Applied Social Psychology, 2011. **33**(4): p. 304-321.

216. Good, C., A. Rattan, and C.S. Dweck, *Why do women opt out? Sense of belonging and women's representation in mathematics.* Journal of personality and social psychology, 2012. **102**(4): p. 700.

217. Stout, J.G., et al. *How a gender gap in belonging contributes to the gender gap in physics participation*. in *AIP conference proceedings*. 2013. American Institute of Physics.

218. Smith, J.L., et al., *When trying hard isn't natural: Women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns.* Personality and Social Psychology Bulletin, 2013. **39**(2): p. 131-143.

219. Aguillon, S.M., et al., *Gender differences in student participation in an active-learning classroom.* CBE—Life Sciences Education, 2020. **19**(2): p. 1-10.

220. Zumbrunn, S., et al., *Support, belonging, motivation, and engagement in the college classroom: A mixed method study.* Instructional Science, 2014. **42**(5): p. 661-684.

221. Solomon, D., et al., *Teacher practices associated with students' sense of the classroom as a community.* Social Psychology of Education, 1996. **1**(3): p. 235-267.

222. Cohen, G.L. and D.K. Sherman, *The psychology of change: Self-affirmation and social psychological intervention.* Annual review of psychology, 2014. **65**: p. 333-371.

223. Cohen, G.L., et al., *Reducing the racial achievement gap: A social-psychological intervention.* science, 2006. **313**(5791): p. 1307-1310.

224. Bowling, K.G., et al., *Impacts of virtual clinical trials simulations on science knowledge and attitudes.* The Electronic Journal for Research in Science & Mathematics Education, 2017. **21**(7).

225. Sari, U.S., M. Alıcı, and Ö. Faruk, *The effect of STEM instruction on attitude, career perception and career interest in a problem-based learning environment and student opinions.* The Electronic Journal for Research in Science & Mathematics Education, 2018. **22**(1).

226. Swarat, S., *What makes a topic interesting? A conceptual and methodological exploration of the underlying dimensions of topic interest.* The Electronic Journal for Research in Science & Mathematics Education, 2008.

227. Yang, L.-h., *Interpersonal relationships and the development of student interest in science.* The Electronic Journal for Research in Science & Mathematics Education, 2016. **20**(1).

228. Hanson, L.M., J.F. Westerlund, and P.W. Vaughn, *Gender differences in student attitudes towards science in secondary school classrooms with and without NSF GK-12 resident scientists.* The Electronic Journal for Research in Science & Mathematics Education, 2020. **24**(4): p. 50-68.

229. Louis, R.A. and J.M. Mistele, *The differences in scores and self-efficacy by student gender in mathematics and science.* International Journal of Science and Mathematics Education, 2012. **10**(5): p. 1163-1190.

230. Hazari, Z., et al., *Factors that affect the physical science career interest of female students: Testing five common hypotheses.* Physical Review Special Topics-Physics Education Research, 2013. **9**(2): p. 020115.

231. George, B.T., S.W. Watson, and M.L. Peters, *The impact of participating in a STEM academy on girls' STEM attitudes and self-efficacy.* The Electronic Journal for Research in Science & Mathematics Education, 2020. **24**(4): p. 22-49.

232. Wheeler, E.R. and S.M. Wischusen, *Developing self-regulation and self-efficacy: A cognitive mechanism for success of biology boot camps.* The Electronic Journal for Research in Science & Mathematics Education, 2014. **18**(1).

233. Tims, H.E., et al., *Cyber value and interest development: Assessment of a STEM career intervention for high school students.* The Electronic Journal for Research in Science & Mathematics Education, 2014. **18**(1).

234. Potvin, G. and Z. Hazari. *The development and measurement of identity across the physical sciences*. in *Proceedings of the 2013 Physics Education Research Conference*. 2013. Portland, OR.

235. Flowers III, A.M. and R. Banda, *Cultivating science identity through sources of self-efficacy.* Journal for Multicultural Education, 2016. **10**(3): p. 405-417.

236. Adams, W.K., et al., *New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey.* Physical Review Special Topics - Physics Education Research, 2006. **2**(1): p. 010101.

237. Shanahan, M.-C. and M. Nieswandt, *Creative Activities and Their Influence on Identification in Science: Three Case Studies.* Journal of Elementary Science Education, 2009. **21**(3): p. 63-79.

238. Nokes-Malach, T.J., et al. *Prior preparation and motivational characteristics mediate relations between gender and learning outcomes in introductory physics*. in *Proceedings of the 2018 Physics Education Research Conference*. 2018. Washington, DC.

