Investigating students' self-identified and reflected appraisal of femininity, masculinity, and androgyny in introductory physics courses

Yangqiuting Li [1] and Eric Burkholder [2]

[1] *Department of Physics, Oregon State University, Corvallis, Oregon 97331, USA*
[2] *Department of Physics, Auburn University, Auburn, Alabama 36849, USA*

(Received 6 September 2023; accepted 16 January 2024; published 27 February 2024)

In the field of physics education research, numerous studies have been dedicated to investigating the relationship between gender identity and physics learning. However, these studies have predominantly employed binary gender measurement methods, which may limit the range of research questions that can be explored and impede the discovery of crucial insights. In this study, we adapted gradational measures from prior research to investigate students' self-identified femininity, masculinity, and androgyny, as well as their reflected appraisal of femininity, masculinity, and androgyny (i.e., perceptions of how others perceive them) in both algebra-based and calculus-based introductory physics courses. The use of gradational measures revealed significant variation in students' self-identified femininity, masculinity, and androgyny within the binary categories of women and men, providing new insights into gender dynamics in physics. We found that self-identified women in the calculus-based courses, where they are underrepresented, tend to perceive themselves as more masculine and less feminine than how they believe others perceive them. Similarly, students of color are also more likely than White students to perceive themselves as more masculine than they believe others perceive them. Using structural equation modeling, we found that students' gender stigma consciousness plays an important role in mediating the effects of identifying as women and students of color on the observed discrepancies. Additionally, we found that women also exhibit a tendency to perceive themselves as more androgynous than they believe others perceive them in both algebra-based and calculus-based physics courses, and this phenomenon is also related to gender stigma consciousness. Moreover, our analyses revealed that students in the calculus-based courses tend to have a higher level of gender stigma consciousness even after controlling for gender and race. Our findings underscore the potential of gradational gender measurements in deepening our understanding of gender-related issues in physics education, shedding light on the complex interplay between students' gender identity, perceptions from others, and their educational experiences in the field.

DOI: 10.1103/PhysRevPhysEducRes.20.010110

I. INTRODUCTION

Physics has long been a field predominantly occupied by men. For instance, in 2020, women earned approximately 65% of all bachelor's degrees in the U.S., while only 25% of physics undergraduate degrees were earned by women [1,2]. The underrepresentation of women persists in higher levels of academia, where they earned 21% of doctoral degrees in physics, 18% of postdoctoral positions, and constituted 14% of faculty members in the U.S. [2,3]. In the field of physics education research, numerous studies have explored the challenges faced by women in physics and have sought to promote inclusiveness, equity, and diversity within the field [4–15]. One important objective of these studies is to bridge the gender gaps disadvantaging women observed in various areas, such as students' academic performance and motivational beliefs [6,10,14,16–23]. Researchers have investigated diverse potential factors contributing to these gender gaps, such as stereotype threat [12,24], lack of role models and recognition [4,25,26], and other systemic biases [27,28]. Furthermore, prior studies have explored a range of approaches to reduce the gender gaps, such as implementing psychological interventions [19,29], adopting active learning strategies [16,30], and revising assessment methods [31,32]. These endeavors aim to create more inclusive learning environments and foster equality and diversity in physics.

While previous studies have provided valuable insights into the gender related issues in physics education, most of them have relied on binary gender measures [4,33–39]. Considering that gender is a social construct that exists along a spectrum rather than a rigid binary [40,41], the use of binary measures may have inadvertently constrained the

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

scope of research questions and limited the interpretation of findings in these previous studies. To contribute to a more comprehensive understanding of gender related issues in physics education, we employed gradational gender measures to investigate students' femininity, masculinity, and androgyny in introductory physics courses; such measures ask students to report their gender identity and expression along a 7-point Likert scale. Specifically, we compared students' response to the gradational gender measure and their response to the traditional binary gender measure, and we also examined the relationship between students' self-identified femininity, masculinity, and androgyny and their perception of how others perceive their femininity, masculinity, and androgyny. This study represents our initial step in exploring gender related issues in physics education using a more nuanced gender measurement framework. By embracing this framework, we aim to uncover novel perspectives and valuable findings that may have remained unexplored in previous studies.

In the next section of this paper (Sec. II), we will provide a review of the background of this study, including a review of prior studies on gender related issues in physics education and a discussion of the potential limitations associated with the use of a binary gender measurement. In Sec. III, we will introduce the theoretical framework employed in this study and the gradational gender measure utilized. Before delving into these sections, it is crucial to clarify the terminology used in this study and the reviewed studies. The following definitions will be used:

- *Gender* is a social construct that encompasses the roles, behaviors, expectations, and identities associated with femininity or masculinity within a particular society or culture. Gender exists along a spectrum rather than being binary and can vary across cultures and evolve over time.
- *Gender binary* is the classification of gender into two distinct and opposite forms of masculine and feminine, typically based on societal norms or cultural beliefs. In many cultures, a binary gender system is used, categorizing individuals as either men or women [42–44].
- *Femininity* is a set of attributes, behaviors, and roles generally associated with women and girls. Femininity can be understood as socially constructed.
- *Masculinity* is a set of attributes, behaviors, and roles generally associated with men and boys. Masculinity can be understood as socially constructed.
- *Androgyny* describes an individual who possesses similar (high) levels of femininity and masculinity [45,46].
- *Self-identified femininity, masculinity, or androgyny* refers to an individual's self-perception of their own degree of *femininity, masculinity, or androgyny.*
- *Reflected appraisal of femininity or masculinity* refers to how individuals perceive their *femininity, masculinity, or androgyny* to be viewed by others [47,48].

II. BACKGROUND

Numerous studies in the field of physics education have investigated the relationship between gender and physics learning. For example, some studies have shown women often reported a lower level of physics motivational beliefs, such as physics self-efficacy and identity [4,11,33,34,49,50]. Additionally, other studies have revealed gender differences in students' performance on some physics concept inventories [35–39]. Previous studies have demonstrated that these gender gaps in physics are closely related to lack of inclusiveness in physics learning environment, lack of women role models, and the pervasive societal stereotypes and biases about who belongs to and can succeed in physics [51–54]. These stereotypes can negatively influence women's learning experiences, lower their physics identity, and lead to withdrawal from physics [55–57]. Considerable research efforts have been dedicated to reducing gender gaps and supporting women in physics by examining and intervening in various factors, such as students' academic preparation, teaching methods, inclusiveness of the learning environment, stereotype threats, and assessment methods [6,19,58–63].

These prior studies offer valuable insights into the gender disparities in physics and make significant contributions to enhancing inclusion and equity in the field. However, we note that most of prior studies have relied solely on the traditional binary gender measure, which may have limited the scope of research questions that could be asked and potentially restricted the findings that could be obtained. For example, it is not known how a person who identifies as a woman, but may feel more masculine, perceives the learning environment compared to a woman who identifies as more feminine. In the following sections, we will discuss some potential limitations associated with the use of a binary gender measure and review several nonbinary gender measures proposed by prior studies.

A. Limitations of binary gender measurement

Prior studies in psychological and educational research have explored the potential limitations of using a binary gender measure [64–68]. These studies highlight several key issues that arise from relying solely on a binary framework. First, the binary gender measure cannot account for individuals who identify outside of the traditional men or women categories. Nonbinary, gender queer, and other gender identities are disregarded or marginalized when using a binary framework, resulting in the erasure of their experiences and perspectives in research [64–66]. Second, relying solely on a binary gender measure may perpetuate and reinforce societal norms and stereotypes that emphasize inherent and fixed differences between gender categories [67], neglecting the influence of social and cultural factors on the construction of gender roles and expectations [67]. By overlooking the fluid and variable nature of gender, the binary gender measure cannot accommodate individuals whose gender identities change cross contexts or evolve over time, thereby misrepresenting their experiences [68]. Furthermore, in the context of physics education, another concern with using the binary gender measure, as pointed out by Traxler *et al.*, is that when a disparity is observed between men and women, the focus tends be on how to help women attain the same level of performance as men. This approach implicitly assumes that the men's experience serves as the ideal standard and overlooks the need to challenge and reshape the culture that privileges specific groups.

To address these limitations, researchers emphasize the significance of adopting more inclusive approaches to understand and measure gender. This involves acknowledging the fluidity and variability of gender experiences [66,68], incorporating alternative gender measurements that extend beyond the binary framework [69,70], and adopting intersectional perspectives that consider the interactions between gender and other social identities [71]. In the next section, we will review several nonbinary gender measures proposed by prior studies.

B. Nonbinary gender measurement

Previous studies have proposed various alternative measures to capture gender diversity beyond the traditional binary categorization. One common alternative measure is to include responses beyond the binary categories of women and men, such as transgender, gender queer, and "a gender not listed here" [72–74]. However, simply adding more categories alone cannot fully address the complex challenges of representing population diversity. Analysts may face difficulties when dealing with a small number of respondents in specific gender categories, which may lead to the exclusion of certain populations from analysis or aggregate them under a single umbrella category, such as transgender [75]. In addition, it is important to note that nontraditional gender practices are not limited to transgender people [76], so relying solely on a categorical measure, regardless of the number of options provided, cannot adequately capture the overlap between gender categories.

Compared to categorical gender measures, gradational gender measures are suggested to offer more analytic potential. Early psychological studies utilizing gradational measures scored femininity and masculinity along a single, bipolar scale [77,78]. In the 1970s, the assumption that femininity and masculinity were mutually exclusive and opposite started to be questioned [79], leading to a shift in later studies that treated masculinity and femininity as separate entities. Moreover, the androgyny construct was first formulated and introduced by Sandra Bem to describe an individual who possesses similarly high levels of femininity and masculinity. Bem also developed a widely used inventory, the Bem Sex Role Inventory (BSRI), to measure femininity, masculinity, and androgyny. The BSRI consisted of a 60-item index assigning feminine, masculine, and androgynous scores based on gendered trait ratings [80]. Later, this inventory was abbreviated to the 30-item BSRI. However, it is important to note that these measures are based on identification with stereotypically feminine and masculine traits [80]. For example, if a person reports on the BSRI that they are "often" or "always or almost always" gentle or compassionate, their femininity score increases, while reporting that one is "often" or "always or almost always" assertive or analytical increases a respondent's masculinity score. Therefore, the resulting scores from these measurements may not necessarily reflect a person's gendered sense of self but rather the extent to which they conform to a set of stereotypes. Additionally, the length of these measures can pose challenges when incorporating them into large-scale surveys, which are important sources of data for social science research.

To overcome the limitations of the aforementioned measurements, a recent sociological study introduced a gradational measure that allows individuals to rate their own level of femininity, masculinity, or androgyny on a 7-point Likert scale [69]. This approach enables individuals to consider the various factors that contribute to their gendered sense of self when providing their responses. By giving respondents control over their gender identification instead of imposing rigid criteria, this measure recognizes that people construct complex gender identities and modes of expression as they navigate a system of gendered expectations and institutions [69,81,82]. In addition to the measure of self-identified femininity, masculinity, or androgyny, this recent study also proposed a measure of reflected appraisal of femininity, masculinity, or androgyny, which is how individuals perceive their femininity, masculinity, or androgyny to be viewed by others [47,48]. In our study presented here, we adapted these measures to investigate students' self-identified and reflected appraisals of femininity, masculinity, and androgyny in introductory physics courses.

III. THEORETICAL FRAMEWORK

A. The theory of gender performativity

Our study is grounded in the theory of gender performativity, which was developed by feminist philosopher Judith Butler and challenges the conventional understanding of gender as a fixed and inherent characteristic [83]. According to Butler, gender is not an individual's essential nature, but rather a social and cultural construct that is created and maintained through repeated acts and performances. Butler argues that gender is performative, meaning that it is constituted through a series of actions, gestures, and speech acts. These performances are not simply individual choices or expressions of an underlying essence; instead, they are shaped by societal norms, expectations, and discourses. Gender, therefore, is an ongoing process of "doing" and "repeating" certain behaviors and expressions that are recognized as masculine or feminine within a specific cultural context.

