Mediating role of personality in the relation of gender to self-efficacy in physics and mathematics

Rachel Henderson,[1,2] Dona Hewagallage,[3] Jake Follmer,[4] Lynnette Michaluk,[5]
Jessica Deshler,[6] Edgar Fuller,[7] and John Stewart[3,*]

[1]Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824, USA
[2]CREATE for STEM Institute, Michigan State University, East Lansing, Michigan 48824, USA
[3]Department of Physics and Astronomy, West Virginia University, Morgantown, West Virginia 26506, USA
[4]Department of Counseling and Learning Sciences, West Virginia University, Morgantown, West Virginia 26506, USA
[5]West Virginia University Center for Excellence in STEM Education, West Virginia University, Morgantown, West Virginia 26506, USA
[6]Department of Mathematics, West Virginia University, Morgantown, West Virginia 26506, USA
[7]Department of Mathematics, Florida International University, Miami, Florida 33199, USA

(Received 5 February 2022; accepted 6 May 2022; published 3 June 2022)

Self-efficacy has emerged as one of the most important noncognitive variables explaining academic behavior. It has been shown to influence students' academic and career decisions as well as their academic performance. Multiple studies have reported differences in self-efficacy between men and women in science, technology, engineering, and mathematics classes. A student's personality, characterized by the five-factor model, is also related to academic performance; some personality facets are substantially different for men and women. This work examines the relations among the five-factor model of personality (agreeableness, conscientiousness, extraversion, neuroticism, and openness), self-efficacy toward physics and mathematics, and course outcomes in university physics and mathematics classes. Women reported significantly higher neuroticism in all classes, a medium to large effect size, and significantly higher conscientiousness in Calculus 1 and Physics 1, small effects. Men reported higher self-efficacy in two-semester Calculus 1, one-semester Calculus 1, Physics 1, and Physics 2, small effects. Conscientiousness and neuroticism had competing mediational effects on the relation of gender to self-efficacy. The path through neuroticism accounted for 25%–47% of the total effect of gender on self-efficacy (increasing self-efficacy for men) and the path through conscientiousness accounted for 12%–23% of the total effect (increasing self-efficacy for women). Self-efficacy mediated the relation of conscientiousness to course grade in all classes, accounting for 30%–45% of the total effect.

DOI: 10.1103/PhysRevPhysEducRes.18.010143

I. INTRODUCTION

Approximately one-half of high school physics students are women [1]; yet, women earned only 24% of physics undergraduate degrees in 2020 [2,3]. Physics education research (PER) has long investigated issues that may influence diversity and inclusion in physics classes. Historically, much of this research has explored differences in conceptual learning between men and women [4,5] while some recent studies have explored differences between other demographic groups [6–9]. While the existence of differences in performance between demographic groups is well established, the reason for the differences is still an active area of research. PER has more recently begun to explore the effect of noncognitive variables such as self-efficacy and its relation to physics outcomes [10–14]. The relation of noncognitive factors and educational outcomes is a broad area of non discipline-based educational research and substantial evidence exists for the relation of a number of noncognitive factors to educational outcomes [15–22]. This study examines the relations between two of the most studied noncognitive factors: self-efficacy and the five-factor model of personality.

Despite substantial research strands identifying differences in self-efficacy of men and women [23–27], the role of science or mathematics anxiety in achievement differences [28–33], and the differences in personality by gender [34], we could identify only one research study exploring the relationships among gender, self-efficacy, and personality [35]; no studies specifically examining STEM students were identified.

[*]jcstewart1@mail.wvu.edu

A. Research questions

The relations among gender, personalty, self-efficacy, and physics course grade are explored within the framework of mediation and moderation. One variable mediates the relation between two other variables if part of the effect of one variable on the other is explained by the mediating variable. A variable moderates the relation of two variables if the relation is different depending on the value of the moderator. Mediation and moderation is explained in detail in Sec. II C. This study seeks to answer five research questions.

RQ1: How does self-efficacy or personality differ for men and women in core university introductory mathematics and physics classes?

RQ2: Does gender moderate the relationships of personality, self-efficacy, and achievement?

RQ3: Does personality mediate the relationship of gender to self-efficacy? If so, how does it mediate the relationship?

RQ4: Does personality mediate the relationship of gender to achievement? If so, how does it mediate the relationship?

RQ5: Does self-efficacy mediate the relationship of personality and gender to achievement? If so, how does it mediate the relationship?

Considerable effort has been directed toward understanding and mitigating performance differences between students who have been historically marginalized in science, technology, engineering, and mathematics (STEM) and those who are not [6,36]. The use of cognitive factors such as high school GPA, ACT, or SAT scores, conceptual pretest or post-test scores, and test averages are prominent in PER. Less work has investigated noncognitive factors in PER—factors that do not directly measure academic achievement at some level. Richardson *et al.* [37] provided an extensive overview of the relations among noncognitive factors and academic achievement in the broader educational research literature.

The majority of this work has examined characteristics of instruction as one factor that may co-explain disparities in representation and performance in physics [10]. Much less work has focused on the roles of students characteristics in explaining female students pursuit of a STEM major or a STEM degree. Noncognitive factors could be particularly important for understanding the underrepresentation of some groups in STEM [38] as well as the retention of more students within various STEM majors. Self-efficacy has long been an important variable in models predicting college persistence [39].

B. Self-efficacy

Self-efficacy has been shown to be an important factor in academic goal setting, motivation, performance, and persistence. Bandura defined self-efficacy as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives" [40,41]. Students self-efficacy is grounded in their beliefs about their performances, their vicarious experiences, and forms of persuasion derived from teachers, parents, and others [42]. As such, self-efficacy towards physics represents a student's belief they can successfully perform physics tasks such as homework, examinations, and laboratory activities. These beliefs are informed by prior success in physics and mathematics classes and success on prior assignments in the same class (mastery experiences), observing others successfully perform physics tasks (vicarious experiences), and receiving support from peers or instructors that they believe the student can successfully complete physics tasks (social persuasion). Researchers have shown self-efficacy to be highly correlated with student performance and achievement in science courses [15,17,37,43,44], students' career aspirations [20–22,45], as well as their persistence within STEM disciplines [16,18,19,46,47]. Students self-efficacy beliefs also influence students' academic choices, the amount of effort they expend toward their studies, how they address and persist in the face of academic challenges, and the ways in which they navigate academic stressors as they engage with a task [27].

Shortly after the introduction of the construct, differences in self-efficacy between men and women began to be reported [23–27]. Women, although generally more successful in science classes, report lower self-efficacy toward science than their male counterparts [48]. Explanations for the variation in self-efficacy beliefs by gender have included considerations based on the manner in which efficacy beliefs are assessed, the influence of gender orientation and stereotyped beliefs about gender, and cumulative impact of differences in expectations for performance by gender [27,49].

1. Self-efficacy in physics

While there have been robust studies performed within various STEM domains, recent research has explored the relation between self-efficacy and physics instruction [10,12,50,51]. Nissen analyzed students' self-efficacy in high-school classrooms. Men reported higher self-efficacy toward physics than women and, after physics instruction, the self-efficacy of women decreased more than men [52]. Marshman *et al.* [10] also found that women reported lower levels of self-efficacy as well as lower value associated with physics after completing an introductory physics course.

Significant gender differences in the self-efficacy of non STEM majors have also been reported in physics classes. Cavallo *et al.* showed that men reported higher self-efficacy than women in an introductory, algebra-based physics course for students majoring in the biological sciences [53]. Lindstrøm and Sharma investigated differences in self-efficacy in a course designed for students without any prior physics instruction; the gender differences in self-efficacy grew from pre-instruction to post-instruction [54].

Other studies have explored gender differences in physics classrooms required for students majoring in STEM. Women, on average, reported lower self-efficacy toward physics-related activities [50,55–57]. In general, from pretest to post-test, women's self-efficacy toward physics decreased more than men's which, in turn, increased the gender gap in self-efficacy over the course of the semester. Furthermore, Sawtelle *et al.* showed that traditional instruction negatively impacted all students' self-efficacy but students' engagement in a Modeling Instruction course positively impacted specific sources of women's self-efficacy [58]. Specifically, it was shown that the vicarious learning source of self-efficacy was positively related to the success of women in these courses; however, the mastery experience and social persuasion sources did not explain any variation in their models. In addition to Modeling Instruction, Miller *et al.* showed that in a class using peer instruction, the lower self-efficacy reported by women prior to instruction fully mediated the effect of female students switching their answers to conceptual physics problems from right to wrong [59].

More recently, researchers have explored gender differences in self-efficacy in relation to how students perceive their own class achievement [12] and what harm these differences have on female students [11]. In general, men and women interpret grades and other measures of STEM performance differently, which may lead to a confidence gap between men and women [60] and, thus, longer term impacts such as persistence toward a degree or career.

2. Self-efficacy in mathematics

In addition to studies done within the context of the physics classroom, researchers have investigated students' self-efficacy in mathematics courses. In a meta-analysis, Huang demonstrated that men reported higher self-efficacy than women toward college mathematics (Hedge's $g = 0.18$)—a similar result was reported for both physics and computer science [24]. The effect size criteria for Hedge's g are similar to Cohen's d; 0.2 is a small effect, 0.5 a medium effect, and 0.8 a large effect. While the results in physics show men consistently reporting higher levels of self-efficacy toward the discipline than women, in mathematics, some studies do not find significant differences in self-efficacy between men and women [61,62]. Further, in-depth examination of students self-efficacy in mathematics contexts suggests that students' efficacy beliefs fluctuate over time and across engagement with mathematics tasks [63].

3. Self-efficacy in other STEM disciplines

Self-efficacy has also been studied in other STEM disciplines, including engineering, biology, and chemistry. In engineering, many studies concluded that men, on average, report higher self-efficacy than women [64–73]. Fewer studies have investigated self-efficacy in chemistry and biology. One study of high-school students showed no difference in self-efficacy toward biology [74]; however, Ainscough *et al.* demonstrated that men reported a higher level of self-efficacy toward biology at the beginning and end of a first-year biology course [75]. In college chemistry, some studies have shown that women report higher self-efficacy toward chemistry tasks than men [19,76].