239. Smith, J.L., C. Sansone, and P.H. White, *The stereotyped task engagement process: The role of interest and achievement motivation.* Journal of Educational Psychology, 2007. **99**(1): p. 99-114.

240. Hewitt, N.M. and E. Seymour, *A long, discouraging climb.* ASEE Prism, 1992. **1**(6): p. 24-28.

241. Ivie, R. and K. Stowe, *Women in Physics, 2000. AIP Report.* 2000.

242. Beilock, S.L., R.J. Rydell, and A.R. McConnell, *Stereotype threat and working memory: Mechanisms, alleviation, and spillover.* Journal of Experimental Psychology: General, 2007. **136**(2): p. 256.

243. Christy, K.R. and J. Fox, *Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance.* Computers & Education, 2014. **78**: p. 66-77.

244. Rydell, R.J., et al., *Stereotype threat prevents perceptual learning.* Proceedings of the National Academy of Sciences, 2010. **107**(32): p. 14042-14047.

245. Tobias, S., *They're Not Dumb, They're Different*. 1990: Research Corporation Tucson, AZ.

246. Eddy, S.L., S.E. Brownell, and M.P. Wenderoth, *Gender gaps in achievement and participation in multiple introductory biology classrooms.* CBE—Life Sciences Education, 2014. **13**(3).

247. Hidi, S. and K.A. Renninger, *The four-phase model of interest development.* Educational Psychologist, 2006. **41**(2): p. 111-127.

248. Bryk, A.S. and S.W. Raudenbush, *Hierarchical linear models: Applications and data analysis methods*. 1992: Sage Publications, Inc.

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

250. Hazari, Z., P.M. Sadler, and G. Sonnert, *The science identity of college students: Exploring the intersection of gender, race, and ethnicity.* Journal of College Science Teaching, 2013. **42**(5): p. 82-91.

251. Cumming, G. and S. Finch, *Inference by eye: confidence intervals and how to read pictures of data.* American Psychologist, 2005. **60**(2): p. 170.

252. Lock, R. and Z. Hazari, *Discussing underrepresentation as a means to facilitating female students' physics identity development.* Physical Review Physics Education Research, 2016. **12**: p. 020101.

253. Rockinson-Szapkiw, A., J.L. Wendt, and J.S. Stephen, *The efficacy of a blended peer mentoring experience for racial and ethnic minority women in STEM pilot study: Academic, professional, and psychosocial outcomes for mentors and mentees.* Journal for STEM Education Research, 2021. **4**(2): p. 173-193.

254. Hazari, Z., et al., *Examining the relationship between informal science experiences and physics identity: Unrealized possibilities.* Physical Review Physics Education Research, 2022. **18**(1): p. 010107.

255. Le, P.T., et al., *Investigating undergraduate biology students' science identity production.* CBE—Life Sciences Education, 2019. **18**(4).

256. Maton, K.I., et al., *Outcomes and processes in the Meyerhoff Scholars Program: STEM PhD completion, sense of community, perceived program benefit, science identity, and research self-efficacy.* CBE—Life Sciences Education, 2016. **15**(3): p. ar48.

257. Camacho, T.C., et al., *Science identity among latinx students in the biomedical sciences: The role of a critical race theory–informed undergraduate research experience.* CBE—Life Sciences Education, 2021. **20**(2).

258. Akoglu, H., *User's guide to correlation coefficients.* Turkish Journal of Emergency Medicine, 2018. **18**(3): p. 91-93.

259. *R: A Language and Environment for Statistical Computing*, ed. R.F.f.S. Computing. 2019, Vienna, Austria: R Foundation.

260. Buck, G.A., *Teaching discourses: Science teachers' responses to the voices of adolescent girls.* Learning Environments Research, 2002. **5**(1): p. 29-50.

261. Salmi, H. and H. Thuneberg, *The role of self-determination in informal and formal science learning contexts.* Learning Environments Research, 2019. **22**(1): p. 43-63.

262. Hughes, R.M., B. Nzekwe, and K.J. Molyneaux, *The single sex debate for girls in science: A comparison between two informal science programs on middle school students' STEM identity formation.* Research in Science Education, 2013. **43**(5): p. 1979-2007.

263. Andersen, H.M., L.B. Krogh, and E. Lykkegaard, *Identity matching to scientists: Differences that make a difference?* Research in Science Education, 2014. **44**(3): p. 439-460.

264. Vincent-Ruz, P. and C.D. Schunn, *Identity complexes and science identity in early secondary: mono-topical or in combination with other topical identities.* Research in Science Education, 2019: p. 1-22.