The theory of gender performativity challenges the notion of a stable and binary gender system, suggesting that gender is fluid and constantly being negotiated and constructed. This perspective sheds light on how gender norms and expectations are imposed upon individuals, limiting their freedom to express themselves outside of these norms. By deconstructing the binary gender system, the theory of gender performativity provides space for embracing gender diversity and nonconformity.

In addition to the performative aspects of gender, prior studies have also emphasized the interactive nature of gender. Individuals "do" their gender by enacting patterns of behavior that are socially understood to be feminine or masculine, and their gender is simultaneously "determined" by others who perceive and interpret those enactments [84]. Research has revealed that while there is a connection between how individuals perform their gender and how others perceive it, discrepancies can arise between individuals' self-perception of their gender and how others interpret it [85], and these discrepancies have been shown to associate with worse health outcomes among both cisgender and transgender people [70,85]. While some studies have compared individuals' self-identified femininity, masculinity, or androgyny with their perceptions of how others see them [69,70], to our knowledge, no prior studies have investigated the discrepancies between individuals' self-identified and reflected appraisal of femininity, masculinity, or androgyny and how these discrepancies might correlate with other characteristics of the individual. In this study, we will focus on this issue in the context of college introductory physics courses.

B. Potential influential factors on self-identified and reflected appraisal of femininity, masculinity, and androgyny

Based on the theory of gender performativity, which emphasizes the socially constructed nature of gender, individuals' sense of femininity, masculinity, and androgyny can be influenced by the environment and context they are in. Previous studies have provided evidence supporting this notion. For instance, a study demonstrated that engaging in care work resulted in all participants perceiving themselves as more feminine. Similarly, research has shown that women who engage in farm work tend to perceive themselves as more masculine [86,87]. Additionally, it has been shown that societal norms and ideals can also influence individuals' perception of their own femininity and masculinity. For example, in one study, women tend to rate themselves as more masculine when asked to consider society's ideal man and woman [88]. Given that androgyny is a combination of high levels of both femininity and masculinity, the factors influencing masculinity and femininity could also influence androgyny.

Considering that physics is a male-dominated field with pervasive stereotypes and biases about who belong and can excel in physics, students' self-identified femininity, masculinity, or androgyny may be influenced by the physics culture. However, to our knowledge, there is no prior quantitative research exploring students' femininity, masculinity, or androgyny using gradational measures in introductory physics courses. Furthermore, while previous research has shown that individuals' self-identified gender may differ from how others perceive their gender, which has been linked to negative health outcomes [70,85], very few studies have explored the alignment between students' self-identified femininity, masculinity, or androgyny and their perception of how others view their femininity, masculinity, or androgyny (i.e., reflected appraisal of femininity, masculinity, or androgyny). Therefore, our study aims to fill these knowledge gaps by quantitatively investigating students' femininity, masculinity, and androgyny in introductory physics courses.

In addition to investigating students' self-identified and reflected appraisal of femininity, masculinity, and androgyny, our study also aims to explore potential factors that can influence these constructs. Previous studies have indicated that women in science and technology often distance themselves from traditional femininity to fit into male-dominated disciplines [89,90]. Therefore, students' sense of belonging may play a role in shaping their self-identified and reflected appraisal of femininity, masculinity, and androgyny. In addition, gender stigma consciousness, which refers to individuals' awareness and recognition of the stigmatization and negative stereotypes associated with gender, has been found to positively correlate with identifying as women after being informed about gender differences in math performance [91]. Thus, gender stigma consciousness may also relate to students' self-identified and reflected appraisal of femininity, masculinity, and androgyny. In addition, prior studies have shown that the gender composition of a professional or academic environment can also impact participants' experiences and perceptions [92,93]. For example, Murphy and colleagues found that when women were exposed to a conference video where men far outnumbered women, they reported greater stress, a diminished sense of belonging, and lower interest in the conference compared to women who watched a video with a balanced gender ratio [92]. Considering that women are underrepresented in calculus-based introductory physics courses while being overrepresented in algebra-based introductory physics courses, we also include the course type factor in our analyses of students' self-identified and reflected appraisal of femininity, masculinity, and androgyny.

IV. THE PRESENT STUDY

In this study, we investigated students' self-identified and reflected appraisal of femininity, masculinity, and androgyny in introductory physics courses. Specifically, we examined the relationships between students' responses to the gradational measure of femininity, masculinity, or androgyny and their responses to the binary gender measure. Additionally, we compared students' self-identified femininity, masculinity, or androgyny and their reflected appraisal of femininity, masculinity, or androgyny. In instances where discrepancies were observed between students' self-identified and reflected appraisal femininity, masculinity, or androgyny, we delved into potential factors contributing to these differences. Moreover, we compared the results of the analyses in the calculus-based and algebra-based introductory physics courses. In summary, our study aimed to answer the following research questions:

RQ1. To what extent do individuals self-identified as women and men (on a binary scale) in introductory physics courses perceive their femininity, masculinity, and androgyny on a gradational scale?

RQ2. What are the correlations between students' responses to the binary gender measure and the gradational measures of self-identified femininity, masculinity, and androgyny? Additionally, what are the correlations among students' self-identified femininity, masculinity, and androgyny?

RQ3. Are there discrepancies between students' self-identified and reflected appraisal of femininity, masculinity, or androgyny? If so, what factors may contribute to these differences?

RQ4. How do the answers to the above research questions differ in the calculus-based and algebra-based introductory physics courses?

V. METHODOLOGY

A. Participants and data collection

The data used in this study were collected from college level algebra-based and calculus-based introductory physics courses at a large public research university in the US over two consecutive school semesters (Fall 2022 and Spring 2023). The algebra-based courses are generally taken by biomedical science majors, while the calculus-based courses are generally taken by engineering and physical science majors. Both the algebra-based and calculus-based sequence consist of two physics courses: physics 1, focusing on mechanics, and physics 2, covering electricity, magnetism, and optics. Physics 1 and physics 2 were offered in both the fall and spring semesters. The courses are four credit-hour courses which include 2.5 h of lecture per week, and a 3-h combined recitation and laboratory led by a teaching assistant that meets once per week. The laboratory sections have 20–24 students working in groups of 3–4, whereas the lecture sections typically have 100–200 students. The recitation periods are typically used for brief weekly quizzes, and the laboratories are standard cookbook-type physics labs [94]. Some faculty include small amounts of group work in the lectures, but the courses are predominantly traditional.

Data collection took place during the first recitation class of each semester for each course. In the semesters studied, there were 889 students taking the algebra-based courses, and 760 (445 women and 315 men) of them participated in our study. There were 1541 students taking the calculus-based course in the semesters studied, and 1284 (279 women and 1005 men) of them participated in our study. Some reasons for nonparticipation include not giving consent to use their responses as research data, not attending the recitations when the survey was implemented, or adding the course after the survey was implemented (the add-drop period is the first few weeks of the course). The students surveyed in the Fall 2022 and Spring 2023 semesters are from two different cohorts. There were 28 students in the sample who repeated the course they took in the fall again in the spring, and 183 students who took physics 1 in the fall and physics 2 in the spring. We have kept the data of these students for both semesters in the analyses because we hypothesized that their survey responses might be related to the classroom environment (e.g., instructor, peers, etc.). Future studies will investigate how individuals may change their responses over time.

Students' demographic data were provided by the university. Most students in the calculus-based courses were 18–20 years old, and the average age was 19.0 years old with a standard deviation of 1.85. In the algebra-based course, students were primarily aged 18 to 21, with an average age of 19.8 years and a standard deviation of 1.59. The majority of students were White (80%), with the remainder identifying as Asian (5%), Hispanic (5%), African American (4%), Multiracial (3%), or Other (3%). Additionally, 12% of the students were first-generation college students. The demographics are relatively even across different lecture sections, with around 18%–25% self-identified women in the calculus-based courses, and around 60%–65% self-identified women in the algebra-based courses. Similarly, there are around 15%–25% non-White students across different sections in both the calculus-based and algebra-based courses. The binary gender information was collected by the university, which did not offer any alternative options to man or woman at the time of data collection.

B. Measurements

1. Gradational measures of femininity, masculinity, and androgyny

In this study, we adapted the gradational measures developed in the previous study [69], as discussed in the introduction section, to assess students' self-identified and reflected appraisal of femininity, masculinity, and androgyny. The survey items and the corresponding Likert scales are shown in Fig. 1. The first question, "In general, how do you see yourself?" asks students to rate their own self-identified femininity, masculinity, and androgyny on a 7-point Likert scale, ranging from "not at all" (coded as 0) to "very" (coded as 6). The second question, "In general, how do most people see you?" prompts students to evaluate how they believe others perceive their femininity, masculinity, and androgyny, also using the 7-point Likert scale from not at all to very. These survey questions pertaining to self-identified and reflected appraisal of femininity, masculinity, and androgyny were placed after the psychological questions concerning students' experiences and attitudes toward physics. However, we did not provide a specific prompt asking students to answer the gender-related questions in the context of the physics classroom.

2. *Psychological survey*

As discussed earlier in the theoretical framework section, we intended to investigate the roles of psychological constructs such as gender stigma consciousness and the sense of belonging in students' self-identified and reflected appraisal of femininity, masculinity, and androgyny. To measure these psychological constructs, we employed a 7-point Likert scale survey consisting of validated items adapted from prior studies [95,96]. The psychological survey was administered alongside the gradational femininity, masculinity, or androgyny scales during the first recitation class of each semester.

We revalidated the survey items using methods such as exploratory factor analysis, confirmatory factor analysis, and Cronbach's alpha [97,98]. Table I presents the survey items for each psychological construct and the corresponding Cronbach's alpha and CFA factor loadings. Table I shows that both constructs have Cronbach's alpha values higher than 0.8, indicating a high level of internal consistency among the survey items [97]. In addition, all factor loadings are higher than 0.4 and most of them are higher than 0.6, indicating that our constructs extract sufficient variance from the items [99]. A student's score for each construct is the average score of all items within that construct, and the score range for both gender stigma consciousness and sense of belonging is 1–7.

C. Data analysis

In this study, we first examined the distribution of self-identified femininity, masculinity, and androgyny among students within each binary gender category. Specifically, we calculated the percentages of women identified in the university data who rated themselves at different levels of femininity, masculinity, and androgyny on the 7-point Likert scale, and we conducted the same analysis for students who identified as men in the university data. In addition, to further investigate the relationship between the binary gender measure and the gradational measures of self-identified femininity, masculinity, and androgyny, we calculated the correlations between students' responses to these measures.