C. Personality

This work utilizes the five-factor model of personality which characterizes personality along five dimensions: agreeableness, conscientiousness, extraversion, neuroticism, and openness [77–79]. Agreeableness is related to an individual's tendency to be cooperative and compassionate. Conscientious individuals are organized, focused, and careful. Extraversion is related to a desire for social interaction and attention. Neuroticism is associated with feeling anxiety or other strong emotions. Neuroticism is sometimes reverse coded as "emotional stability." Openness (to experience) involves a willingness to embrace new ideas and experiences. This work measures the five-factor model using the big five inventory (BFI) [80,81]. A considerable body of research has shown strong relationships among facets of the five-factor model and academic performance [82]; however, little research has investigated students at the college level in STEM fields. While personality does change with time, this change occurs slowly at the average age of students in this study [34].

A large nonacademic study ($N > 10^6$) reported differences in some personality facets between men and women [34]. These differences were particularly large for the neuroticism facet, with women reporting 58 out of 100 on a percent of maximum possible (POMP) scale while men reported 46. A POMP scale projects the 5-point Likert scale into the range from 0 to 100. If these values are converted to a 5-point Likert scale, women report a neuroticism of 3.3 and men 2.8. As such, personality, and specifically neuroticism, may be an important construct to explain some of the gender differences reported in other constructs such as self-efficacy. It may also be important to the understanding of a large body of research reporting academic achievement differences between men and women, with women generally outperforming men [48,83]. A 2012 meta-analysis, for example, showed both conscientiousness and self-efficacy to be positively correlated with academic achievement ($r = 0.19$ and $r = 0.31$, respectively, where r is the correlation coefficient). Neuroticism was slightly negatively correlated with achievement ($r = -0.01$) [37].

1. Self-efficacy and personality

A number of studies have explored the relationship between personality and self-efficacy [84,85], with some studies exploring differences between men and women. In a prior study, the five personality facets were measured to have different regression coefficients, β, predicting self-efficacy: neuroticism $\beta = -0.25$, extraversion $\beta = 0.27$, openness $\beta = 0.13$, agreeableness $\beta = -0.06$, and conscientiousness $\beta = 0.18$ [86]. Self-efficacy has been shown to mediate the relationship between engagement in studying and openness to experience [87].

Prosocial behaviors are behaviors which benefit others; prosociality is the tendency to engage in those beneficial behaviors. A significant correlation between agreeableness and prosociality, which differed between men and women (men $r = 0.42$, women $r = 0.33$), has also been reported [88]. Self-efficacy and prosociality were also correlated (men $r = 0.57$, women $r = 0.35$). Self-efficacy was shown to mediate the relationship between agreeableness and prosociality, explaining 30% of the relation for men and 19% for women.

A significant positive interaction between gender and emotional stability (reverse-coded neuroticism) was reported in these two variables' effect on self-efficacy [35]. Further, in a meta-analysis examining the correlates with complex task performance in the workplace, both conscientiousness and cognitive ability were positively correlated with complex task performance; conscientiousness was also positively correlated with self-efficacy ($r = 0.27$). Self-efficacy mediated the relation between conscientiousness, cognitive ability, and simple task performance but was not a mediator for complex task performance [89].

Studies have also examined links among students' self-regulation and self-efficacy skills, personality, motivation, and achievement. In an examination of profiles of college students, Dörrenbächer and Perels found that students' low or high in self-regulated learning varied in dimensions of motivation and personality [90]. Specifically, achievement was found to be significantly higher among students high in both self-regulated learning and motivation. Importantly, students who are higher in self-regulated learning skills, such as self-efficacy, also demonstrated lower neuroticism and higher conscientiousness, agreeableness, and openness to experience, suggesting key links among students' regulation, efficacy, and dimensions of personality.

D. Science and mathematics anxiety

Both mathematics anxiety [28,29] and science anxiety [30–33] have been explored as explanations for differences in quantitative examination performance. In a meta-analysis, higher levels of mathematics anxiety were negatively correlated with performance ($r = -0.27$); this result was independent of gender [28]. Differences in mathematics anxiety for men and women ($d = 0.28$) have been shown to be of the same effect size as differences in self-efficacy ($d = 0.33$); both differences are substantially larger than academic performance differences ($d = 0.11$). Effect size is characterized by Cohen's d. STEM majors report lower levels of science anxiety than nonscience majors [91]; however, within STEM, women report higher levels of science anxiety than men.

Sources of anxiety within the physics classroom have also been investigated. Students with higher communication apprehension produced lower normalized gains on the Force Concept Inventory [92,93]. Students in classes where the physics instructor allowed more autonomy had less anxiety about the course and higher achievement in the course [94].

E. Theoretical framework

In this research, we adopt Bandura's social cognitive theory (SCT) as the primary theoretical framework for this study [40]. SCT proposes a recursive relationship between task achievement, goal setting, and self-efficacy, producing a construct that evolves in time due to external feedback and influences how an individual addresses future goals. For academic self-efficacy, one of the primary sources of performance feedback is academic achievement in classes. For STEM students, performance on course examinations often form a substantial part of overall course grades. A substantial literature suggests experiencing stress or anxiety during an examination degrades performance; thus, the personality facet neuroticism may be related to examination performance and affect self-efficacy by modifying examination performance. Science and mathematics anxiety represent anxiety specifically experienced during mathematics and science experiences, often examinations, and are different from the general tendency to feel anxiety measured by the neuroticism facet. However, it seems reasonable to hypothesize that students who are more likely to feel anxiety in general are also more likely to feel anxiety toward mathematics and science experiences.

Conversely, additional anxiety might make a student prepare for the examination more thoroughly, potentially increasing performance and, later, self-efficacy. Course grades in STEM classes also generally require the completion of assignments, such as homework, in addition to examinations. The conscientiousness facet may influence whether a student consistently completes assigned tasks, thus affecting homework and other assignment grades. Homework is generally designed to affect test performance, further suggesting conscientiousness may influence overall class grade.

Developing a model for the relation of self-efficacy to academic achievement is challenging because a college STEM student's self-efficacy towards STEM academic situations has been under development for a decade before they enter an introductory physics or mathematics class. According to Bandura's model, a student's current self-efficacy should be informed by their history of both prior academic achievement and prior levels of self-efficacy as self-efficacy is adjusted according to performance feedback. As such, current self-efficacy is partially a measure of past academic achievement which naturally affects future grades. Self-efficacy at a certain time is then properly modeled as a student's current belief in their capability to perform some action as well as their interpretation of the past experiences which inform that belief.

F. Summary

The connections between gender, personality, self-efficacy, and achievement are complex as detailed above. Of the myriad relations discussed earlier, the following most directly impact this work. Many studies have identified differences in self-efficacy by gender in STEM fields. Self-efficacy is a strong correlate to academic achievement. Gender differences in personality have also been identified in a very large non STEM study. Personalty facets, particularly conscientiousness, also correlate with academic achievement. Multiple studies have shown a negative correlation with STEM achievement and the tendency to feel anxiety in STEM testing situations.

II. METHODS

A. Sample

This study was performed from fall 2015 to fall 2019 at a large land-grant university in the eastern United States. The university's general undergraduate population reported ACT scores ranging from 21 to 26 (25th to 75th percentile) [95]. The undergraduate demographic composition was 80% White, 6% international, 4% African American, 4% Hispanic, 4% students reporting two or more races, 2% Asian, and other groups each with 1% or less [95].

This study includes data from two introductory physics classes (Physics 1 and Physics 2) and three introductory calculus classes (Calculus 1, Calculus 1A, and Calculus 1B). Calculus 1 is the traditional one semester calculus 1 class. Science and engineering students who are "math ready" enroll in Calculus 1 their first semester. Calculus 1A and Calculus 1B are a two-semester sequence which covers the material of Calculus 1 along with some precalculus and is designed for students who are not academically ready to take Calculus 1. Physics 1 is the introductory calculus-based mechanics course taken by scientists and engineers and has Calculus 1 or Calculus 1A as its prerequisite. Physics 2 is the introductory calculus-based electricity and magnetism course and has Physics 1 as its prerequisite. Data were also collected in Workshop Mathematics, a remedial mathematics class to help students prepare to take college algebra. Over 90% of the students in the calculus and physics classes were pursuing STEM majors, while only 70% of the students in Workshop Mathematics were pursuing STEM majors. Workshop students represent a population more representative of the university in general.

The mathematics classes were taught by many instructors over the course of the study. Many of these instructors used a variety of active learning strategies to support their students. Each physics course was led by a single instructor with an expertise in PER over the course of the study. The lectures used clicker questions and research-based pedagogy to engage students. Multiple other instructors team taught the courses and adopted this pedagogy; the courses offered multiple lecture sections per semester. Both physics classes had a 3-h per week required lab which used small group problem solving, white boarding, and hands-on inquiry based activities.

A total of 6286 students completed the physics classes from Fall 2015 to Fall 2019 and a total of 8937 students completed the mathematics classes from Fall 2016 to Spring 2019. Of these, only domestic students with ACT or SAT scores who completed both the personality and self-efficacy surveys were retained. For physics, 3334 students met these criteria (1783 Physics 1 and 1551 Physics 2). For mathematics, 3977 students met these criteria (1074 Workshop Mathematics, 765 Calculus 1A, 563 Calculus 1B, and 1575 Calculus 1) completed both surveys. These students form the sample for this study.

B. Instruments

The big five inventory (BFI) [80,81,96,97] was used to measure students' personality. The BFI measures the five-factor personality model based on the following facets: agreeableness, conscientiousness, extraversion, neuroticism, and openness. The BFI has been extensively used in a broad set of studies [98]. This work focuses on the conscientiousness and neuroticism facets.