265. Dou, R. and H. Cian, *The Relevance of Childhood Science Talk as a Proxy for College Students' STEM Identity at a Hispanic Serving Institution.* Research in Science Education, 2020: p. 1-13.

266. Chen, S. and B. Wei, *Development and Validation of an Instrument to Measure High School Students' Science Identity in Science Learning.* Research in Science Education, 2020. **50**(2): p. 1-16.

267. Krogh, L.B. and H.M. Andersen, *"Actually, I may be clever enough to do it". Using identity as a lens to investigate students' trajectories towards science and university.* Research in Science Education, 2013. **43**(2): p. 711-731.

268. Ladson-Billings, G. and W.F. Tate IV, *Toward a critical race theory of education*, in *Critical Race Theory in Education*, A.D. Dixson, C.K.R. Anderson, and J.K. Donnor, Editors. 1995.

269. Rios-Aguilar, C., *The changing context of critical quantitative inquiry.* New directions for institutional research, 2014. **2013**(158): p. 95-107.

270. Yosso, T.J., *Whose culture has capital? A critical race theory discussion of community cultural wealth.* Race ethnicity and education, 2005. **8**(1): p. 69-91.

271. Larry, T. and J.L. Wendt, *Predictive relationship between gender, ethnicity, science self-efficacy, teacher interpersonal behaviors, and science achievement of students in a diverse urban high school.* Learning Environments Research, 2021. **25**: p. 1-17.

272. Serrano Corkin, D.M., S.F. Lindt, and P.S. Williams, *Effects of positive college classroom motivational environments on procrastination and achievement.* Learning Environments Research, 2021. **24**(2): p. 299-313.

273. Bandura, A., *Social cognitive theory of self-regulation.* Organizational Behavior and Human Decision Processes, 1991. **50**(2): p. 248-287.

274. Meltzer, D.E. and K. Manivannan, *Transforming the lecture-hall environment: The fully interactive physics lecture.* American Journal of Physics, 2002. **70**(6): p. 639-654.

275. Hammer, D., *Epistemological beliefs in introductory physics.* Cognition and Instruction, 1994. **12**(2): p. 151-183.

276. Singh, C., *Impact of peer interaction on conceptual test performance.* American Journal of Physics, 2005. **73**(5): p. 446-451.

277. Sayer, R., E. Marshman, and C. Singh. *The impact of peer interaction on the responses to clicker questions in an upper-level quantum mechanics course*. in *Physics Education Research Conference*. 2016. Sacramento, CA: AIP.

278. Good, M., et al., *Physics teaching assistants' views of different types of introductory problems: Challenge of perceiving the instructional benefits of context-rich and multiple-choice problems.* Physical Review Physics Education Research, 2018. **14**(2): p. 020120.

279. Johnson, A., *The mathematics of sex: How biology and society conspire to limit talented women and girls.* Science Education, 2012. **96**(5): p. 960-962.

280. Kellner, D., *Toward a critical theory of education.* Democracy & Nature, 2003. **9**(1): p. 51-64.

281. Tolbert, S., A. Schindel, and A.J. Rodriguez, *Relevance and relational responsibility in justice-oriented science education research.* Science Education, 2018. **102**(4): p. 796-819.

282. Estrada, M., A. Eroy-Reveles, and J. Matsui, *The influence of affirming kindness and community on broadening participation in STEM career pathways.* Social Issues and Policy Review, 2018. **12**(1): p. 258-297.

283. Green, A.M., B.R. Brand, and G.E. Glasson, *Applying actor–network theory to identify factors contributing to nonpersistence of African American students in STEM majors.* Science Education, 2019. **103**(2): p. 241-263.

284. Ketenci, T., A. Leroux, and M. Renken, *Beyond student factors: A study of the impact on STEM career attainment.* Journal for STEM Education Research, 2020. **3**(3): p. 368-386.

285. Han, J., T. Kelley, and J.G. Knowles, *Factors Influencing Student STEM learning: Self-efficacy and outcome expectancy, 21st century skills, and career awareness.* Journal for STEM Education Research, 2021. **4**(2): p. 117-137.

286. Kosiol, T., S. Rach, and S. Ufer, *(Which) mathematics interest is important for a successful transition to a university study program?* International Journal of Science and Mathematics Education, 2019. **17**(7): p. 1359-1380.

287. Mujtaba, T. and M.J. Reiss, *A survey of psychological, motivational, family and perceptions of physics education factors that explain 15-year-old students'aspirations to study physics in post-compulsory English schools.* International Journal of Science and Mathematics Education, 2014. **12**(2): p. 371-393.