Next, we compared students' self-identified femininity, masculinity, and androgyny with their reflected appraisal of femininity, masculinity, and androgyny. Specifically, we calculated the difference between students' self-identified

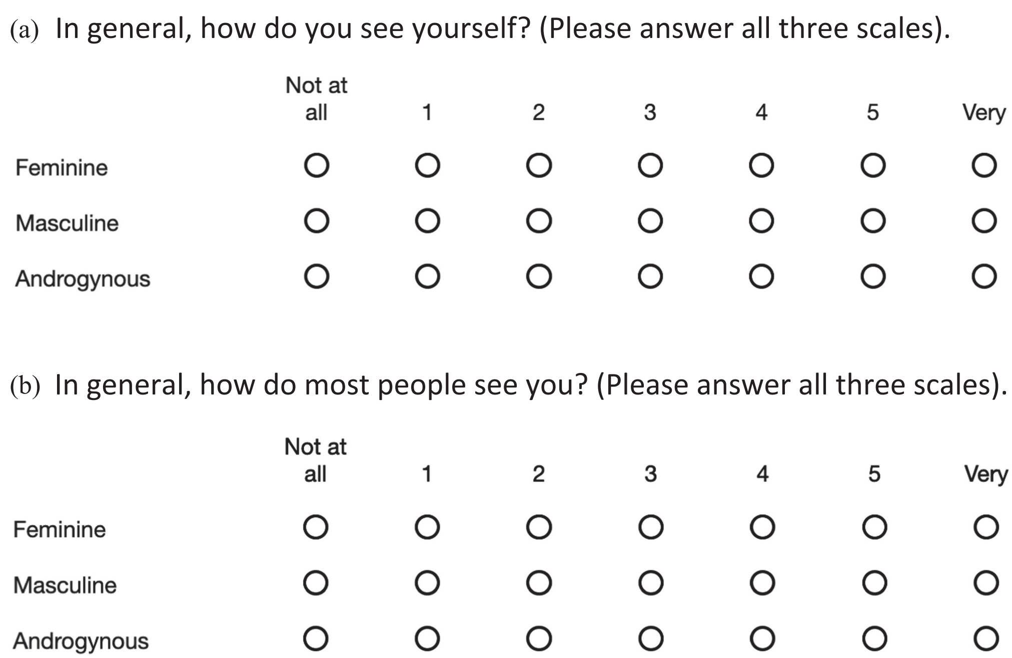

FIG. 1. Survey questions for (a) self-identified femininity, masculinity, or androgyny and (b) reflected appraisal of femininity, masculinity, or androgyny.

TABLE I. Survey items for gender stigma consciousness and sense of belonging and the corresponding Cronbach's alphas and CFA factor loadings (Lambda and p values of the significance test for each item loading). †The second and fifth belonging items were reverse coded to ensure that a higher score in these two items represents a higher sense of belonging.

Construct and item	Lambda	p value
Gender stigma consciousness (Cronbach's alpha = 0.92)		
My gender affects how my peers interact with me	0.931	<0.001
Most people judge me on the basis of my gender	0.946	<0.001
Physics sense of belonging (Cronbach's alpha = 0.85)		
I feel like I belong in this class.	0.790	<0.001
I feel like an outsider in this class†	0.710	<0.001
I feel comfortable in this class.	0.849	<0.001
I feel like I can be myself in this class.	0.651	<0.001
Sometimes I worry that I do not belong in this physics class†	0.700	<0.001
I am comfortable making a comment or asking a question during class discussion	0.468	<0.001

femininity, masculinity, or androgyny and their reflected appraisal of femininity, masculinity, or androgyny. For example, if a student rated their self-identified masculinity as 2 and their reflected appraisal of masculinity as 4, the difference would be coded as −2. Conversely, if the difference is positive, it means that the student perceives themselves as more masculine than how they believe others view them. To investigate whether the differences between students' self-identified and reflected appraisal of femininity, masculinity, or androgyny are statistically significant, we used the Wilcoxon signed-rank test. The Wilcoxon signed-rank test is a nonparametric statistical test commonly used to compare two matched samples when the normality assumption is not satisfied or when the data are ordinal [100]. Considering that women are underrepresented in the calculus-based courses and overrepresented in the algebra-based courses, we conducted the above analyses for these two course types separately and compared the results between them.

If statistically significant discrepancies between students' self-identified and reflected appraisal of femininity, masculinity, or androgyny were observed, we then explored potential factors that may contribute to these differences. Specifically, we considered demographic factors, such as gender on a binary scale (women vs men), race (non-White vs White), and first-generation status (first generation vs not first generation), as well as course-level factors, such as course type (calculus based vs algebra based) and semester (fall vs spring—corresponding to on-sequence or off-sequence course taking). Additionally, we investigated potential psychological factors, such as sense of belonging and gender stigma consciousness, which are previously discussed in the theoretical framework section.

We identified several factors that are either significantly correlated with the observed discrepancies or show the potential of associating with these discrepancies based on literature and our theoretical framework. Subsequently, we conducted logistic regression analyses to examine the effects of each factor on the likelihood of a discrepancy existing between self-identified and reflected appraisal of femininity, masculinity, or androgyny. In our logistic regression models, each psychological construct was represented by the mean score of the items comprising that construct, while the demographic and course level factors were coded as binary variables. For example, self-identified women were coded as 1 and men as 0, non-White students were coded as 1 and White students as 0, and calculus-based courses were coded as 1 and algebra-based courses as 0. In addition to logistic regression, we used structural equation modeling (SEM) to investigate the chains of relationships among the studied constructs. SEM includes two parts: confirmatory factor analysis (CFA) and path analysis [98]. As discussed in the measurements section, the confirmatory factor analysis results indicate that our constructs extract sufficient variance from the survey items, which allowed us to perform path analysis using these constructs. The path analysis part of SEM provided standardized regression coefficients (β) for the paths connecting each pair of constructs, offering a measure of the strength of these relationships.

VI. RESULTS

A. Distribution of self-identified femininity, masculinity, and androgyny within binary gender categories

In Table II, we present the percentages of men and women who indicated their levels of femininity and masculinity as "0 (not at all)" or "6 (very)". The data illustrates that only 34% of men identified themselves as purely masculine (i.e., femininity = 0 and masculinity = 6), and only 32% of women identified themselves as purely feminine (i.e., femininity = 6 and masculinity = 0). We also observed that 5% of respondents indicated an equal level of femininity and masculinity. Furthermore, among the respondents, 19 women perceived themselves as more masculine than feminine, and 6 men perceived themselves as more feminine than masculine.

TABLE II. Percentages of men and women who indicated their level of femininity and masculinity as 0 (not at all) or 6 (very), along with the results of two-sample proportion z tests to compare the percentages in the algebra-based and calculus-based courses. Cohen suggested that typically values of 0.2, 0.5, and 0.8 represent small, medium, and large effect sizes for two sample proportion z tests [101].

Binary measure	Gradational measure	Overall ($N = 2044$)	Algebra based ($N = 760$)	Calculus based ($N = 1284$)	Statistics	
					p value	Effect size
Men	Femininity = 0	62%	62%	62%	1.000	0.00
	Masculinity = 6	40%	41%	40%	0.656	0.02
	Femininity = 0 and Masculinity = 6	34%	33%	34%	0.644	0.02
Women	Femininity = 6	45%	49%	39%	<0.001	0.20
	Masculinity = 0	44%	49%	38%	<0.001	0.22
	Femininity = 6 and Masculinity = 0	32%	38%	27%	<0.001	0.24

When comparing the distribution of self-identified femininity and masculinity between algebra-based and calculus-based courses, we note that fewer women in calculus-based courses chose high levels of femininity or low levels of masculinity compared to algebra-based courses, whereas there is almost no difference in the percentages of men between algebra-based and calculus-based courses who indicated their level of femininity as not at all and masculinity as very. These findings are supported by the results of two-sample proportion z tests [101]. For example, as shown in Table II, 38% of women in algebra-based courses identified themselves as purely feminine (i.e., femininity = 6 and masculinity = 0), while only 27% of women in calculus-based courses did so. The two-sample

FIG. 2. Distribution of self-identified (a) femininity, (b) masculinity, and (c) androgyny of self-identified women and men.

TABLE III. Correlations between students' responses to the binary gender measure and the gradational measures of self-identified femininity, masculinity, and androgyny, along with the results of Fisher's z tests comparing these correlations in the algebra-based and calculus-based courses. p values for the correlations are indicated by *** for $p < 0.001$, ** for $0.001 \leq p < 0.01$, * for $0.01 \leq p < 0.05$, and ^{ns} for $p > 0.05$ (not statistically significant).

				Statistics	
Correlations	Overall	Algebra-based	Calculus-based	p value	Effect size
Gender ~ Femininity	0.87^{***}	0.89^{***}	0.84^{***}	<0.001	0.20
Gender ~ Masculinity	-0.84^{***}	-0.89^{***}	-0.76^{***}	<0.001	−0.43
Femininity ~ Masculinity	-0.85^{***}	-0.90^{***}	-0.77^{***}	<0.001	−0.45
Androgyny ~ Gender	0.05^{*}	-0.03^{ns}	0.11^{***}	0.001	−0.14
Androgyny ~ Femininity	0.09^{***}	-0.09^{*}	0.23^{***}	<0.001	−0.32
Androgyny ~ Masculinity	0.01^{ns}	0.18^{***}	-0.10^{**}	<0.001	0.28

proportion z test indicates a statistically significant difference between these two percentages ($p < 0.001$), with the effect size being 0.24.

To visually represent the data, Fig. 2 displays the distribution of women's and men's self-identified femininity, masculinity, and androgyny on the 7-point Likert scale. We note that respondents utilized the full range of the Likert scale to indicate their levels of femininity and masculinity, with overlap between the scale responses of women and men. In addition, Fig. 2 shows that around 25% of women and 22% of men indicated a nonzero level of androgyny.

B. Correlations between students' responses to the binary gender measure and the gradational measures of self-identified femininity, masculinity, and androgyny

In this section, we examined the correlations between students' responses to the binary gender measure (with women coded as 1 and men as 0) and the 7-point Likert scale measures of self-identified femininity, masculinity, and androgyny. As shown in the Table III, while there are correlations between students' responses to the binary gender and self-identified femininity and masculinity, the correlations are not 1 or −1, indicating that the Likert scale measures are not entirely equivalent to the binary measure. Moreover, the correlation between self-identified femininity and masculinity is not −1, suggesting that they are not mutually exclusive and opposite, which is consistent with prior studies [79,80]. Additionally, we note that the absolute correlations among the binary gender measurement and the gradational measurements of femininity and masculinity are slightly higher for students in algebra-based courses compared to those in calculus-based courses. For example, the correlation between the binary gender measurement and self-identified femininity is −0.89 in algebra-based courses and −0.76 in calculus-based courses. The difference between these two correlations is statistically significant, as determined by Fisher's z test ($p < 0.001$, effect size $= -0.43$). Similarly, the correlation between self-identified femininity and masculinity is stronger in algebra-based courses than in calculus-based courses ($p < 0.001$, effect size $= -0.45$).

In addition, Table III shows that the correlations between students' self-identified androgyny and their response to the binary gender measure as well as their self-identified femininity and masculinity are small. We note there are statistically significant correlations between self-identified androgyny and femininity in calculus-based physics courses ($r = 0.23$), and between self-identified androgyny and masculinity in algebra-based physics courses ($r = 0.18$).

C. Comparison between students' self-identified and reflected appraisal of masculinity, femininity, or androgyny

In this section, we calculated the difference between students' self-identified masculinity, femininity, or androgyny and their reflected appraisal of masculinity, femininity, or androgyny (i.e., their perceptions of how others view them in terms of masculinity, femininity, or androgyny) for women and men in the algebra-based and calculus-based courses. The results are presented in Tables IV, V, and VI for masculinity, femininity, and androgyny, respectively. In the tables, positive differences (e.g., +1, +2, and +3) indicate that the self-identified level of masculinity, femininity, or androgyny is higher than the reflected appraisal of masculinity, femininity, or androgyny, while negative differences (e.g., −1, −2, and −3) indicate that the self-identified level is lower than the reflected appraisal. A value of 0 indicates that the self-identified level of masculinity, femininity, or androgyny matches the reflected appraisal of masculinity, femininity, or androgyny. In addition, we conducted Wilcoxon signed-rank tests to examine whether there are statistically significant differences between students' self-identified masculinity, femininity, or androgyny and their reflected appraisal of masculinity, femininity, or androgyny. The results of the Wilcoxon signed-rank tests are also shown in Tables IV, V, and VI.

Table IV shows that 70% of women and 78% of men in calculus-based courses identified their level of masculinity as consistent with how they perceive others see them.