Neuroticism is related to the tendency to feel stress, anxiety, or other strong emotions. The BFI measures neuroticism using eight items measured on a five-point Likert scale. The student is asked to rate how true the statement is for them; some items are reversed coded. The items are

- Is depressed, blue
- Is relaxed, handles stress well (reversed)
- Can be tense
- Worries a lot
- Is emotionally stable, not easily upset (reversed)
- Can be moody
- Remains calm in tense situations (reversed)
- Gets nervous easily

Conscientiousness is related to the tendency to follow instructions, to work hard and carefully, and to meet outside expectations. The items in the conscientiousness scale in the BFI are

- Does a thorough job
- Can be somewhat careless (reversed)
- Is a reliable worker
- Tends to be disorganized (reversed)
- Tends to be lazy (reversed)
- Perseveres until the task is finished
- Does things efficiently

- Makes plans and follows through with them
- Is easily distracted (reversed)

Self-efficacy was measured with the self-efficacy for learning and performance subscale from the motivated strategies for learning questionnaire (MSLQ) [99]. This eight-item scale was reduced to six items and specialized to the class environment by specifying either physics or mathematics classes [100]. The scale was reduced to remove one item asking about reading comprehension and one item that was very similar to a second item as part of a larger project to measure self-efficacy in multiple STEM domains. The resulting physics self-efficacy subscale is

- I believe that I will receive an excellent grade in this physics class.
- I'm confident I can understand the basic concepts taught in this physics class.
- I'm confident I can understand the most complex material presented by the instructor in this physics class.
- I'm confident I can do an excellent job on the assignments and tests in this physics class.
- I'm certain I can master the skills being taught in this physics class.
- Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this physics class.

The modified survey items were extensively revalidated. The method of word substitution to modify the MSLQ for specific domains has been used in prior studies [70].

Both of the surveys were administered once per semester and the students received a small amount of course credit for completing each survey. Informed consent was collected from all participants and all procedures were approved by the Institutional Review Board.

Academic achievement was characterized by physics and mathematics course grades measured on a numeric scale with $\mathrm{F} = 0$ and $\mathrm{A} = 4$. Academic preparation was measured by ACT and SAT mathematics percentile scores (ACTM%). Gender was accessed from university records where it was recorded as a binary variable. This work coded gender as a dichotomous variable with levels 0 (women) and 1 (men). This coding of gender is not optimal, but is consistent with other work in PER. For a more nuanced discussion of gender, see Traxler *et al.* [101].

C. Mediation and moderation

Mediation and moderation form a powerful framework for investigating the relationships between variables affecting educational achievement. To investigate the relation of personality facets and self-efficacy with achievement, the mediation framework developed by Baron and Kenny [102] was used. Figure 1 represents the mediational model for the relations of the dependent variable (Dep), independent variable (Indep), and the mediator (Med). The dependent variable of each regression is the node at the tail of the directed line; the independent variables for the regression are all nodes at the head of lines directed at the dependent variable.

The total effect, C, is measured through the regression

$$\mathrm{Dep} = \beta_1 + C \times \mathrm{Indep} + \epsilon_1, \tag{1}$$

where β_i is the intercept and ϵ_i is the residual error.

With the mediator, Dep is predicted through two paths: the direct path characterized by C' and the indirect path through the mediator composed of a path from Indep to Med (A) and the path from Med to Dep (B). These parameters are measured by

$$\mathrm{Med} = \beta_2 + A \times \mathrm{Indep} + \epsilon_2. \tag{2}$$

$$\mathrm{Dep} = \beta_3 + C' \times \mathrm{Indep} + B \cdot \mathrm{Med} + \epsilon_3. \tag{3}$$

Significant mediation exists if A, B, and C are significant regression coefficients and if the direct effect C' is less than the total effect C. If $C' < C$, part of the overall effect of Indep on Dep is a result of the relation of both variables with the mediator. To further test for significant mediation, the total indirect effect ($A \times B$) was calculated by bootstrapping with 1000 replications. The mediation is significant if a t tests shows this product is significantly different than zero. The total effect of the independent variable on the dependent variable C is thus partitioned into two parts: one resulting from the mediator ($A \times B$) and one not resulting from the mediator (C'). The total effect can be expressed as a sum of these two contributions, $C = C' + A \times B$. The percentage of the total effect C which is the result of the mediator is then $A \times B/C \times 100\%$.

Figure 1 uses the notation $C \to C'$ to represent the total effect C of Indep on Dep without the mediator and the remaining direct effect C'. This notation is intended to represent the change in the C coefficient to C' after adding the mediating variables. This relation is more traditionally represented with the notation $C(C')$; however, the more traditional representation does not naturally generalize to the multistage mediation analysis presented in this work.

Moderation occurs when one variable, the moderator (Mod), influences the relation between two other variables. For example, it may be that the relation of self-efficacy to course grade is different for men and women; in this case,

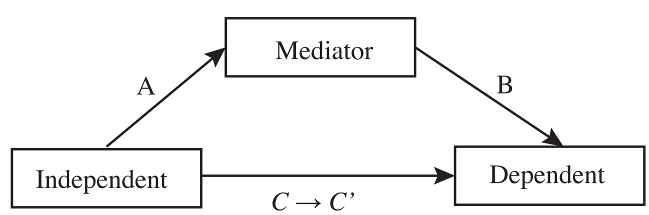

FIG. 1. Mediation process.

gender may moderate the relationship between self-efficacy and course grade. To detect moderation, the moderator is added to the regression equation as a product term as

$$\mathrm{Dep} = \beta_4 + \beta_5 \times \mathrm{Indep} + D \times \mathrm{Indep} \times \mathrm{Mod} + \epsilon_4. \quad (4)$$

The moderation is significant if D is significant. If the moderator is dichotomous, the effect of moderation is to produce different slopes, β_5 and $\beta_5 + D$, depending on the level of the moderator (0 or 1, respectively).

III. RESULTS

Table I shows the descriptive statistics for each course. The effect size difference between men and women for each variable is measured by Cohen's d. For each variable, a t test was performed to determine if the difference between men and women was significant; the result of the t test is presented as a superscript on d. A Bonferroni correction was applied to each of Tables I–IV individually to correct for the inflation of type I error. The significance threshold was divided by the number of statistical tests performed in the table. For example, in Table I the p threshold was divided by 36 for the 36 statistical tests performed in the columns from self-efficacy to openness, which are the focus of this work. The grade and ACTM% columns are presented for reference and as a general measure of differences in the student populations of each class.

The data presented have some striking features. In all courses, women report higher levels of neuroticism with effect sizes ranging from $d = 0.64$ to 0.80, from a medium to a large effect. The 0.4–0.6 Likert point difference on the neuroticism scale was very consistent between the physics and calculus classes. In addition, the Workshop Mathematics class, which is taken by a population of students with a lower percentage intending on majoring in STEM and students with less mathematics preparation (as measured by ACTM %), shows a similar gender difference. As such, it seems likely that these differences represent a general feature of college age students, not a specific feature of students pursuing physical science and engineering majors.

TABLE I. Descriptive statistics. Cohen's d measures the effect size for the difference between men and women for each quantity. The significance of a t-test of the difference between men and women is shown as a superscript on d. Note: "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$. A Bonferroni correction was applied to the significance levels.

	N	ACTM%	Grade	Self-efficacy	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
					Workshop mathematics				
Men	752	41.0 ± 12.5	2.9 ± 1.2	3.9 ± 0.8	3.8 ± 0.5	3.6 ± 0.6	3.2 ± 0.7	2.8 ± 0.7	3.5 ± 0.5
Women	322	39.6 ± 11.7	2.9 ± 1.3	3.9 ± 0.9	3.9 ± 0.5	3.6 ± 0.6	3.3 ± 0.7	3.3 ± 0.8	3.4 ± 0.5
d		0.12	0.00	0.01	0.30^c	0.05	0.16	0.70^c	0.18
					Calculus 1A				
Men	387	71.2 ± 16.4	2.6 ± 1.2	3.9 ± 0.8	3.8 ± 0.6	3.6 ± 0.6	3.2 ± 0.7	2.7 ± 0.7	3.6 ± 0.5
Women	378	67.6 ± 17.3	2.5 ± 1.3	3.7 ± 0.9	3.9 ± 0.6	3.7 ± 0.6	3.2 ± 0.8	3.3 ± 0.8	3.5 ± 0.6
d		0.21	0.05	0.27^b	0.18	0.18	0.04	0.69^c	0.05
					Calculus 1B				
Men	304	72.6 ± 16.8	2.7 ± 1.1	3.9 ± 0.8	3.7 ± 0.6	3.7 ± 0.6	3.2 ± 0.7	2.7 ± 0.7	3.6 ± 0.5
Women	259	70.1 ± 16.4	2.8 ± 1.1	3.8 ± 0.9	3.9 ± 0.6	3.8 ± 0.6	3.2 ± 0.8	3.2 ± 0.7	3.5 ± 0.6
d		0.15	0.05	0.13	0.22	0.14	0.00	0.80^c	0.04
					Calculus 1				
Men	1020	83.8 ± 12.9	2.3 ± 1.3	3.9 ± 0.8	3.8 ± 0.6	3.6 ± 0.6	3.2 ± 0.8	2.7 ± 0.7	3.6 ± 0.5
Women	555	82.2 ± 12.8	2.6 ± 1.2	3.6 ± 1.0	3.9 ± 0.6	3.7 ± 0.6	3.3 ± 0.8	3.2 ± 0.8	3.6 ± 0.6
d		0.12	0.21^c	0.28^c	0.11	0.19^a	0.10	0.73^c	0.07
					Physics 1				
Men	1284	80.1 ± 15.3	2.9 ± 1.0	3.9 ± 0.8	3.8 ± 0.6	3.7 ± 0.6	3.2 ± 0.7	2.6 ± 0.7	3.6 ± 0.5
Women	499	81.3 ± 14.6	2.9 ± 1.0	3.6 ± 1.0	3.9 ± 0.6	3.9 ± 0.6	3.3 ± 0.8	3.1 ± 0.7	3.7 ± 0.6
d		0.08	0.01	0.36^c	0.24^c	0.32^c	0.07	0.70^c	0.08
					Physics 2				
Men	1186	81.5 ± 15.1	2.9 ± 1.0	3.9 ± 0.8	3.8 ± 0.6	3.7 ± 0.6	3.2 ± 0.7	2.7 ± 0.7	3.6 ± 0.5
Women	365	83.5 ± 12.8	3.1 ± 1.0	3.8 ± 0.9	3.9 ± 0.6	3.8 ± 0.6	3.2 ± 0.7	3.1 ± 0.8	3.7 ± 0.6
d		0.14	0.15	0.23^a	0.20	0.19	0.07	0.64^c	0.01

Women also consistently report higher levels of conscientiousness in calculus and physics classes with effect sizes for the differences ranging from below a small effect $d = 0.14$ to a small effect $d = 0.32$; however, these differences were significant only in Calculus 1 and Physics 1.