288. Gao, J., *Sources of mathematics self-efficacy in Chinese students: a mixed-method study with Q-sorting procedure.* International Journal of Science and Mathematics Education, 2020. **18**(4): p. 713-732.

289. McKinney, J., M.-L. Chang, and D. Glassmeyer, *Why females choose STEM majors: Understanding the relationships between major, personality, interests, self-eficacy, and anxiety.* Journal for STEM Education Research, 2021. **4**(3): p. 278-300.

290. Uitto, A., *Interest, attitudes and self-efficacy beliefs explaining upper-secondary school students' orientation towards biology-related careers.* International Journal of Science and Mathematics Education, 2014. **12**(6): p. 1425-1444.

291. Crouch, C.H. and E. Mazur, *Peer Instruction: Ten years of experience and results.* American Journal of Physics, 2001. **69**(9): p. 970-977.

292. Li, Y. and C. Singh, *Maybe it's just me: The impact of perceived recognition by physics instructors on women's self-efficacy and interest.* submitted.

293. Center, N.R. *Science & Engineering Degree Attainment: 2006-2016*. 2017; Available from: https://nscresearchcenter.org/snapshotreport-science-and-engineering-degree-completion-by-gender/.

294. Kalender, Z.Y., et al. *Motivational characteristics of underrepresented ethnic and racial minority students in introductory physics courses*. in *Proceedings of the 2017 Physics Education Research Conference*. 2017. Cincinnati, OH.

295. *Science and Engineering Indicators 2022: The State of U.S. Science and Engineering*. 2022; NSB-2022-1]. Available from: https://ncses.nsf.gov/pubs/nsb20221.

296. Hewlett, S.A., M. Marshall, and L. Sherbin, *How diversity can drive innovation.* Harvard Business Review, 2013. **91**(12): p. 30-30.

297. Ellison, S.F. and W.P. Mullin, *Diversity, social goods provision, and performance in the firm.* Journal of Economics & Management Strategy, 2014. **23**(2): p. 465-481.

298. Sáiz-Manzanares, M.C., et al., *Systematic review on inclusive education, sustainability in engineering: An analysis with mixed methods and data mining techniques.* Sustainability, 2020. **12**(17): p. 6861.

299. Rizki, P.N.M., et al., *Promoting self-regulated learning for students in underdeveloped areas: The case of Indonesia Nationwide Online-Learning Program.* Sustainability, 2022. **14**(7): p. 4075.

300. Evans, T.L., *Competencies and pedagogies for sustainability education: A roadmap for sustainability studies program development in colleges and universities.* Sustainability, 2019. **11**(19): p. 5526.

301. Bian, L., et al., *Messages about brilliance undermine women's interest in educational and professional opportunities.* Journal of Experimental Social Psychology, 2018. **76**: p. 404-420.

302. Sayer, R., E. Marshman, and C. Singh, *Case study evaluating Just-In-Time Teaching and peer instruction using clickers in a quantum mechanics course.* Physical Review Physics Education Research, 2016. **12**(2): p. 020133.

303. Cwik, S. and C. Singh, *Not feeling recognized as a physics person by instructors and teaching assistants is correlated with female students' lower grades.* Physical Review Physics Education Research, 2022. **18**(1): p. 010138.

304. Cwik, S. and C. Singh, *Students' sense of belonging in introductory physics course for bioscience majors predicts their grade.* Physical Review Physics Education Research, 2022. **18**(1): p. 010139.

305. E Akins, E., et al., *Sustainability education and organizational change: A critical case study of barriers and change drivers at a higher education institution.* Sustainability, 2019. **11**(2): p. 501.

306. Birt, J.A., et al., *Fostering agency to overcome barriers in college science teaching: Going against the grain to enact reform‐based ideas.* Science Education, 2019. **103**(4): p. 770-798.

307. Sonnert, G. 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, 2012. **83**(1): p. 73-101.

308. Wigfield, A. and J.S. Eccles, *Expectancy–value theory of achievement motivation.* Contemporary Educational Psychology, 2000. **25**(1): p. 68-81.

309. Zohar, A. and A. Gershikov, *Gender and performance in mathematical tasks: Does the context make a difference?* International Journal of Science and Mathematics Education, 2008. **6**(4): p. 677-693.

310. Meeuwisse, M., S.E. Severiens, and M.P. Born, *Learning environment, interaction, sense of belonging and study success in ethnically diverse student groups.* Research in Higher Education, 2010. **51**(6): p. 528-545.