TABLE IV. Differences between students' self-identified and reflected appraisals of masculinity in the algebra-based and calculus-based courses, along with the results of Wilcoxon signed-rank tests to assess the statistical significance of these differences. Cohen suggested that values of 0.1, 0.3, and 0.5 typically represent small, medium, and large effect sizes, respectively, for Wilcoxon signed-rank tests [101].

		Difference between self-identified and reflected appraisal of masculinity								
		−3 and lower	−2	−1	0	+1	+2	+3 and higher	p value	Effect size
Calculus based	Women	0%	1%	8%	70%	18%	3%	0%	0.001	0.16
	Men	0%	1%	6%	78%	12%	2%	1%	<0.001	0.11
Algebra based	Women	0%	2%	9%	73%	12%	3%	1%	0.088	0.07
	Men	0%	0%	5%	82%	12%	1%	0%	0.004	0.12

TABLE V. Differences between students' self-identified and reflected appraisals of femininity in the algebra-based and calculus-based courses, along with the results of Wilcoxon signed-rank tests to assess the statistical significance of these differences. Cohen suggested that values of 0.1, 0.3, and 0.5 typically represent small, medium, and large effect sizes, respectively, for Wilcoxon signed-rank tests [101].

		Difference between self-identified and reflected appraisal of femininity								
		−3 and lower	−2	−1	0	+1	+2	+3 and higher	p value	Effect size
Calculus based	Women	1%	2%	12%	76%	8%	1%	0%	0.027	−0.10
	Men	1%	1%	7%	82%	8%	1%	0%	0.568	0.01
Algebra based	Women	1%	1%	8%	78%	11%	1%	0%	0.221	0.04
	Men	1%	1%	7%	82%	8%	1%	0%	0.960	−0.00

TABLE VI. Differences between students' self-identified and reflected appraisals of androgyny in the algebra-based and calculus-based courses, along with the results of Wilcoxon signed-rank tests to assess the statistical significance of these differences. Cohen suggested that values of 0.1, 0.3, and 0.5 typically represent small, medium, and large effect sizes, respectively, for Wilcoxon signed-rank tests [101].

		Difference between self-identified and reflected appraisal of androgyny								
		−3 and lower	−2	−1	0	+1	+2	+3 and higher	p value	Effect size
Calculus based	Women	0%	1%	1%	83%	10%	3%	2%	<0.001	0.20
	Men	1%	0%	3%	90%	3%	1%	1%	0.343	0.02
Algebra based	Women	1%	1%	2%	86%	7%	2%	1%	0.002	0.13
	Men	0%	1%	4%	90%	4%	1%	0%	0.999	0.00

However, a sizable percentage of women (21%) and men (15%) perceive themselves as more masculine than how they believe others perceive them, while only a small proportion of women (9%) and men (7%) perceive themselves as less masculine than how they believe others perceive them. The results of Wilcoxon signed-rank tests support our findings, indicating that in the calculus-based courses, both men and women exhibit statistically significant tendencies to perceive themselves as more masculine than they believe others perceive them ($p = 0.001$ for women and $p < 0.001$ for men). Moreover, the effect size of this discrepancy is larger for women (effect size $= 0.16$) than for men (effect size $= 0.11$). Table IV also shows that 16% of women and 13% of men in the algebra-based courses perceive themselves as more masculine than how they believe others perceive them, while only 11% percent of women and 5% of men perceive themselves as less masculine than how they believe others perceive them. Wilcoxon signed-rank tests results indicate that only men in the algebra-based courses exhibit a statistically significant tendency to perceive themselves as more masculine than they believe others perceive them ($p = 0.004$), while this difference is not statistically significant for women ($p = 0.088$).

Additionally, we compared students' self-identified and reflected appraisal of femininity. The results are shown in Table V. Wilcoxon signed-rank tests results indicate no statistically significant difference between these two measurements for men in either calculus-based or algebra-based courses. However, for women, there is a statistically significant tendency to perceive themselves as less feminine than they believe others perceive them ($p = 0.027$, effect size $= -0.10$) in the calculus-based courses, although such a trend was not observed in the algebra-based course.

Regarding androgyny (Table VI), Wilcoxon signed-rank tests results suggest no statistically significant difference between men's self-identified androgyny and their reflected appraisal of androgyny in either algebra-based or calculus-based courses. However, among women in the calculus-based courses, 15% perceive themselves as more androgynous than they believe others perceive them, whereas only 2% perceive themselves as less androgynous than they believe others perceive them. This tendency among women to perceive themselves as more androgynous than how they believe others perceive them holds statistical significance based on the Wilcoxon signed-rank test ($p < 0.001$, effect size $= 0.20$). Similarly, women also exhibit a statistically significant tendency to perceive themselves as more androgynous than they believe others perceive them ($p = 0.002$, effect size $= 0.13$) in the algebra-based courses.

D. Potential factors associated with the discrepancies between students' self-identified and reflected appraisal of masculinity, femininity, or androgyny

In the previous section, we observed that both self-identified women and men in the calculus-based courses have a tendency to perceive themselves as more masculine than they believe others perceive them. For self-identified women, there is also a tendency to perceive themselves as more androgynous than they believe others perceive them in both calculus-based and algebra-based courses. Moreover, self-identified women also tend to view themselves as less feminine than they believe others perceive them in the calculus-based courses.

Since prior studies have shown that individual's sense of femininity, masculinity, or androgyny can be influenced by the environment and context they are in [86,87], we hypothesize that discrepancies observed in the previous section are also associated with various of factors. To explore the potential factors associated with the discrepancies observed, we create a binary variable (self-identified masculinity > reflected appraisal masculinity) coded as 1 if students' self-identified level of masculinity was higher than reflected appraisal of masculinity, and 0 otherwise. We also created a binary variable (self-identified androgyny > reflected appraisal androgyny) coded as 1 if students' self-identified level of androgyny was higher than reflected appraisal of androgyny, and 0 otherwise. Similarly, we created a binary variable (self-identified femininity < reflected appraisal femininity) coded as 1 if students' self-identified level of femininity was lower than reflected appraisal of femininity, and 0 otherwise. The factors we focus on in this study include students' demographic factors and psychological factors as well as course-level factors. The demographic variables available in our data include gender on a binary scale (women vs men), race (non-White vs White), and first-generation status (first generation vs not first generation). The psychological variables available in our data include physics interest, mindset belief, sense of belonging in the class, and gender stigma consciousness, etc. The course-level variables include course type (calculus based vs algebra based), semester (fall vs spring), and course content (Physics 1 vs Physics 2). The criteria we used for selecting potential factors for further investigations include statistically significantly correlated with any of the three discrepancy binary variables (Self-identified masculinity > reflected appraisal masculinity, self-identified femininity < reflected appraisal femininity, and Self-identified androgyny > reflected appraisal androgyny), supported by literature, and demonstrated significant interest in our previous analyses. The above factors satisfying one or more of the three criteria were identified as potential factors for further investigations. In particular, we found that gender and race are statistically significantly correlated with the discrepancy binary variables. The psychological factor gender stigma consciousness and sense of belonging are statistically correlated with at least one of the three discrepancy binary variables and are also supported by literature as discussed in the theoretical framework section. The course type (calculus based vs algebra based) statistically significantly correlates with the discrepancy in androgyny (self-identified androgyny > reflected appraisal androgyny), and we found statistically significant differences between the two types of courses in the analyses discussed earlier. Therefore, we included the following variables as potential factors for further study: Gender on a binary scale (women coded as 1 vs men coded as 0), race (non-White coded as 1 vs White coded as 0), course type (calculus-based coded as 1 vs algebra-based coded as 0), sense of belonging, and gender stigma consciousness. The other factors that were not included in further analyses did not show significant interest based on our criteria.

Table VII shows the correlation coefficients between the discrepancy binary variables and the potential factors discussed above. As shown in the table, the masculinity discrepancy binary variable is positively correlated with self-identifying as women (gender) and non-White (race) as well as gender stigma consciousness. Similarly, the femininity discrepancy binary variable is positively correlated with gender and gender stigma consciousness while negatively correlated with sense of belonging. In addition, the androgyny discrepancy binary variable is positively correlated with gender, race, and gender stigma consciousness while negatively correlated with being in the calculus-based courses. Moreover, Table VII shows that gender

TABLE VII. Correlation coefficients between potential predictors of the discrepancy between self-identified and reflected appraisal of masculinity. p values for the correlations are indicated by *** for $p < 0.001$, ** for $0.001 \leq p < 0.01$, * for $0.01 \leq p < 0.05$, and ^{ns} for $p > 0.05$ (not statistically significant).

Observed variable	1	2	3	4	5	6	7	8
1. Self-identified > reflected appraisal masculinity	...	...	...	...	...	...	...	...
2. Self-identified < reflected appraisal femininity	0.12^{***}	...	...	...	...	...	...	...
3. Self-identified > reflected appraisal androgyny	0.29^{***}	-0.05^{*}	...	...	...	...	...	...
4. Gender	0.05^{*}	0.07^{**}	0.07^{***}	...	...	...	...	...
5. Race	0.05^{*}	-0.02^{ns}	0.06^{**}	0.02^{ns}	...	...	...	...
6. Calculus based	0.02^{ns}	0.01^{ns}	-0.06^{**}	-0.37^{***}	0.03^{ns}	...	...	...
7. Sense of belonging	-0.03^{ns}	-0.08^{***}	-0.13^{ns}	-0.25^{***}	-0.08^{***}	0.18^{***}	...	...
8. Gender stigma consciousness	0.11^{***}	0.08^{***}	0.06^{**}	0.25^{***}	0.11^{***}	-0.01^{ns}	-0.17^{***}	...

negatively correlates with calculus-based and sense of belonging, while positively correlates with gender stigma consciousness. These results indicate that there were fewer women in the calculus-based courses than in the algebra-based courses, and women had a lower average sense of belonging and a higher average gender stigma consciousness. In addition, race negatively correlates with sense of belonging and positively correlates with gender stigma consciousness, which means that non-White students had a lower average sense of belonging and a higher average gender stigma consciousness. Additionally, calculus-based is positively correlated with sense of belonging, which means that students in the calculus-based physics courses had a higher average sense of belonging than students in the algebra-based physics courses. In addition, sense of belonging is negatively correlated with gender stigma consciousness, which means that students who had a higher level of gender stigma consciousness are likely to have a lower level of sense of belonging.

Table VII shows how students' sense of belonging in the class and their gender stigma consciousness are correlated with self-identified gender, race, and the type of courses they are enrolled in. However, it remains unclear what specific effect each demographic and course factor has on sense of belonging and gender stigma consciousness after controlling for the other factors. To address this, we conducted multiple linear regression analyses, with gender stigma consciousness and sense of belonging as the predicted variables, respectively. The unstandardized regression coefficients for each linear regression model are presented in Table VIII. Unstandardized regression coefficients for binary indicators can be interpreted as the mean difference in the dependent variable between the two groups defined by the binary predictor while holding the other predictors constant. For instance, the unstandardized regression coefficient for gender is 0.99. This indicates that, on average, the gender stigma consciousness for self-identified women is 0.99 higher than that for self-identified men, even after accounting for the effects of race and course type. Similarly, non-White students exhibited 0.47 higher gender stigma consciousness than White students, even after controlling for gender and course type. Additionally, students in calculus-based courses had, on average, 0.32 higher gender stigma consciousness than those in algebra-based courses, even after accounting for the effects of gender and race. On the other hand, Table VIII illustrates that gender and non-White status are negatively correlated with sense of belonging, while being enrolled in calculus-based courses is positively correlated with sense of belonging. These effects persist even after controlling for other predictors.

TABLE VIII. Unstandardized regression coefficients of two linear regression models with gender stigma consciousness and sense of belonging being the predicted variable, respectively. p values for the regression coefficients are indicated by *** for $p < 0.001$.