Men report higher levels of self-efficacy toward their calculus and physics classes with differences ranging from an effect size of 0.13–0.36 with the difference in the range of a small effect in all calculus and physics classes except Calculus 1B. The differences were statistically significant in all calculus and physics classes except Calculus 1B. Self-efficacy toward the remedial Workshop Mathematics class was approximately equal for men and women. Men reported the same level of self-efficacy as that reported in the more challenging calculus and physics classes. Women in Workshop Mathematics reported higher levels of self-efficacy than women in the calculus and physics classes. This may have resulted from the class being fairly easy with simply completing assignments all that was required for a passing grade.

Overall, the averages of the personality facets for men and women were strikingly similar for classes requiring substantially different high school preparation and bridging the first two years of college. Significant differences between men and women were also measured for the agreeableness facet in Workshop Mathematics and Physics 1 with women reporting higher levels of this facet.

The sample contains two two-class course sequences: Calculus 1A and 1B and Physics 1 and 2. For both sequences, the difference between the self-efficacy of men and women was smaller for the second class in the course sequence. The self-efficacy of women increased between the first class in the sequence and the second class. This change could be the result of either increased self-efficacy in women with longer exposure to physics and mathematics. It could also be caused by women with lower self-efficacy choosing not to enroll in the second class in the sequence.

To further investigate this effect, a paired sample was extracted consisting of students who had taken both courses in the course sequence. For Physics 1 and Physics 2, 865 students took both courses (men 635, women 230); the self-efficacy for women did not significantly increase from Physics 1 (3.7 ± 1.0) to Physics 2 (3.8 ± 0.9). The effect size of this difference is $d = 0.24$. The self-efficacy for men also did not increase from Physics 1 (4.0 ± 0.8) to Physics 2 (4.0 ± 0.8); the effect size is $d = 0.10$. A similar matched sample was extracted for the Calculus 1A and 1B sequence ($N = 312$, men 157, women 155). The self-efficacy of women did not significantly increase from Calculus 1A (3.8 ± 0.9) to Calculus 1B (3.8 ± 0.9), $d = 0.08$; the self-efficacy of men also did not significantly change from from Calculus 1A (4.0 ± 0.7) to Calculus 1B (3.9 ± 0.8), $d = 0.21$. This result was similar to the finding of Cwik and Singh who reported a consistent self-efficacy at the beginning and at the end of the course [103].

A. Full path model

Figure 2 shows two possible path models for the relation of gender, the personality facets conscientiousness and neuroticism, self-efficacy, and physics course grade. Only conscientiousness and neuroticism were examined because of the significant differences observed in Table I, prior work relating these variables to academic achievement, and for the theoretical reasons discussed in Sec. I E. Model 2 contains an additional element; the curved line between self-efficacy and grade represents the correlation between these variables. Treated as structural equation models (SEM), these path models are mathematically equivalent. In SEM, reversing the direction of an edge or replacing an edge with a correlation does not change the overall fit of the model. Two models are presented to allow the investigation of different assumptions of the relation of self-efficacy to grade. A path model encodes the relational hypotheses of the researcher.

In both models, our relational hypothesis is that gender influences personality which in turn influences self-efficacy. This hypothesis is supported by the consistent gender difference in the neuroticism facet observed in both STEM and non STEM samples, that gender identity usually develops prior to adult personality, and that personality develops generally prior to STEM self-efficacy. It is also supported by causal arguments relating higher anxiety or conscientiousness to academic performance which informs the development of self-efficacy.

For model 1, we additionally assume self-efficacy influences grade, but that personality also influences grade directly and through its effect on self-efficacy. This model is supported theoretically by the generally positive effect of believing one can do something on actually doing that thing. It is also supported temporally; self-efficacy is measured midsemester while grades are assigned at the end of the semester.

For model 2, we discard the hypothesis that changes in self-efficacy imply changes in grades, and relax it to the assumption that self-efficacy covaries with grades. This model can be theoretically justified by observing self-efficacy is related to prior course success which should influence grades. Modifying self-efficacy should influence future grades; however, modifying self-efficacy will not modify prior academic performance which also is related to course grades.

In models 1 or 2, personality could affect self-efficacy either by modifying how past experience is processed into current beliefs, by modifying the past experiences themselves, or by doing both. The effect of personality on self-efficacy through either mechanism has been acting for many years; as such, the structure in models 1 and 2 relating gender, personality, and self-efficacy should be viewed as a cumulative effect acting over many years.

The central difference between models 1 and 2 is that in model 1 changes in self-efficacy should produce changes in

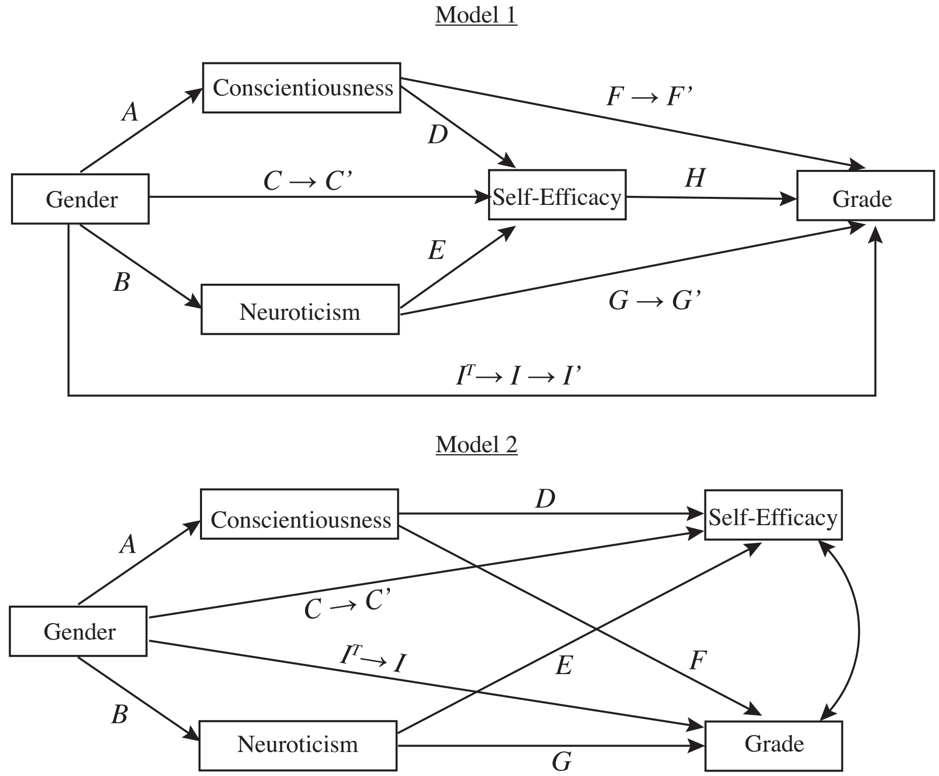

FIG. 2. Path models showing the relation of gender, personality, self-efficacy, and physics course grade. The notation $C \to C'$ shows the change in the coefficient relating gender to self-efficacy before the addition of mediating variables (C) and after the addition of these variables (C'). The notation $I^T \to I \to I'$ shows the relation between gender and grade without mediators (I^T), with only personality as mediating variables (I), and with both personality and self-efficacy as mediating variables (I').

course grades while in model 2 differences in self-efficacy should be related to differences in grades. There are two possible interpretations of model 1. The first views both self-efficacy and grade as quantities measured at a single instance in time and suggests that if some intervention could improve self-efficacy prior to the end of the course then this change would have an effect on the current course grade (as well as future course grades) suggested by the coefficient of the path model. We argue that this interpretation is unlikely (particularly given the size of the measured coefficients). An intervention to modify self-efficacy at a given time would not change the past experiences (academic achievement) which informed self-efficacy. A second interpretation acknowledges the recursive nature of Bandura's model and views both self-efficacy and academic achievement (measured by a college course grade) as variables that have developed over time. In this view, personality has influenced self-efficacy over the student's development which has in turn influenced their general academic achievement.

This work investigates three potential mediational relationships shown in model 1; two of these relations are also present in model 2. The first relation explores the mediation of the combination of conscientiousness and neuroticism of the relationship between gender and self-efficacy. This mediation model is composed of edges A, B, $C \to C'$, D, and E. The notation $C \to C'$ indicates that the coefficient C representing the total effect of gender on self-efficacy is changed to C' by the action of the mediating variables. The second mediational relation investigates whether the relation of gender to grade is mediated by personality. This model is composed of edges: A, B, F, G, and $I^T \to I$. The third possible mediating relationship shown only in model 1 investigates whether self-efficacy mediates the relation of gender and personality to grade; this model requires the full path model 1. These three analyses will be discussed in Secs. III C–III E.

The path models in Fig. 2 were analyzed with traditional multiple linear regression analysis. It could also be analyzed as a structural equation model (SEM) that yields the same results (it is a saturated or just-identified model so all model fit statistics are perfect). For the SEM model, the two personality facets are assumed to co-vary.