	Predicted variable	
	Gender stigma consciousness	Sense of belonging
Predictive variable		
Gender	$0.99\ (0.08)^{***}$	$-0.49\ (0.05)^{***}$
Race	$0.47\ (0.09)^{***}$	$-0.22\ (0.06)^{***}$
Calculus-based	$0.32\ (0.08)^{***}$	$0.24\ (0.05)^{***}$

E. The effect of different factors on the discrepancies between students' self-identified and reflected appraisal of masculinity

In the previous sections, we identified various potential factors associated with the difference between students' self-identified and reflected appraisal of masculinity, as well as the relationships between these factors. In this section, we conducted logistic regression analysis to examine the effect of each potential factor on the difference between students' self-identified and reflected appraisal of masculinity. As mentioned earlier, we utilized a binary dependent variable (self-identified masculinity > reflected appraisal masculinity) in the logistic regression, coding it as 1 if students' self-identified level of masculinity was higher than their reflected appraisal of masculinity and 0 otherwise. To better investigate the effect of each factor, we initially tested a model with only demographic factors (gender and race), then we added the course-level factor (calculus based) to the model, and finally we included the psychological factors (sense of belonging and gender stigma consciousness). We assessed changes in the model fit and how the addition of predictors influences the coefficients of the other predictors. Specifically, we used AIC (Akaike information criterion) value to compare different models, and smaller values of AIC indicate a better balance between goodness of fit and model complexity. Therefore, the model with the lowest AIC is often considered the best among the models compared [102].

We first tested a model (model 1) to investigate the effects of the demographic factors on the discrepancy between students' self-identified and reflected appraisal of masculinity. The demographic factors include gender on a binary scale (women coded as 1 v. men coded as 0) and race (non-White coded as 1 vs White coded as 0). The results of model 1 are presented in Table IX. As shown in the table, both gender and race are statistically significant predictors. The odds ratio (OR) in logistic regression represents the multiplicative change in the odds of the event occurring associated with a one-unit change in the predictor variable, while holding other variables constant. In the context of model 1, the odds ratio is 1.32 for gender, indicating that women were 1.32 times more likely than men to perceive themselves as more masculine than how they believe others view them after controlling for race. Likewise, with an odds ratio of 1.38, non-White students were 1.38 times more likely than White students to perceive themselves as more masculine than how they believe others view them after controlling for gender.

Then, we added the course level factor (calculus based) to the model (model 2). As shown in Table IX, the effect of calculus-based is not statistically significant, which is consistent with the correlation coefficients in Table VII. In addition, we note that the effect of race becomes not statistically significant even though the odds ratio ($\mathrm{OR} = 1.36$) for race is almost the same as that ($\mathrm{OR} = 1.38$) in model 1. Moreover, we note that after adding the predictor calculus-based, the AIC value almost does not change compared with model 1.

Finally, we added the psychological factors (sense of belonging and gender stigma consciousness) to the model (model 4). Model 4 shows that sense of belonging is not a statistically significant predictor of the discrepancy between students' self-identified and reflected appraisal of masculinity, while gender stigma consciousness is a statistically significant predictor. In particular, the odds ratio for gender stigma consciousness is 1.15, indicating that a one-unit increase in gender stigma consciousness is associated with 1.15 times higher (or 15% increase in) odds of perceiving oneself as more masculine than one believes others see them. In addition, we note that model 3 has a significantly lower AIC value compared with models 1 and 2, making it a more favorable choice for capturing the underlying patterns in the dataset [102]. Furthermore, we observed that after adding gender stigma consciousness to the model, the effect of gender is no longer statistically significant and the odds ratio for race significantly decreases, indicating that the effect of gender and race on the discrepancy between students' self-identified and reflected appraisal of masculinity might be mediated by gender stigma consciousness.

To test our hypothesis, we conducted structural equation modeling (SEM) analyses to further investigate the relationships between gender, race, gender stigma consciousness, and the discrepancy between students' self-identified and reflected appraisal of masculinity. The advantage of SEM is that it can show not only the direct regression relation between two constructs but also all the indirect relations mediated through other constructs [98]. The results of the SEM model are presented in Fig. 3 (model 4). The regression coefficients in the SEM models in this study are standardized to be in units of standard deviations. Thus, model 4 shows that women have a gender stigma consciousness score that is 0.25 standard deviations higher than men. The link between gender stigma consciousness and the final outcome in model 4 indicates that, for each standard deviation increased in gender stigma consciousness, the odds of identifying oneself as more masculine than one thinks others perceive them increases by a factor of $e^{0.14} = 1.15$. Although the direct effects of gender and race on the discrepancy between students' self-identified

TABLE IX. Results of logistic regression analysis with the binary variable (self-identified masculinity > reflected appraisal masculinity) as the dependent variable.

	Model 1		Model 2		Model 3	
Predictor	p-value	Odds ratio	p-value	Odds ratio	p-value	Odds ratio
Intercept	<0.001	0.16	<0.001	0.13	<0.001	0.10
Gender	0.041	1.32	0.015	1.43	0.227	1.21
Race	0.049	1.38	0.060	1.36	0.147	1.27
Calculus	⋯	⋯	0.138	1.25	0.247	1.19
Sense of belonging	⋯	⋯	⋯	⋯	0.765	0.98
Gender stigma consciousness	⋯	⋯	⋯	⋯	<0.001	1.15
AIC	1546.2		1546		1531.9	

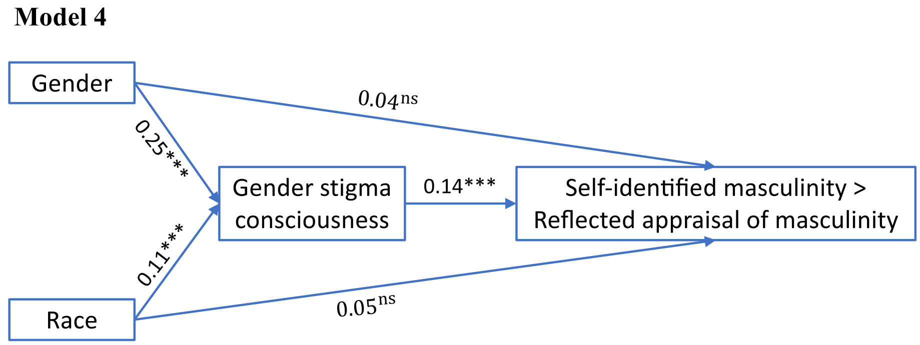

FIG. 3. Results of the path analysis part of the structural equation modeling, illustrating the relationships between gender, race, gender stigma consciousness, and the discrepancy between students' self-identified and reflected appraisal of masculinity. Self-identified masculinity > reflected appraisal of masculinity is a binary variable coded as 1 when students' self-identified masculinity is higher than their reflected appraisal of masculinity, and 0 otherwise. The solid lines with a single arrowhead represent regression paths. The numbers represent standardized regression coefficients with $p < 0.001$ indicated by ***, and $p > 0.05$ indicated by ns.

and reflected appraisal of masculinity are not statistically significant, they indirectly predict the discrepancy through gender stigma consciousness. This model fits the data very well (CFI = 1.0, TLI = 1.0, RMSEA = 0, SRMR = 0) [98]. Since gender and race are statistically significant predictors as shown in Model 1 in Table IX, and the effects of gender and race become not statistically significant after adding gender stigma consciousness (as shown in model 4), the effects of gender and race are fully mediated through gender stigma consciousness, which confirms our hypothesis [103].

F. The effect of different factors on the discrepancies between students' self-identified and reflected appraisal of femininity

As discussed earlier, Table V shows that women in the calculus-based physics courses exhibited a statistically significant tendency to perceive themselves as less feminine than they believe others perceive them, while there was no such a tendency in the algebra-based physics courses. To explore this further, we also conducted logistic regression analyses to investigate the effects of the potential factors on the discrepancy between students' self-identified and reflected appraisal of femininity. In the logistic regression analysis, we used a binary dependent variable coded as 1 if students' self-identified level of femininity was lower than reflected appraisal of femininity, and 0 otherwise. Similar to the earlier logistic regression analyses for masculinity, we initially tested a model with only the demographic factors as predictors. Subsequently, we introduced the course-level factor, and finally, incorporated the psychological factors into the analysis.

The results of the logistic regression analyses are shown in Table X. As shown in the table, when we only included gender and race as predictors (model 5), only gender emerged as a statistically significant predictor. Specifically, the odds ratio for gender was 1.70, indicating that self-identified women were 1.70 times more likely than self-identified men to perceive themselves as less feminine than how they believe other people view them. This result is consistent with the results in Table VII showing that gender is positively correlated with the femininity discrepancy binary variable. Then, we included the course level factor (calculus based) as a predictor (model 6), and we found the effect of calculus based is not statistically significant ($p = 0.076$). In addition, we note that the AIC value of model 6 is almost the same as that of model 5, indicating that adding calculus based as a predictor does not significantly improve the model.

TABLE X. Results of logistic regression analysis with the binary variable (self-identified femininity < reflected appraisal femininity) as the dependent variable.

	Model 5		Model 6		Model 7	
Predictor	p value	Odds ratio	p value	Odds ratio	p value	Odds ratio
Intercept	<0.001	0.08	<0.001	0.06	<0.001	0.15
Gender	0.001	1.70	<0.001	1.91	0.025	1.52
Race	0.339	0.81	0.302	0.80	0.119	0.70
Calculus	⋯	⋯	0.076	1.38	0.058	1.41
Sense of belonging	⋯	⋯	⋯	⋯	0.001	0.78
Gender stigma consciousness	⋯	⋯	⋯	⋯	0.043	1.10
AIC	1165.1		1163.9		1136.7	

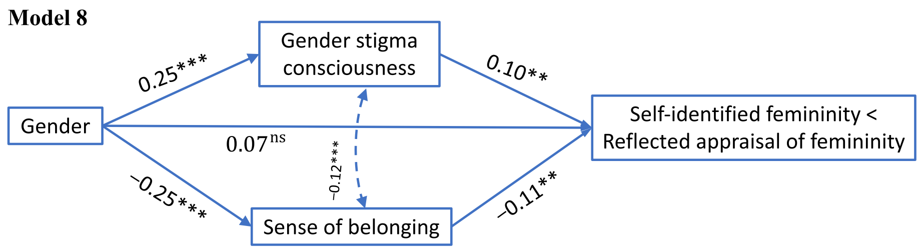

FIG. 4. Results of the path analysis part of the structural equation modeling, illustrating the relationships between gender (on a binary scale), gender stigma consciousness, and the discrepancy between students' self-identified and reflected appraisal of femininity. Self-identified femininity < reflected appraisal of femininity is a binary variable coded as 1 when students' self-identified femininity is lower than their reflected appraisal of femininity, and 0 otherwise. The solid lines with a single arrowhead represent regression paths. The curved dashed line with two arrowheads represents a residual covariance. The numbers next to the regression paths represent standardized regression coefficients with $p < 0.001$ indicated by $***$, $0.0001 \le p < 0.01$ indicated by **, and $p > 0.05$ indicated by ns.

Finally, we included sense of belonging and gender stigma consciousness to the model (model 7), and we found that both sense of belonging and gender stigma consciousness are statistically significant predictors. Specifically, the odds ratio is 0.78 for sense of belonging and 1.10 for gender stigma consciousness. These results indicate that a one-unit increase in sense of belonging is associated with a 22% (which is 1–0.78) decrease in the odds of perceiving oneself as less feminine than one believes others see them, while one-unit increase in gender stigma consciousness is associated with a 10% increase in the odds of having this perception. We note that the AIC value of model 7 is significantly higher than that for both models 5 and 6, suggesting that model 7 offers a more accurate and efficient representation of the relationships among the variables under consideration [102].