The coefficients in the path models were estimated using a series of linear regressions shown in Eqs. (5)–(11). For all analyses that follow, all continuous variables have been normalized by subtracting the mean of each variable and dividing by the standard deviation.

The regression coefficients for the mediation of personality on the relation of gender to self-efficacy were computed using the following equations, where SEF is self-efficacy, Nrt is neuroticism, Cns is conscientiousness, β_i^0 is the intercept, and ϵ_i the residual error:

$$\mathrm{SEF} = \beta_1^0 + C \times \mathrm{Gender} + \epsilon_1, \tag{5}$$

$$\mathrm{Cns} = \beta_2^0 + A \times \mathrm{Gender} + \epsilon_2, \tag{6}$$

$$\mathrm{Nrt} = \beta_3^0 + B \times \mathrm{Gender} + \epsilon_3, \tag{7}$$

$$\mathrm{SEF} = \beta_4^0 + C' \times \mathrm{Gender} + D \times \mathrm{Cns} + E \times \mathrm{Nrt} + \epsilon_4. \tag{8}$$

The overall effect of gender on course grade (I^T) is estimated by

$$\mathrm{Grade} = \beta_5^0 + I^T \times \mathrm{Gender} + \epsilon_5. \tag{9}$$

The coefficients relating to the mediation of this effect by personality are estimated with

$$\mathrm{Grade} = \beta_6^0 + F \times \mathrm{Cns} + G \times \mathrm{Nrt} + I \times \mathrm{Gender} + \epsilon_6. \tag{10}$$

The coefficients of the mediation of self-efficacy on the relation between gender, personality, and course grade are estimated by

$$\begin{aligned}\mathrm{Grade} = \beta_7^0 &+ F' \times \mathrm{Cns} + G' \times \mathrm{Nrt} + H \times \mathrm{SEF} \\ &+ I' \times \mathrm{Gender} + \epsilon_7.\end{aligned} \tag{11}$$

B. Moderation

The models of the previous section assume that all relations are linear in the independent variables, that no products of independent variables are important. To partially test this assumption, product terms involving gender were added to Eqs. (8) and (11). These product terms, for example Gender × Cns, tested whether the relations outlined in Fig. 2 were moderated by gender.

Gender was added as a moderator to Eq. (8) to determine if the effect of personality on self-efficacy was different for men and women. To do this, Eq. (8) was modified to

$$\begin{aligned}\mathrm{SEF} = \beta_8^0 &+ C' \times \mathrm{Gender} + D \times \mathrm{Cns} \\ &+ M_D \times \mathrm{Gender} \times \mathrm{Cns} \\ &+ E \times \mathrm{Nrt} + M_E \times \mathrm{Gender} \times \mathrm{Nrt} + \epsilon_8,\end{aligned} \tag{12}$$

where M_D and M_E are the regression coefficients of the interaction terms (moderators). If either regression coefficient is significant, the relation of either conscientiousness or neuroticism to self-efficacy is different for men and women. For example, if M_D is significant, the slope of the relation between conscientiousness and self-efficacy is D for women and $D + M_D$ for men. No statistically significant moderation was found in any of the courses (the regression coefficients M_D and M_E were not significant in any course). This is especially important and shows that, while women report higher mean levels of conscientiousness and neuroticism than men, the relation of both to self-efficacy is the same for men and women.

The moderation of the relation of conscientiousness, neuroticism, and self-efficacy to grade was tested with

$$\begin{aligned}\mathrm{Grade} = \beta_9^0 &+ I' \times \mathrm{Gender} + F' \times \mathrm{Cns} \\ &+ M_F \times \mathrm{Gender} \times \mathrm{Cns} \\ &+ G \times \mathrm{Nrt} + M_G \times \mathrm{Gender} \times \mathrm{Nrt} \\ &+ H \times \mathrm{SEF} + M_H \times \mathrm{Gender} \times \mathrm{SEF} + \epsilon_9,\end{aligned} \tag{13}$$

where M_F, M_G, and M_H are the regression coefficients of the interaction terms. No statistically significant moderation was found in any of the courses.

As such, the relation of gender, personality, self-efficacy, and course grade is not moderated by gender. The path models in Fig. 2 do not leave out important higher order terms involving gender. The relations of all these quantities are the same for men and women.

C. Mediation of the relation of gender to self-efficacy

Sections III C–III E present three separate mediation analyses. The overall results of these analyses for model 1 and 2 are presented in the path model in Fig. 3 for Physics 1. The path models presenting model 1 for the other classes are shown in Fig. 4; the path models for model 2 are presented in Fig. 5. Physics 1 is presented in the same figure to allow comparison between the two models. Results will be discussed for all classes; when a specific example would be helpful, Physics 1 is used.

Differences in personality may affect how a student interacts with academic situations. Higher conscientiousness may lead a student to complete more of his or her assignments or to work harder on those assignments, thus increasing academic achievement. Higher neuroticism may add to the anxiety felt during testing, lowering academic achievement, or may cause a student to feel excess concern about the class causing an increase in effort increasing achievement. Either lower or higher past academic achievement should influence future self-efficacy in Bandura's model. As such, it is possible personality mediates the relation of gender and self-efficacy.

The mediation analysis of the relation of gender, personality, and self-efficacy for all physics and calculus courses is shown in Table II in the Appendix. Beyond the coefficients presented in Figs. 3–5, this table also contains the calculation of the indirect effects through each path and the percentage of the total effect that is explained by the mediator.

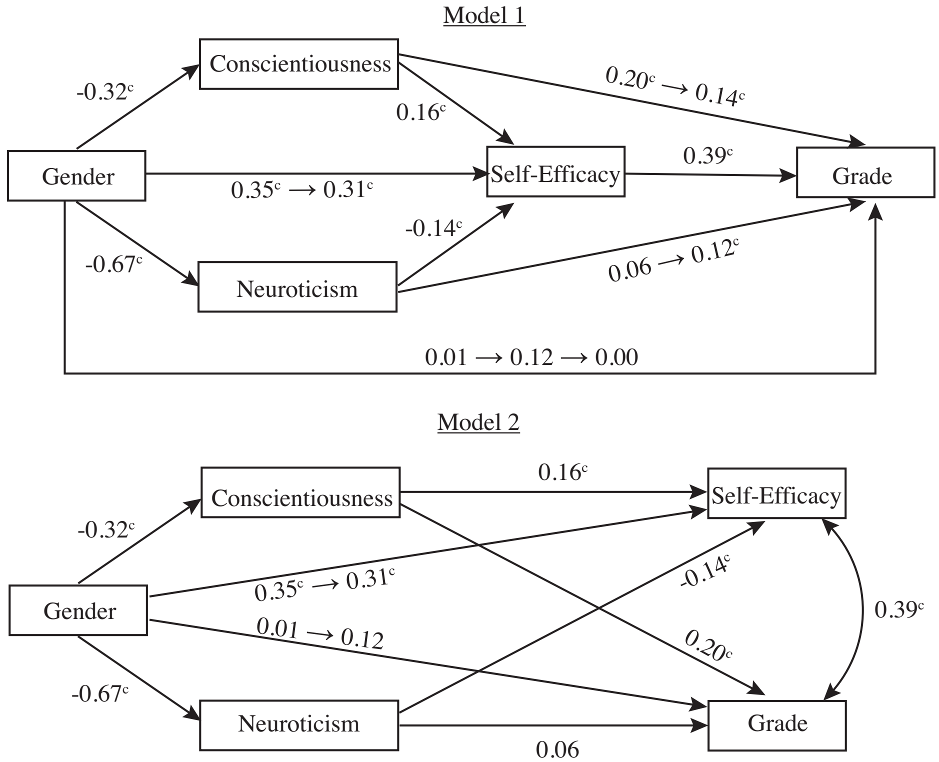

FIG. 3. Path models showing the relation of gender, personality, and self-efficacy for students in Physics 1. Gender was coded with women as zero, men as one. The number on each path is the value of the regression coefficient. The notation #1 → #2 shows the change in the coefficient before (#1) and after (#2) the addition of the mediating variables. Compare the figure with Fig. 2 for the symbolic variable related to each number. Note: “a” denotes $p < 0.05$, “b” $p < 0.01$, and “c” $p < 0.001$. A Bonferroni correction was applied to the significance levels.

The beta coefficient for the linear relation between a dichotomous and a normalized continuous variable represents the difference in the average of the continuous variable in standard deviation units between the two levels of the dichotomous variable and is related to Cohen’s d. The linear relation between two normalized continuous variables is related to the correlation between the two variables.

The relation of gender to self-efficacy was significant for all classes except Calculus 1B. For all four of these classes, the total effect of gender on self-efficacy (C) was reduced; however, in many classes the amount of reduction was fairly small. Examination of the two indirect paths from gender to self-efficacy shows that this was the result of paths through neuroticism and conscientiousness partially cancelling with men having lower conscientiousness which led to lower self-efficacy (14.4% of the total effect for Physics 1) while the lower level of neuroticism in men led to higher self-efficacy (26.7% of the total effect in Physics 1). For all classes, the indirect path through neuroticism accounted for a higher percentage of the total effect of gender on self-efficacy than the path through conscientiousness. For three of the classes, the path through neuroticism accounted for about 25% of the total effect; for Calculus 1 and Physics 2, 40% of the total effect. As such, a substantial part of the gender difference in self-efficacy was explained by differences in neuroticism. The path through conscientiousness explained about 15% of the total effect.

The direct effect of gender on personality shows the same pattern as was observed in Table I. The gender difference in conscientiousness was generally significant, but was less than a small effect in all classes except Physics 1. The direct effect of gender on neuroticism was much larger, in the range of a medium effect. The pattern of direct effects was different for Calculus 1A and 1B and the other classes. Note, for dichotomous variables regression coefficient β is related to Cohen’s d but they are not identical; d normalizes the difference in level by the pooled standard deviation while β uses the aggregated standard deviation.