In addition, we found that the effect of gender decreases after adding sense of belonging and gender stigma consciousness, which indicates that the effect of gender on the outcome variable might be mediated by the two psychological factors. To further test our hypothesis, we conducted structural equation modeling analyses. We include gender, sense of belonging, and gender stigma consciousness to the SEM model, as they are statistically significant predictors in Model 7. The results of the SEM model are presented in Fig. 4 (model 8). Model 8 also fits the data well: CFI = 1.0, TLI = 1.0, RMSEA = 0, SRMR = 0 [98]. We note that even though gender directly predicts the outcome variable in models 5 and 6, this direct effect becomes not statistically significant after adding sense of belonging and gender stigma consciousness (model 8), which means that the effect of gender on the discrepancy between students' self-identified and reflected appraisal of femininity is fully mediated through sense of belonging and gender stigma consciousness.

G. The effect of different factors on the discrepancies between students' self-identified and reflected appraisal of androgyny

As discussed earlier, Table VI shows that women in both calculus-based and algebra-based physics courses exhibited a statistically significant tendency to perceive themselves as more androgynous than they believe others perceive them. To explore this further, we conducted logistic regression analyses to investigate the effects of the potential factors on the discrepancy between students' self-identified and reflected appraisal of androgyny. In the logistic regression analysis, we used a binary dependent variable coded as 1 if students' self-identified level of androgyny was higher than reflected appraisal of androgyny, and 0 otherwise.

Similar to the logistic regressions discussed earlier, we initially tested a model (model 9) to investigate the effects

TABLE XI. Results of logistic regression analysis with the binary variable (self-identified androgyny > reflected appraisal androgyny) as the dependent variable.

	Model 9		Model 10		Model 11	
Predictor	p value	Odds ratio	p value	Odds ratio	p value	Odds ratio
Intercept	<0.001	0.40	<0.001	0.47	<0.001	0.32
Gender	<0.001	1.40	0.015	1.29	0.046	1.25
Race	0.010	1.36	0.008	1.38	0.011	1.37
Calculus	…	…	0.044	0.81	0.017	0.77
Sense of belonging	…	…	…	…	0.350	1.04
Gender stigma consciousness	…	…	…	…	0.050	1.06
AIC	2558.1		2556.1		2492.5	

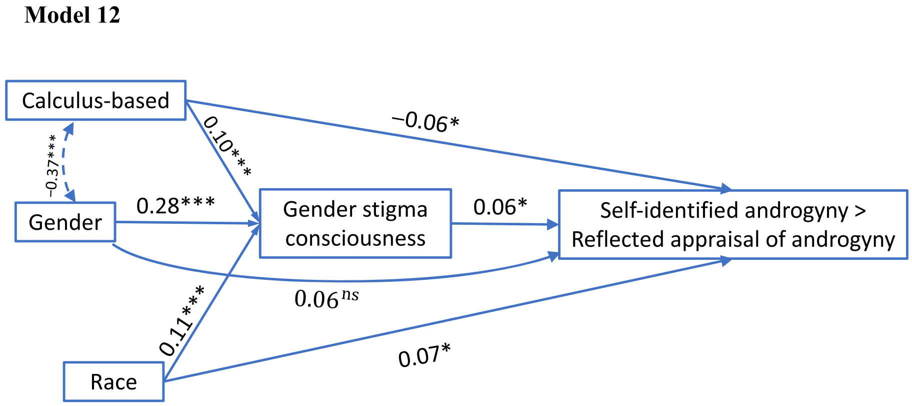

FIG. 5. Results of the path analysis part of the structural equation modeling, illustrating the relationships between gender (on a binary scale), gender stigma consciousness, and the discrepancy between students' self-identified and reflected appraisal of androgyny. Self-identified androgyny > reflected appraisal of androgyny is a binary variable coded as 1 when students' self-identified androgyny is higher than their reflected appraisal of androgyny, and 0 otherwise. The solid lines with a single arrowhead represent regression paths. The curved dashed line with two arrowheads represents a residual covariance. The numbers next to the regression paths represent standardized regression coefficients with $p < 0.001$ indicated by ***, $0.0001 \leq p < 0.01$ indicated by **, $0.01 \leq p \leq 0.05$ indicated by **, and $p > 0.05$ indicated by ns.

of the demographic factors on the discrepancy between students' self-identified and reflected appraisal of androgyny. The results of model 9 are presented in Table XI. As shown in the table, both gender (on a binary scale) and race are statistically significant predictors. Specifically, the odds ratio for gender is 1.40, indicating that self-identified women were 1.40 times more likely than self-identified men to perceive themselves as more androgynous than how they believe other people view them. Similarly, that self-identified non-White students were 1.36 times more likely than White students to hold this perception. Then we added the course level factor (calculus based) to the model (model 10), and found it is also a statistically significant predictor with odds ratio 0.81, which means that students in the calculus-based courses have 19% lower odds of holding this perception. Finally, we introduced sense of belonging and gender stigma consciousness into the model (model 11). The results of model 11 reveal that gender stigma consciousness is a statistically significant predictor, while the effect of sense of belonging is not statistically significant. The odds ratio for gender stigma consciousness is 1.06, which means that one-unit increase in gender stigma consciousness is associated with a 6% increase in the odds of perceiving oneself as more androgynous than how they believed other people viewed them. Furthermore, we note that the AIC value of model 10 is slightly lower than that of model 9, while model 11 has significantly a lower AIC value compared with both models 9 and 10, which indicates that model 11 is a more favorable choice for explaining the variability in the data than models 9 and 10.

By comparing model 11 with models 9 and 10, we observed that after adding gender stigma consciousness as a predictor, the effect of gender decreases, indicating that the effect of gender on the discrepancy between students' self-identified and reflected appraisal of androgyny might be mediated by gender stigma consciousness. To test our hypothesis, we conducted structural equation modeling analyses with the statistically significant predictors in model 11. The results of the SEM model are presented in Fig. 5. Figure 5 (model 12) shows the direct effect of gender on the outcome variable is not statistically significant, while this effect is statistically significant in model 9, indicating that the effect of gender is fully mediated through gender stigma consciousness, while the effect of race is partially mediated through gender stigma consciousness. The SEM model also fits the data well: CFI = 1.0, TLI = 1.0, RMSEA = 0, SRMR = 0 [98].

VII. DISCUSSION

In this study, we investigated students' self-identified and reflected appraisals of masculinity, femininity, and androgyny in introductory physics courses using gradational measures. This approach allowed us to capture a broader range of variations in students' femininity and masculinity identities and perceptions within the context of physics courses. By going beyond the traditional binary gender measures, we were able to explore research questions that might have remained hidden, thereby gaining valuable insights into the multifaceted nature of gender. Our findings can serve as a stepping stone for future research endeavors aimed at deepening our understanding of gender-related issues in physics education.

In response to RQ1, our study revealed significant variation in students' self-identified femininity and masculinity within the binary categories of women and men. In particular, we found that only about 35% of students identified themselves as exclusively feminine or exclusively masculine, indicating that the majority of students embraced both femininity and masculinity as aspects of their identity. In addition, around 23% of students indicated a nonzero level of androgyny. These results are consistent with prior studies [66,69,70,83] showing that gender is a social construct that goes beyond a simple binary classification and individuals can experience and express a wide range of femininity and masculinity.

In response to RQ2, our results show a positive correlation between identifying as a woman and higher self-identified femininity, as well as between identifying as a man and higher self-identified masculinity. These results align with traditional understandings of gender roles and expectations. However, we note that identifying as a woman or man and self-identified femininity or masculinity are not perfectly correlated, and our study revealed instances of students identifying as women but identifying a higher level of masculinity than femininity. Furthermore, we found that even though femininity and masculinity identities are negatively correlated, they are not perfectly correlated suggesting that they are not mutually exclusive and opposite. In other words, the presence of high femininity does not preclude the possibility of also having high masculinity, and vice versa. These findings are consistent with prior studies showing that stereotypically masculine and feminine interests and behaviors are not mutually exclusive and can even be positively associated [79,80]. Our study contributes to the existing body of research by showing that physics students' self-identified femininity and masculinity are also not mutually exclusive, which highlights the multidimensionality of gender identity and emphasizes the importance of moving beyond binary gender measurement in physics education research [66,104]. By recognizing and embracing the complexity of gender identity, we may foster a more inclusive and comprehensive understanding of gender-related issues in physics education.

In response to RQ3, our study reveals discrepancies between students' self-identified masculinity, femininity, or androgyny and reflected appraisal of masculinity, femininity, or androgyny. Specifically, we found that many students in introductory physics courses perceive themselves as more masculine than they believe others view them, with this divergence being particularly pronounced among students who identify as women in the calculus-based physics courses, where women are underrepresented. In addition, we found that women also tend to perceive themselves as more androgynous than they believe others perceive them in both the algebra-based and calculus-based physics courses. On the other hand, women in the calculus-based courses exhibit a tendency to perceive themselves as less feminine than they believe others perceive them. Prior research has shown that individuals' self-identified gender may not always align with how others perceive them [70,85]. Our study contributes to the existing literature by showing that there are also discrepancies between students' self-identified masculinity, femininity, and androgyny and their perception of how others view them in the context of introductory physics courses.

One potential factor that may contribute to these discrepancies is the prevailing masculine culture in physics. Prior studies have shown that women in science and technology often distance themselves from traditional femininity as a means of fitting into disciplines dominated by men [89,90]. For instance, a study on women in physics graduate programs found that some participants perceived feminine characteristics as conflicting with the logic and nature of physics itself. Participants explained that feminine attire, such as dresses and heels, felt out of place in the physics lab, and women may even face pressure to adopt more masculine traits to fit into the environment [66,105]. Therefore, women may worry about not performing enough masculinity to be accepted by others in the physics learning environment. This concern may contribute to our finding that many women in the physics courses feel that their masculinity or androgyny is not adequately perceived by others compared to their self-perception and their femininity is overemphasized by others compared to their self-identified level of femininity.

Given the potential influence of the masculine culture in physics on students' gender identities and perceptions, we examined the effects of several potential factors on the observed discrepancies (RQ3). These factors include demographic factors (gender and race), course level factor (calculus based vs Algebra based), and psychological factors (sense of belonging and gender stigma consciousness). The results of the logistic regression analyses indicate that having a high level of gender stigma consciousness positively predicts the odds of perceiving oneself as more masculine and androgynous, and less feminine than one believes others view them. Similarly, having a low level of a sense of belonging positively predicts the odds of perceiving oneself as less feminine than they believe other people view them. These results suggest that individuals who are more conscious of gender norms and stereotypes, or who have a low sense of belonging within the field of physics, may face increased pressure to conform to those stereotypical expectations, ultimately leading to a discrepancy between their self-perception and how they believe others perceive them. Moreover, our study shows that self-identifying as women positively predicts the likelihood of perceiving oneself as more masculine, more androgynous, and less feminine than how one believes other people see them, and gender stigma consciousness mediated these effects of identifying as women on these discrepancies. Similarly, sense of belonging also mediates the effect of identifying as women on the discrepancy between self-identified and reflected appraisal of femininity. These findings indicate that self-identified women in the physics courses may encounter a higher level of stereotypes and stigmas and a lower level of a sense of belonging, exacerbating the challenges they face in reconciling their own sense of masculinity, femininity, or androgyny with societal expectations.