D. Mediation of the relation of gender to achievement

For this section, we consider the overall effect of gender on course grade (I^T) which is estimated by Eq. (9) and whether the personality facets conscientiousness and neuroticism mediate this relationship. This mediation model is formed of the edges A, B, F, G, and $I^T \rightarrow I$ in Fig. 1 model 2; these edges also appear in model 1 with F, G, and I the values of the regression coefficients before taking into

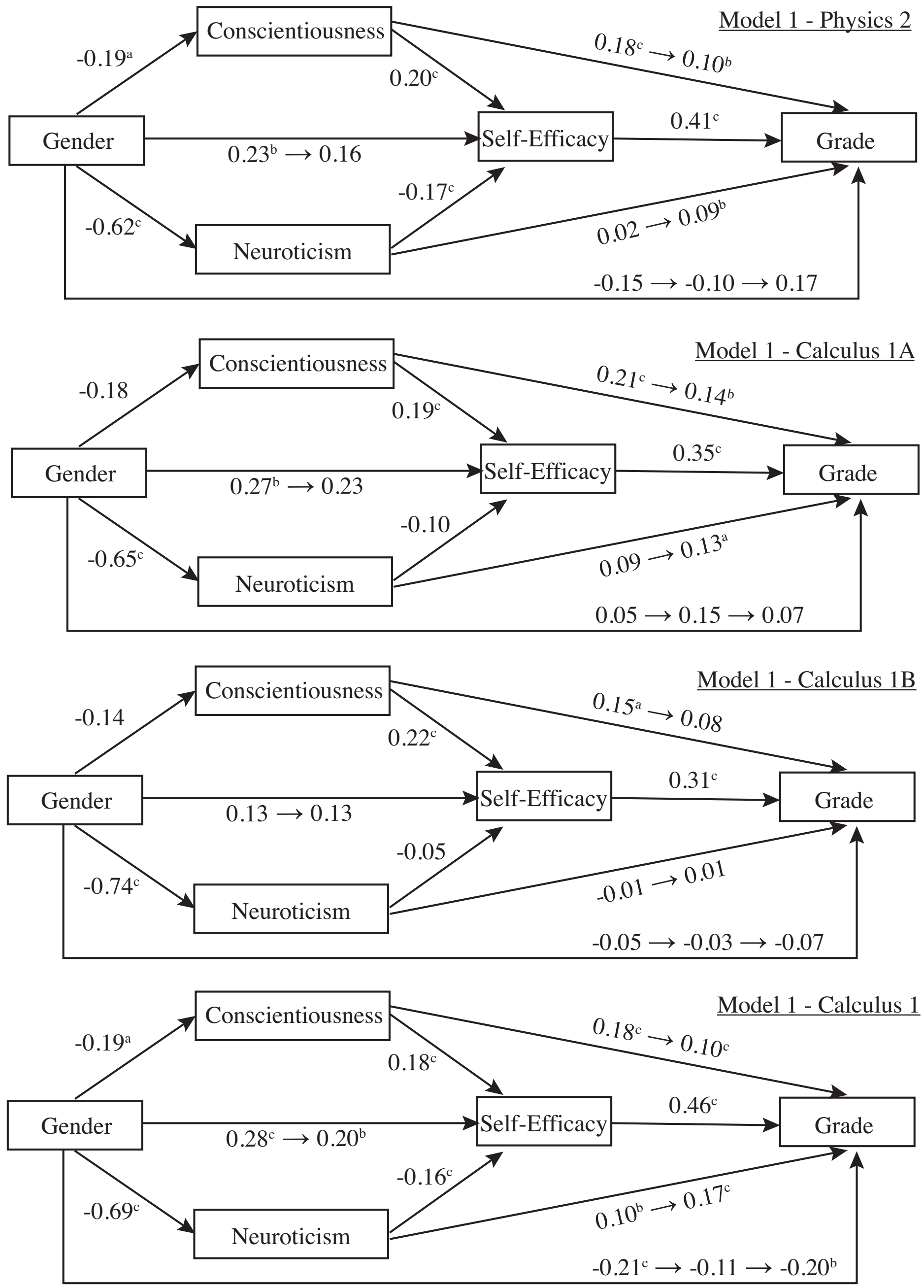

FIG. 4. Path models showing the relation of gender, personality, and self-efficacy for students in Physics 2, Calculus 1A, Calculus 1B, and Calculus 1. The number on each path is the value of the regression coefficient. The notation #1 → #2 shows the change in the coefficient before (#1) and after (#2) the addition of the mediating variables. Compare the figure with Fig. 2 for the symbolic variable related to each number. Note: "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$.

account the mediation of self-efficacy. The coefficients F, G, and I are estimated by Eq. (10).

Table III in the Appendix summarizes the mediation analysis. Figures 3–5 present the path models for the analysis. The total effect of gender on grade, I^T, was significant only in Calculus 1 at the level of a small effect. The effect of the personality variables either reduced the female advantage in grades or increased the male advantage. Examination of the indirect effects showed that the reason for this change is that women gain a small advantage in course grades through both the indirect path through conscientiousness and neuroticism. The advantage through the path through conscientiousness was expected and is supported by many general education studies [37]. The advantage gained through the path through neuroticism was less expected, but still understandable. Any disadvantage accrued through higher neuroticism causing increased anxiety in testing situations must be offset by positive

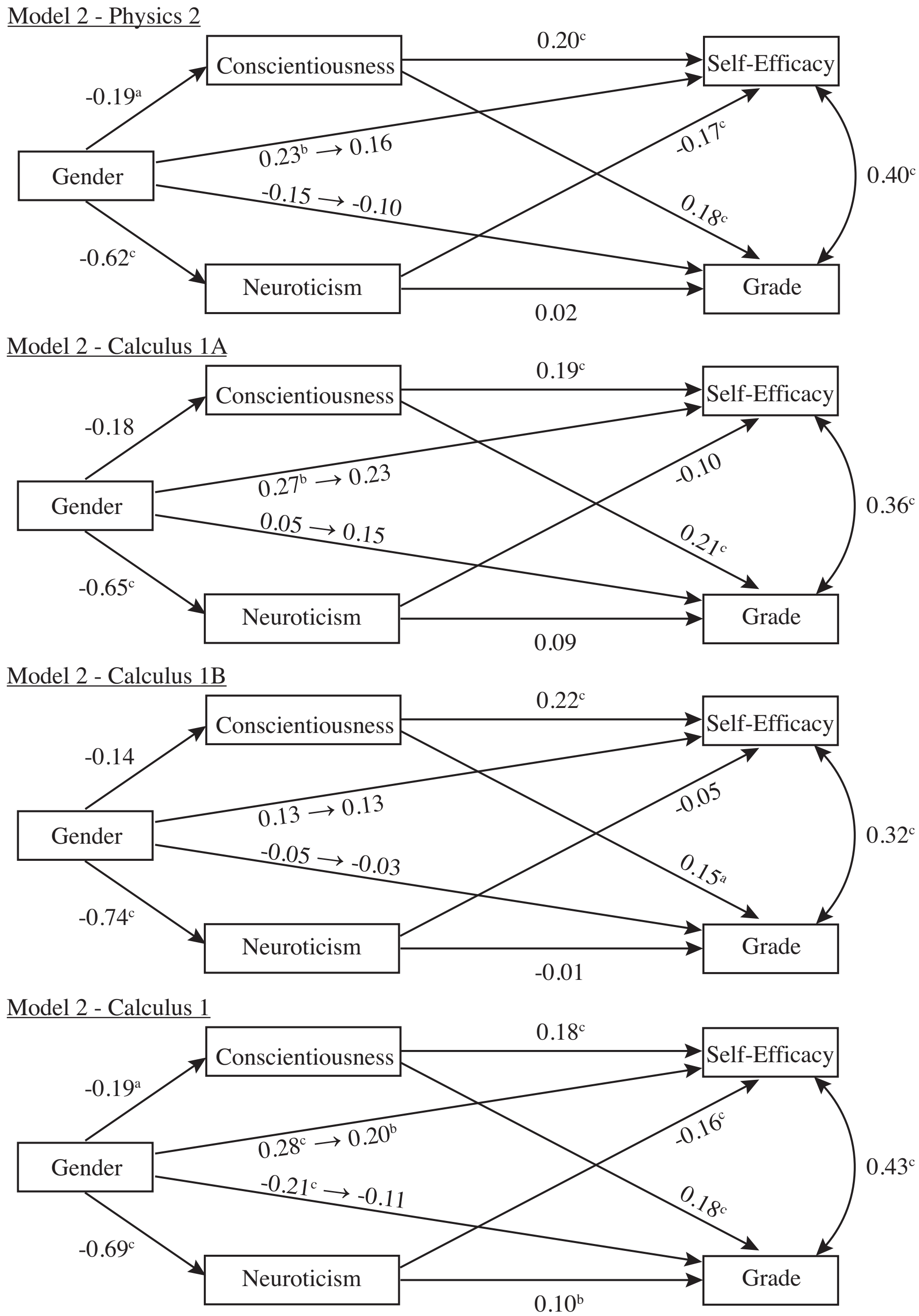

FIG. 5. Model 2 path models showing the relation of gender, personality, and self-efficacy for students in Physics 2, Calculus 1A, Calculus 1B, and Calculus 1. The number on each path is the value of the regression coefficient. The notation #1 → #2 shows the change in the coefficient before (#1) and after (#2) the addition of the mediating variables. Compare the figure with Fig. 2 for the symbolic variable related to each number. Note: "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$.

impacts of feeling anxiety or other strong emotions. For example, additional anxiety prior to a test may cause a student to prepare more thoroughly for the test.

E. Mediation of the relation of personality and self-efficacy to achievement

Substantial research has demonstrated a relation between personality and academic achievement [37]. The previous sections demonstrated that personality, particularly the neuroticism facet, mediated gender differences in self-efficacy and that the personality facets modified the relation of gender to grade. Self-efficacy has been reliably demonstrated as one of the most important noncognitive factors in explaining academic achievement [37]. It is, therefore, possible that the reported relation between personality and academic achievement actually exists because personality affected self-efficacy which affected achievement.