In addition to women, we note that students self-identifying as non-White are also more likely than White students to perceive themselves as more masculine and androgynous than they believe others view them, and gender stigma consciousness also mediates these effects. This result may be related to the cultural nature of gender and the influence of societal norms. Prior studies have shown that popular culture often portrays White men as standard bearers of masculinity, associated with traits such as confidence, individualism, and competition [106]. These traits have also been shown to be commonly perceived as components of physics culture [107,108]. Prior studies suggest that the prevailing stereotypical image of a physicist as an eccentric genius working in isolation contributes to the culture of isolation and competition within physics [109,110]. However, other cultural values may not align with these specific traits. For instance, Asian culture values humility and communalism, which may not fit within the framework of traditional White notions of masculinity [106]. Prior studies have shown that Asian men are often rated as less masculine than White men by both men and women [111,112]. Consequently, Asian men may face the dilemma of conforming to White notions of masculinity or accepting the perception that they are not masculine enough in an environment dominated by the White masculine culture, such as in the field of physics. Research has shown that Asian men often prefer to align themselves with the dominant hegemonic masculinity rather than aligning with other marginalized groups, as patriarchal rewards and advantages are associated with conforming to the dominant group [106,113]. Similarly, some prior studies have shown that Black and Mexican-American women scored lower on masculinity than White women, as measured by the BSRI, which suggests that non-White women exhibit a lower level of stereotypical masculine traits than White women [114,115]. Therefore, in male-dominated disciplines like physics, where masculinity is highly valued, students from non-White racial or ethnic minority groups may also experience concerns about not being perceived as masculine enough to fit into this environment. Future work with more diverse samples will consider the intersectional nature of race and gender as it relates to gender stigma consciousness in physics.

In response to RQ4, we compared the results for the above research questions between the algebra-based and calculus-based physics courses. While the overall findings for RQ1 and RQ2 are consistent between the two types of courses, we do find some small differences. For example, we found that self-identified women in the calculus-based courses are less likely than those in the algebra-based courses to identify themselves as exclusively feminine. Additionally, we note that the correlations among the binary gender measurement and the gradational measurements of femininity and masculinity are slightly higher for students in the algebra-based courses compared to those in the calculus-based courses. With regard to RQ3, we found that women tend to perceive themselves as more masculine and less feminine in the calculus-based physics courses, while these tendencies were not found in the algebra-based courses. After including calculus based as a predictor in the logistic regression models, we found that being in the calculus-based course is positively associated with the odds of perceiving oneself as more androgynous than they believe others view them, while it is not statistically significantly associated with the observed discrepancies regarding masculinity and femininity. In addition, we found that students in the calculus-based courses had a higher average level of gender stigma consciousness and sense of belonging even after controlling for gender and race, while gender stigma consciousness and sense of belonging are negatively correlated with each other. The similarities between the algebra-based courses and calculus-based courses suggest that some of the findings may be inherent to the physics classroom and the masculine culture of physics. The differences in the findings between the calculus-based and algebra-based courses indicate that students' self-identified gender and perception of how others see them might be associated with the differences in these two types of courses, such as the representation of different gender groups. Moreover, there might also be population differences contributing to the findings. For example, women in the algebra-based courses, on average, do not identify as more masculine and less feminine than they think others perceive them while calculus-based women do. This may be due to the differences in disciplinary culture, with most women in the algebra-based courses being prehealth majors (which prior research shows is related to higher levels of femininity [86,87]) and most women in the calculus-based courses being engineers. Students' learning experiences might also differ in these two types of courses with distinct gender representation and disciplinary culture, potentially influencing our findings. Further studies are needed to investigate the factors contributing to the observed differences between the algebra-based and calculus-based courses. We also note that the findings which are consistent across calculus-based and algebra-based courses may also be due to broader institutional culture, which warrants further investigation.

Previous research has shown that when there is a disparity between one's gender identity and the expectations or judgments of others, it can lead to various psychosocial and health-related consequences such as depression, anxiety, and stress [70,85]. In our study, we found there are also discrepancies between students' self-identified and reflected appraisals of masculinity, femininity, and androgyny in introductory physics courses, however, the impact of these discrepancies on students' physics learning remains unexplored. It is unclear whether and how this divergence affects various aspects of students' physics education, including their motivational beliefs, academic performance, and persistence in the field of physics. Further investigations can explore the potential impacts of these discrepancies on students' academic and motivational outcomes and shed light on strategies for fostering inclusivity and promoting equitable experiences in physics classes.

VIII. LIMITATIONS AND FUTURE DIRECTIONS

In this study, we used quantitative research methods to investigate students' femininity, masculinity, or androgyny. In future studies, it would be valuable to incorporate qualitative methods such as interviews to gain a deeper understanding of the reasons behind students' choices in self-identified and reflected appraisals of femininity, masculinity, or androgyny and their relationship with the culture within physics. Qualitative approaches can help us delve into individuals' experiences and perceptions, providing a more comprehensive understanding of gender-related issues in physics. In addition, in this study, we did not provide a specific prompt to ask students to answer the gender-related questions in the context of the physics classroom, therefore, it is unclear whether our findings are related to students' experiences in the physics courses. Future studies could investigate whether and how students' responses to the survey questions change when such prompts are provided. Moreover, even though we found that the gradational measures of femininity, masculinity, and androgyny can capture a broader range of variations in students' gender identities and perceptions compared with the binary gender measure, these measures may still not be able to represent the full spectrum of gender identities. Future studies could delve into further investigations aimed at refining the gender measurement frameworks. In addition, in this study, we investigated the effects of potential demographic factors, course level factors, and psychological factors on the observed discrepancies, future studies could also explore the effects of other factors such as academic performance and peer interaction. Additionally, our study was conducted at a single public research university in the U.S. To examine the generalizability of our findings, it would be valuable to conduct similar investigations in other types of institutions and across different countries and cultures. Furthermore, investigating students' self-identified femininity, masculinity, or androgyny and reflected appraisals of femininity, masculinity, or androgyny in other disciplines, such as chemistry and math, would provide valuable insights into the influence of the learning environment and disciplinary culture on students' gender identities and perceptions.

ACKNOWLEDGMENTS

We thank all students and instructors whose data were analyzed, as well as Dr. Emily Driessen and all reviewers, for their constructive feedback on the manuscript.

[1] J. Blue, A. L. Traxler, and X. C. Cid, Gender matters, Phys. Today **71**, No. 3, 40 (2018).

[2] American Physical Society, *Newest Data Shows Mixed Progress for Women and Marginalized Groups*, Physics Higher Ed (2022), https://www.aps.org/publications/apsnews/202211/newest-data.cfm#:~:text=Women%20earned%201%20in%204,to%20just%2019%25%20in%202015.

[3] R. Ivie and A. Porter, *Women in Physics and Astronomy* (American Institute of Physics, New York, 2019), https://www.aip.org/statistics/reports/women-physics-and-astronomy-2019.

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

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

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

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

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

[9] I. Rodriguez, G. Potvin, and L. H. Kramer, How gender and reformed introductory physics impacts student success in advanced physics courses and continuation in the physics major, Phys. Rev. Phys. Educ. Res. **12**, 020118 (2016).

[10] N. I. Karim, A. Maries, and C. Singh, Do evidence-based active-engagement courses reduce the gender gap in introductory physics?, Eur. J. Phys. **39**, 025701 (2018).

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

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

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

[14] A. Madsen, S. B. McKagan, and E. C. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap?, Phys. Rev. ST Phys. Educ. Res. **9**, 020121 (2013).

[15] C. Lindstrøm and M. D. Sharma, Self-efficacy of first year university physics students: Do gender and prior formal instruction in physics matter?, Int. J. Innov. Sci. Math. Educ. **19**, 1 (2011), https://core.ac.uk/download/pdf/229407173.pdf.

[16] S. J. Pollock, N. D. Finkelstein, and L. E. Kost, Reducing the gender gap in the physics classroom: How sufficient is interactive engagement?, Phys. Rev. ST Phys. Educ. Res. **3**, 010107 (2007).

[17] L. E. Kost, S. J. Pollock, and N. D. Finkelstein, Characterizing the gender gap in introductory physics, Phys. Rev. ST Phys. Educ. Res. **5**, 010101 (2009).

[18] P. B. Kohl and H. V. Kuo, Introductory physics gender gaps: Pre- and post-studio transition, AIP Conf. Proc. **1179**, 173 (2009).

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

[20] J. G. Stout, T. A. Ito, N. D. Finkelstein, and S. J. Pollock, How a gender gap in belonging contributes to the gender gap in physics participation, AIP Conf. Proc. **1513**, 402 (2013).

[21] T. Espinosa, K. Miller, I. Araujo, and E. Mazur, Reducing the gender gap in students' physics self-efficacy in a team- and project-based introductory physics class, Phys. Rev. Phys. Educ. Res. **15**, 010132 (2019).

[22] R. Henderson, J. Stewart, and A. Traxler, Partitioning the gender gap in physics conceptual inventories: Force concept inventory, force and motion conceptual evaluation, and conceptual survey of electricity and magnetism, Phys. Rev. Phys. Educ. Res. **15**, 010131 (2019).

[23] R. Henderson, G. Stewart, J. Stewart, L. Michaluk, and A. Traxler, Exploring the gender gap in the Conceptual Survey of Electricity and Magnetism, Phys. Rev. Phys. Educ. Res. **13**, 020114 (2017).

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

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

[26] Y. Li and C. Singh, Impact of perceived recognition by physics instructors on women's self-efficacy and interest, Phys. Rev. Phys. Educ. Res. **19**, 020125 (2023).

[27] C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham, and J. Handelsman, Science faculty's subtle gender biases favor male students, Proc. Natl. Acad. Sci. U. S. A. **109**, 16474 (2012).

[28] A. A. Eaton, J. F. Saunders, R. K. Jacobson, and K. West, How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates, Sex Roles **82**, 127 (2020).

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

[30] B. A. Adegoke, Impact of interactive engagement on reducing the gender gap in quantum physics learning outcomes among senior secondary school students, Phys. Educ. **47**, 462 (2012).

[31] L. McCullough, Gender, context, and physics assessment, J. Int. Women's Stud. **5**, 20 (2004), https://vc.bridgew.edu/jiws/vol5/iss4/2.

[32] R. Henderson, P. Miller, J. Stewart, A. Traxler, and R. Lindell, Item-level gender fairness in the Force and Motion Conceptual Evaluation and the Conceptual Survey of Electricity and Magnetism, Phys. Rev. Phys. Educ. Res. **14**, 020103 (2018).

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

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

[35] L. McCullough, Gender differences in student responses to physics conceptual questions based on question context, in *Proceedings of ASQ Advancing the STEM Agenda in Education, the Workplace and Society, University of Wisconsin-Stout* (2011), https://www.per-central.org/items/detail.cfm?ID=13428.

[36] R. R. Hake, Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization, in *Physics Education Research Conference* (2002), p. 1, https://web.physics.indiana.edu/hake/PERC2002h-Hake.pdf.

[37] J. Docktor and K. Heller, Gender differences in both force concept inventory and introductory physics performance, AIP Conf. Proc. **1064**, 15 (2008).

[38] C. T. Richardson and B. W. O'Shea, Assessing gender differences in response system questions for an introductory physics course, Am. J. Phys. **81**, 231 (2013).

[39] S. Bates, R. Donnelly, C. MacPhee, D. Sands, M. Birch, and N. R. Walet, Gender differences in conceptual understanding of Newtonian mechanics: A UK cross-institution comparison, Eur. J. Phys. **34**, 421 (2013).

[40] A. A. Casper, N. Rebolledo, A. K. Lane, L. Jude, and S. L. Eddy, "It's completely erasure": A qualitative exploration of experiences of transgender, nonbinary, gender nonconforming, and questioning students in biology courses, CBE Life Sci. Educ. **21**, ar69 (2022).

[41] S. Monro, Beyond male and female: Poststructuralism and the spectrum of gender, Int. J. Transgenderism **8**, 3 (2005).

[42] B. B. Frey, *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (SAGE Publications, Inc., Thousand Oaks, CA, 2018).

[43] C. K. Sigelman and E. A. Rider, *Life-Span Human Development* (Cengage Learning, Boston, MA, 2014).