The unmediated model for this analysis removes the self-efficacy node from the path model in Fig. 2 model 1 and was investigated in the previous section. It contains the edges A estimated by Eq. (6), B estimated by Eq. (7), and F, G, and I estimated by Eq. (10).

The full model in Fig. 2 model 1 forms the mediated model containing self-efficacy. The addition of self-efficacy potentially modified the effect of conscientiousness on grade changing F to F', neuroticism on grade changing G to G' and the effect of gender on grade changing I to I'. These coefficients as well as the direct effect of self-efficacy on grade (H) are estimated by Eq. (11).

Each of the total effects can be partitioned into a remaining direct effect and an effect through the mediator (SEF): $F = F' + D \times H$, $G = G' + E \times H$, and $I = I' + C' \times H$. The fraction of the total effect that acts through the mediator is then calculated. The results for all classes are shown in Table IV in the Appendix. The path models for the analysis are shown in Figs. 3 and 4.

The total effect of gender on grade controlling for personality (I) was small and not significant in all classes after a Bonferroni correction was applied. Self-efficacy modified this effect (I') exposing a significant advantage toward women in course grade in Calculus 1 controlling for personality and self-efficacy. In all classes, the mediated effect of gender on grade was more advantageous to women than the unmediated effect.

The total effect of conscientiousness on grade controlling for gender and neuroticism (F) was significant and at or near the level of a small effect in all classes. This was consistent with a substantial body of research showing the importance of this facet in explaining academic performance [37]. This effect was strongly mediated by self-efficacy (F') with the path through self-efficacy accounting for 30%–45% of the total effect of conscientiousness on grade. As such, a substantial portion of this facet's effect on academic performance can be explained by its effect on self-efficacy. This is consistent with Bandura's model where the prior academic achievement experienced by conscientious students leads to higher levels of self-efficacy.

The total effect of neuroticism on grade (G) controlling for conscientiousness and gender was small in all classes and significant at the $p < 0.01$ level in only one class. The addition of self-efficacy exposed a significant positive effect of neuroticism on grade in four of the five classes; the effect was at or near the level of a small effect.

IV. DISCUSSION

This study was conducted to investigate five research questions. Each question will be discussed in order in this section.

RQ1: Does self-efficacy or personality differ for men and women in core university introductory mathematics and physics classes? The self-efficacies of men and women were significantly different in Calculus 1A, Calculus 1, Physics 1, and Physics 2. The results for these courses were consistent with Huang's meta-analysis which showed the self-efficacy of men was higher than women in STEM classes [24]. The effect size for the difference in self-efficacy for Physics 1 was reduced in Physics 2, similarly the effect size of the difference for Calculus 1A was reduced in Calculus 1B and was no longer significant.

Multiple personality facets were different for men and women in some classes. Agreeableness was significantly different in Workshop Mathematics and Physics 1 with women reporting higher levels of agreeableness in each of these courses; the differences represented a small effect. Women reported significantly higher conscientiousness in Calculus 1 and Physics 1, a small effect. Neither openness nor extraversion were significantly different for men and women in any class.

Women reported significantly higher neuroticism in all classes, a medium to large effect. These values differed by 0.4–0.6 points on a 5-point Likert scale. The observed differences in neuroticism were similar to those reported in a large ($N > 10^6$) nonacademic study [34]. The consistency of the differences in this facet for all classes and compared to a national sample suggest the difference in neuroticism is a common feature of college age students, not a feature specific to STEM students.

RQ2: Does gender moderate the relationships of personality, self-efficacy, and achievement? No significant moderation was detected in any model. The relationship of conscientiousness and neuroticism to self-efficacy is the same for men and women, as is the relationship of conscientiousness, neuroticism, and self-efficacy to achievement. As shown in Table I, both men and women report a range of values for both the neuroticism and conscientiousness facets. While women report, on average, higher levels of both facets, physics and mathematics classes contain both men and women with high or low levels of each of these facets. As such, modifications to the classes to reduce anxiety or to promote more conscientious behaviors will help a broad group of students.

The failure to find significant moderation, particularly of the neuroticism facet has another important implication. Women report on average higher levels of neuroticism than men. It could have been that experiencing higher levels of anxiety was in general less important for the development of self-efficacy for women than for men. If the moderation coefficient, M_E in Eq. (12), was significant and positive, then the slope of the relation of neuroticism to self-efficacy would be smaller for women than for men. If this was the case, the higher levels of neuroticism reported by women would be partially compensated for by a weaker effect of neuroticism on self-efficacy. Unfortunately, this was not the case. The higher levels of neuroticism for women affect self-efficacy just as much as men with higher levels of neuroticism.

These observations suggest that physics and mathematics classes consider the levels of anxiety various class structures generate and act to minimize the anxiety. This would help all students predisposed to feel higher level of anxiety, but would more strongly help women, and might serve to lower difference in self-efficacy between men and women.

RQ3: Does personality mediate the relationship of gender to self-efficacy? If so, how does it mediate the relationship? The mediation of self-efficacy by the personality facets conscientiousness and neuroticism was investigated in Table II. For Calculus 1A, Calculus 1, Physics 1, and Physics 2 the total effect (C) of gender on self-efficacy was significant (a small effect) and fairly similar with β ranging from 0.23 to 0.35. With neuroticism and conscientiousness added as mediating variables, these total effects were reduced somewhat by 0.04–0.07 to produce remaining direct effects for 0.16–0.31. Examination of the path model showed this weak mediation partially resulted from the competition of the two facets. The lower conscientiousness of men led to lower self-efficacy accounting for an average of 16% of the total effect (C). The lower neuroticism of men led to higher self-efficacy accounting for an average of 33% of the total effect.

RQ4: Does personality mediate the relationship of gender to achievement? If so, how does it mediate the relationship? This model is summarized in Table III. The relation of gender to achievement was significant in only one class, so in general the relation fails Baron and Kenny's test of mediation. The direct effect of gender on conscientiousness was significant in three of the five classes at or near the level of a small effect. We note the coefficients for Calculus 1A and 1B are also near a small effect and the failure to find a significant effect in these classes is likely the result of the smaller sample size. The coefficient was significant in Calculus 1A before Bonferroni correction. The direct effect of gender on neuroticism was significant and large in all classes at the level of medium to a large effect. The direct effect of conscientiousness on grade was significant and positive in all classes at the level of a small effect. The direct effect of neuroticism on grade was generally positive but significant in only one class. The indirect effects through both conscientiousness and neuroticism were significant and negative in all classes except Calculus 1B. The sum of these effects were −0.10 in three of the classes; half the size of a small effect. This indicates in general that women have an advantage in achievement both due to higher levels of conscientiousness and neuroticism, reflecting a small effect in the first class in each sequence (Calculus 1A, Calculus 1, and Physics 1).

RQ5: Does self-efficacy mediate the relationship of personality and gender to achievement? If so, how does it mediate the relationship? Self-efficacy could potentially mediate the relation of three variables to achievement in this model; the total effect of gender on grade (I), the total effect of conscientiousness on grade (F), and the total effect of neuroticism on grade (G). The effect of gender on grade was small and nonsignificant in all classes; the addition of self-efficacy exposed a significant (small effect size) advantage to women, but only in Calculus 1. Conscientiousness had a significant positive total effect on grade (F) in all classes, a small effect. The β coefficients were strongly reduced in all classes, and became insignificant in Calculus 1B; self-efficacy strongly mediated the relation of conscientiousness to grade explaining on average 39% of the total effect. As such, a substantial part of the effect of conscientiousness on grade can be explained by its prior effect on self-efficacy. The total effect of neuroticism on grade (G) was small in all classes; it was significant at the level of a small effect only in Calculus 1. With the addition of self-efficacy as a mediator, a significant remaining positive direct effect (G') was uncovered in four or the five classes at or near the level of a small effect. As such, higher neuroticism actually improves grades once the negative effect of self-efficacy is accounted for. Conscientiousness and neuroticism produced competing indirect effects on grade by their action on self-efficacy.

V. IMPLICATIONS AND RECOMMENDATIONS

This work combined multiple well-established research strands: the relation of anxiety to test performance, the relation of achievement to self-efficacy, general differences between how men and women report the tendency to experience anxiety, general differences in conscientiousness between men and women, and the relation of conscientiousness to academic success. Together, these strands suggest that a substantial amount of the often-reported differences in self-efficacy between men and women may result from competing gender differences in the tendency to experience anxiety and the tendency to conscientiously complete tasks. Self-efficacy has long been an important construct in models explaining career choice and persistence and is a significant contributor to academic success. As such, variation in physics and mathematics self-efficacy by gender may be one source of differences in the representation of men and women in STEM fields requiring these classes.

This work advanced a more nuanced definition of self-efficacy for students in college science and mathematics classes. For students just starting their journey in the sciences, their belief they can succeed may be separate from experiences informing that belief. Interventions that change self-efficacy can act on those beliefs without the confounding variable of prior success. Students in the mathematics and physics classes studied have long experience with mathematics and science classes and generally a history of success in those classes. Interventions to modify self-efficacy may change beliefs, but cannot modify prior experiences upon which those beliefs are grounded. As such, care should be taken as interpreting the relation between self-efficacy and achievement as causal as implied in Fig. 2 model 1 as opposed to correlational as shown in model 2. An intervention increasing self-efficacy will likely not increase grade to the extent implied in model 1 because a substantial part of the relation of self-efficacy to grade must result from the relation of prior achievement to grade which informs self-efficacy but which also affects grades.

The work presented suggests that some modification of self-efficacy is needed. Model 2 in Fig. 2 shows that while the higher conscientiousness of women is related to higher grades as well as higher self-efficacy as Bandura's model suggests should be the case, the higher neuroticism of women, which was also related to higher grades, was related to lower self-efficacy.

Having identified differences in how personality differences affect achievement and self-efficacy, and identified substantial prior academic experiences as an important component of the academic self-efficacy of college STEM students, one can re-examine interventions designed to improve self-efficacy.