[44] J. E. Maddux and B. A. Winstead, *Psychopathology: Foundations for a Contemporary Understanding* (Routledge, New York, 2015).
[45] S. L. Bem, Theory and measurement of androgyny: A reply to the Pedhazur-Tetenbaum and Locksley-Colten critiques, J. Pers. Soc. Psychol. **37**, 1047 (1979).
[46] J. T. Spence and R. L. Helmreich, Androgyny versus gender schema: A comment on Bem's gender schema theory, Psychol. Rev. **88**, 365 (1981).
[47] G. H. Mead, *Mind, Self, and Society* (University of Chicago Press, Chicago, IL, 1934), Vol. 111.
[48] C. H. Cooley, *Human Nature and the Social Order* (Transaction Publishers, Piscataway, NJ, 1902).
[49] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Damage caused by women's lower self-efficacy on physics learning, Phys. Rev. Phys. Educ. Res. **16**, 010118 (2020).
[50] Y. Li and C. Singh, Do female and male students' physics motivational beliefs change in a two-semester introductory physics course sequence?, Phys. Rev. Phys. Educ. Res. **18**, 010142 (2022).
[51] L. Archer, J. Moote, B. Francis, J. DeWitt, and L. Yeomans, The "exceptional" physics girl: A sociological analysis of multimethod data from young women aged 10–16 to explore gendered patterns of post-16 participation, Am. Educ. Res. J. **54**, 88 (2017).
[52] A. T. Danielsson, Exploring woman university physics students 'doing gender' and 'doing physics,', Gender Educ. **24**, 25 (2012).
[53] A. J. Gonsalves, Exploring how gender figures the identity trajectories of two doctoral students in observational astrophysics, Phys. Rev. Phys. Educ. Res. **14**, 010146 (2018).
[54] Y. Li and C. Singh, Sense of belonging is an important predictor of introductory physics students' academic performance, Phys. Rev. Phys. Educ. Res. **19**, 020137 (2023).
[55] E. Seymour, Tracking the processes of change in US undergraduate education in science, mathematics, engineering, and technology, Sci. Educ. **86**, 79 (2002).
[56] N. M. Hewitt and E. Seymour, A long, discouraging climb, ASEE Prism **1**, 24 (1992).
[57] E. Seymour and N. M. Hewitt, *Talking About Leaving: Why Undergraduates Leave the Sciences* (Westview Press, Boulder, CO, 1997).
[58] R. J. Beichner and J. M. Saul, Introduction to the SCALE-UP (student-centered activities for large enrollment undergraduate programs) project, in *Proceedings of the International School of Physics "Enrico Fermi," Varenna, Italy* (2003), Vol 1.
[59] E. Etkina and A. Van Heuvelen, Investigative science learning environment–A science process approach to learning physics, Research-based Reform of University Physics **1**, 1 (2007), https://per-central.org/per_reviews/media/volume1/ISLE-2007.pdf.
[60] V. P. Coletta, Reducing the FCI gender gap, presented at PER Conf. 2013, College Park, MD, 10.1119/perc.2013.pr.013.
[61] K. R. Binning, N. Kaufmann, E. M. McGreevy, O. Fotuhi, S. Chen, E. Marshman, Z. Y. Kalender, L. Limeri, L. Betancur, and C. Singh, Changing social contexts to foster equity in college science courses: An ecological-belonging intervention, Psychol. Sci. **31**, 1059 (2020).
[62] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, Sci. Educ. **91**, 847 (2007).
[63] Y. Li and C. Singh, Inclusive learning environments can improve student learning and motivational beliefs, Phys. Rev. Phys. Educ. Res. **18**, 020147 (2022).
[64] S. D. Strong, S. Wallace, C. Feldman, and J. R. Welch, Cultural competency with non-binary and genderqueer individuals: Results from a qualitative participatory action research pilot study, in *Trans. Health* (2022), pp. 135–156, https://doi.org/10.14361/9783839450826-011.
[65] G. N. Rider, J. A. Vencill, D. R. Berg, R. Becker-Warner, L. Candelario-Pérez, and K. G. Spencer, The gender affirmative lifespan approach (GALA): A framework for competent clinical care with nonbinary clients, Int. J. Transgenderism **20**, 275 (2019).
[66] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).
[67] J. S. Hyde, R. S. Bigler, D. Joel, C. C. Tate, and S. M. van Anders, The future of sex and gender in psychology: Five challenges to the gender binary, Am. Psychol. **74**, 171 (2019).
[68] S. Jaroszewski, D. Lottridge, O. L. Haimson, and K. Quehl, "Genderfluid" or "Attack Helicopter" responsible HCI research practice with non-binary gender variation in online communities, in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, New York, NY, 2018), p. 1, https://doi.org/10.1145/3173574.3173881.
[69] D. Magliozzi, A. Saperstein, and L. Westbrook, Scaling up: Representing gender diversity in survey research, Socius **2**, 1 (2016).
[70] C. G. Hart, A. Saperstein, D. Magliozzi, and L. Westbrook, Gender and health: Beyond binary categorical measurement, J. Health Soc. Behav. **60**, 101 (2019).
[71] Z. Nicolazzo, 'It's a hard line to walk': Black non-binary trans* collegians' perspectives on passing, realness, and trans*-normativity, Int. J. Qual. Studies Educ. **29**, 1173 (2016).
[72] J. Harrison-Quintana, J. M. Grant, and I. G. Rivera, Boxes of our own creation: A trans data collection wo/manifesto, Transgender Stud. Q. **2**, 166 (2015).
[73] N. Ingraham, V. Pratt, and N. Gorton, Counting trans* patients: A community health center case study, Transgender Stud. Q. **2**, 136 (2015).
[74] K. Schilt and J. Bratter, From multiracial to transgender? Assessing attitudes toward expanding gender options on the US Census, Transgender Stud. Q. **2**, 77 (2015).
[75] T. B. Singer, The profusion of things: The "transgender matrix" and demographic imaginaries in US public health, Transgender Stud. Q. **2**, 58 (2015).
[76] J. Butler, *Bodies that Matter: On the Discursive Limits of Sex* (Taylor & Francis, London, 2011).
[77] H. G. Gough, Identifying psychological femininity, Educ. Psychol. Meas. **12**, 427 (1952).

[78] M. Terman and C. C. Miles, *Sex and personality: Studies in masculinity and femininity* (McGraw-Hill, New York, 1936).
[79] A. Constantinople, Masculinity-femininity: An exception to a famous dictum?, Psychol. Bull. **80**, 389 (1973).
[80] S. L. Bem, The measurement of psychological androgyny, J. Consulting Clinical Psychol. **42**, 155 (1974).
[81] B. J. Risman, Gender as a social structure: Crossing disciplinary boundaries to advance science and equality, AG About Gender-Int. J. of Gender Studies **1**, 1 (2012).
[82] B. J. Risman, *Gender as a Social Structure* (Springer, New York, 2018).
[83] J. Butler, *Gender Trouble: Feminism and the Subversion of Identity* (Routledge, New York, 2011).
[84] L. Westbrook and K. Schilt, Doing gender, determining gender: Transgender people, gender panics, and the maintenance of the sex/gender/sexuality system, Gender Soc. **28**, 32 (2014).
[85] L. R. Miller and E. A. Grollman, The social costs of gender nonconformity for transgender adults: Implications for discrimination and health, *Sociological Forum* (Wiley Online Library, New York, 2015), p. 809.
[86] B. Brandth, Agricultural body-building: Incorporations of gender, body and work, J. Rural Studies **22**, 17 (2006).
[87] J. D. Smyth, A. Swendener, and E. Kazyak, Women's work? The relationship between farmwork and gender self-perception, Rural Sociol. **83**, 654 (2018).
[88] J. D. Smyth and K. Olson, Male/female is not enough: Adding measures of masculinity and femininity to general population surveys, *Understanding Survey Methodology: Sociological Theory and Applications*, edited by Philip S. Brenner (University of Nebraska, Lincoln, 2020), pp. 247–275.
[89] B. Davies and R. Harré, Positioning: The discursive production of selves, J. Theory Soc. Behav. **20**, 43 (1990).
[90] W. Faulkner, Doing gender in engineering workplace cultures. II. Gender in/authenticity and the in/visibility paradox, Eng. Stud. **1**, 169 (2009).
[91] R. P. Brown and E. C. Pinel, Stigma on my mind: Individual differences in the experience of stereotype threat, J. Exp. Soc. Psychol. **39**, 626 (2003).
[92] M. C. Murphy, C. M. Steele, and J. J. Gross, Signaling threat: How situational cues affect women in math, science, and engineering settings, Psychol. Sci. **18**, 879 (2007).
[93] A. A. Niler, R. Asencio, and L. A. DeChurch, Solidarity in STEM: How gender composition affects women's experience in work teams, Sex Roles **82**, 142 (2020).
[94] N. G. Holmes and C. E. Wieman, Introductory physics labs: We can do better, Phys. Today **71**, No. 1, 38 (2018).
[95] S. M. Aguillon, G.-F. Siegmund, R. H. Petipas, A. G. Drake, S. Cotner, and C. J. Ballen, Gender differences in student participation in an active-learning classroom, CBE Life Sci. Educ. **19**, ar12 (2020).
[96] PERTS Academic Mindsets Assessment (2020), https://www.perts.net/orientation/ascend (Accessed 3 February 2019).
[97] L. J. Cronbach, Coefficient alpha and the internal structure of tests, Psychometrika **16**, 297 (1951).
[98] R. B. Kline, *Principles and Practice of Structural Equation Modeling* (Guilford Publications, New York, 2015).
[99] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis: International Version* (Pearson, Upper Saddle River, NJ, 2010).
[100] J. Pallant, *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (Routledge, New York, 2020).
[101] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, Hillsdale, NJ, 1988).
[102] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Chapman and Hall/CRC, London, 2018).
[103] C. J. Ballen and S. Salehi, Mediation analysis in discipline-based education research using structural equation modeling: Beyond "what works" to understand how it works, and for whom, J. Microbiol. Biol. Educ. **22**, e00108-21 (2021).
[104] W. Wood and A. H. Eagly, Two traditions of research on gender identity, Sex Roles **73**, 461 (2015).
[105] A. J. Gonsalves, "Physics and the girly girl—There is a contradiction somewhere": Doctoral students' positioning around discourses of gender and competence in physics, Cult. Stud. Sci. Educ. **9**, 503 (2014).
[106] Y. L. Shek, Asian American masculinity: A review of the literature, J. Men's Studies **14**, 379 (2007).
[107] M. Ong, Body projects of young women of color in physics: Intersections of gender, race, and science, Soc. Probl. **52**, 593 (2005).
[108] J. M. Nissen, I. H. M. Horses, and B. Van, Dusen, Investigating society's educational debts due to racism and sexism in student attitudes about physics using quantitative critical race theory, Phys. Rev. Phys. Educ. Res. **17**, 010116 (2021).
[109] M. A. Weitekamp, The image of scientists in The Big Bang Theory, Phys. Today **70**, No. 1, 40 (2017).
[110] C. Frayling, *Mad, Bad and Dangerous?: The Scientist and the Cinema* (Reaktion Books, London, 2013).
[111] J. Webber, Masculine characteristics across race/ethnicity, gender, and income, Honors thesis, University of North Carolina, 2022.
[112] Y. J. Wong, A. J. Horn, and S. Chen, Perceived masculinity: The potential influence of race, racial essentialist beliefs, and stereotypes, Psychol. Men Masculinity **14**, 452 (2013).
[113] J. Chan, *Chinese American Masculinities: From Fu Manchu to Bruce Lee* (Taylor & Francis, London, 2001).
[114] E. Vazquez-Nuttall, I. Romero-Garcia, and B. D. Leon, Sex roles and perceptions of femininity and masculinity of Hispanic women: A review of the literature, Psychol. Women Q. **11**, 409 (1987).
[115] S. B. Zeff, A cross-cultural study of Mexican American, Black American, and White American women at a large urban university, Hispanic J. Behav. Sci. **4**, 245 (1982).