Positive effects on self-efficacy have been shown among varied learners engaged in a range of intervention activities including the development of targeted learning goals, self-evaluation, and the incorporation of formative and progress-based feedback. Much of this research is performed with noncollege age students or in non STEM environments. Generally, intervention work in this area suggests that students self-efficacy can be promoted by providing them with mastery experiences (experiences where the student successfully complete a relevant task), exposing them to successful models of performance, and providing goal-oriented, formative, and positive feedback [42]. Instructionally, the majority of previous research has relied on the use of effective social models (peer and role models) in demonstrating and explaining tasks [104–106] as well as the use of models to support efficacy and strategy instruction. Such intervention work is directly grounded in social cognitive theory [40] and has been found to be effective when students engage with models who are similar to them, who admit to making errors and approach their academic work with openness, and who demonstrate coping strategies in the face of academic challenges [27]. These considerations are imposing given that gender disparities in physics are due in part to there being few female role models in physics academic communities [2]. In addition, it has been shown that other sources of self-efficacy identified in Bandura's work such as vicarious learning (watching others perform a relevant task) and social persuasion (support from others that you can perform a relevant task) have positively predicted the success of women in introductory physics courses [50]. Therefore, it may be productive to shift our attention from a mastery experience intervention within courses to other interventions that would impact these specific sources of self-efficacy. It could be possible that these could reduce levels of anxiety and, in turn, reduce the gender gap in self-efficacy.

Similar research in engineering has found that the use of classroom interventions focused on peer modeling and the role of engineering in effecting positive change in society resulted in significant increases in female students' positive perceptions of engineering and engineering self-efficacy [107]. Such course-based intervention efforts align with previous research in physics showing that an instructional emphasis on applying physics to the solving of real-world problems resulted in improved conceptual understanding for women and positively predicted physics identity for both men and women [14,108].

In the context of physics, research has also found that women demonstrated improved physics self-concept and efficacy beliefs when they adopted a malleable view of intelligence and ability and when they engaged in communal learning environments that emphasized small group interaction, discussion, and hands-on activities [2,10]. Further, research by Lock and Hazari [1] found that the development of female students' physics identity was promoted by explicit, classroom-based discussions of the underrepresentation of women in science. Taken together, the results of this research have implications for the development of course-based intervention supports aimed to promote underrepresented students' positive and productive attitudes and beliefs about learning physics [57].

Taken as a whole, the substantial work on interventions provides a means to mitigate self-efficacy differences between men and women after they develop. The combination of the positive effects of neuroticism on grades and the negative effect on self-efficacy suggest that men and women are processing the results of prior academic achievement differently partially because they feel differing levels of anxiety during the experience and toward future experiences. The identification of anxiety as a factor in the development of self-efficacy differences may allow a different strand of research that seeks modifications to existing educational structures to prevent self-efficacy differences by gender from emerging. If a tendency to experience anxiety is a substantial factor contributing to differences in self-efficacy, then, to promote inclusion, instructional structures should be examined to determine how they can be modified to reduce anxiety. This may involve modifying how assignments are constructed, how they are evaluated, how examinations are constructed, and how examinations are given. Some pedagogical practices such as randomly calling on students to elicit opinions on problems may need to be reevaluated. Researchers have previously examined postsecondary physics and chemistry curricula [109,110] and we guide the reader to those articles for specific recommendations. In addition, related fields such as chemistry and biology where fewer gendered differences in self-efficacy are reported and differences in representation are lower may offer models.

VI. LIMITATIONS

This work was performed at a single research university in calculus and calculus-based physics classes. Additional research at other institutions including primarily teaching focused institutions with different student populations is needed to understand if the results obtained in this research can be generalized to represent physics students nationally. Additional research should also investigate algebra-based physics classes. Further, this study used a single observation of self-efficacy collected midsemester. Multiple measurements taken at different times during the semester and longitudinally in different classes would allow a more thorough characterization of the recursive development of self-efficacy predicted by Bandura's model. The study also captures generally self-efficacy toward the class; this self-efficacy could be differentiated between differing tasks within the class.

VII. CONCLUSIONS AND FUTURE RESEARCH

This study identified differences in personality as a potential origin for the differences in the self-efficacy beliefs of men and women in physics and mathematics courses. Personality may also explain differences often reported for men and women in engineering classes; most of the students in the current study were engineering majors. Similar differences in self-efficacy are not reported in chemistry and biology while the students in these classes almost certainly also have the same differences in conscientiousness and neuroticism reported in this study. Future research should be conducted to understand the features of course environments that both promote and constrain the development of students' physics and mathematics self-efficacy beliefs. Beyond these possible directions, a qualitative study could shed further light on the self-efficacy difference of men and women in physics classes.

This work examined the conscientiousness and neuroticism facets of the five-factor model of personality and self-efficacy towards physics and mathematics for students in introductory physics and mathematics classes. Women reported substantially higher neuroticism in all courses studied, near a large effect. This was consistent with the results of a large national study suggesting the result is general. Women also reported higher conscientiousness and lower self-efficacy in many of the classes studied, small effects. Neuroticism mediated the relation of gender to self-efficacy substantially in most classes; the path through the mediator explained from 25% to 47% of the total effect. Conscientiousness mediated the effect of gender on self-efficacy more weakly explaining for 12%–23% of the total effect.

The relation of personality to self-efficacy and self-efficacy to course grade was generally consistent for men and women; significant moderation was not measured in any class. As such, the negative relation of neuroticism to self-efficacy is the same for men and women.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grants No. ECR-1561517 and No. HRD-1834569.

APPENDIX: FULL REGRESSION AND MEDIATION ANALYSIS

This appendix presents the full mediation tables for all classes. Beyond the values represented in the path models, these tables also present the percentage of each effect that acts through each mediating pathway.

TABLE II. The mediation by neuroticism and conscientiousness of the relation of gender to self-efficacy. The regression coefficient β and its standard error (SE) are presented. Women are coded as zero, men as one. For indirect effects, the product of the path coefficients $\beta_i\beta_j$ is presented and the standard deviation (SD) of the product. Conscientiousness is abbreviated Cns, neuroticism Nrt, and self-efficacy SEF. Note that "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$. A Bonferroni correction was applied to the significance levels.

	Calculus 1A		Calculus 1B		Calculus 1		Physics 1		Physics 2	
	β	SE	β	SE	β	SE	β	SE	β	SE
Total effect and remaining effect										
Gender → SEF (C)	0.27^b	0.07	0.13	0.08	0.28^c	0.05	0.35^c	0.05	0.23^b	0.06
$C = C' + A \times D + B \times E$										
Gender → SE (C')	0.23	0.08	0.13	0.09	0.20^b	0.05	0.31^c	0.05	0.16	0.06
Direct effects										
Gender → Cns (A)	−0.18	0.07	−0.14	0.08	-0.19^a	0.05	-0.32^c	0.05	-0.19^a	0.06
Gender → Nrt (B)	-0.65^c	0.05	-0.74^c	0.08	-0.69^c	0.05	-0.67^c	0.05	-0.62^c	0.06
Cns → SEF (D)	0.19^c	0.05	0.22^c	0.04	0.18^c	0.03	0.16^c	0.02	0.20^c	0.03
Nrt → SEF (E)	−0.10	0.04	−0.05	0.05	-0.16^c	0.03	-0.14^c	0.03	-0.17^c	0.03
Indirect effects										
	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD
Gender → Cns → SE ($A \times D$)	-0.03^c	0.01	-0.03^c	0.02	-0.03^c	0.01	-0.05^c	0.01	-0.05^c	0.01
% of Total effect ($A \times D/C$)	−12.9%		−22.9%		−12.0%		−14.4%		−16.4%	
Gender → Nrt → SE ($B \times E$)	0.07^c	0.03	0.04^c	0.03	0.11^c	0.02	0.09^c	0.01	0.11^c	0.01
% of total effect ($B \times E/C$)	24.9%		26.6%		39.0%		26.7%		46.5%	

TABLE III. The mediation by neuroticism and conscientiousness of the relation of gender to grade. The regression coefficient β and its standard error are presented. Women are coded as zero, men as one. For indirect effects, the product of the path coefficients $\beta_i\beta_j$ is presented and the standard deviation of the product. Conscientiousness is abbreviated Cns, neuroticism Nrt, and self-efficacy SEF. Note that "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$. A Bonferroni correction was applied to the significance levels.

	Calculus 1A		Calculus 1B		Calculus 1		Physics 1		Physics 2	
	β	SE	β	SE	β	SE	β	SE	β	SE
Total effect and remaining effect										
Gender → Grade (I^T)	0.05	0.08	−0.05	0.08	-0.21^c	0.05	0.01	0.05	−0.15	0.06
$I^T = A \times F + B \times G + I$										
Gender → Grade (I)	0.15	0.08	−0.03	0.09	−0.11	0.06	0.12	0.06	−0.10	0.06
Direct effects										
Gender → Cns (A)	−0.18	0.07	−0.14	0.08	-0.19^a	0.05	-0.32^c	0.05	-0.19^a	0.06
Gender → Nrt (B)	-0.65^c	0.05	-0.74^c	0.08	-0.69^c	0.05	-0.67^c	0.05	-0.62^c	0.06
Nrt → Grade (G)	0.09	0.04	−0.01	0.05	0.10^b	0.05	0.06	0.03	0.02	0.03
Cns → Grade (F)	0.21^c	0.04	0.15^a	0.04	0.18^c	0.03	0.20^c	0.02	0.18^c	0.03
Indirect effects										
	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD
Gender → Cns → Grade ($A \times F$)	-0.04^c	0.02	-0.02^c	0.02	-0.04^c	0.01	-0.06^c	0.01	-0.03^c	0.01
Gender → Nrt → Grade ($B \times G$)	-0.06^c	0.02	0.01	0.04	-0.06^c	0.02	-0.04^c	0.02	-0.01^c	0.02

TABLE IV. The mediation by self-efficacy of the relation of gender, neuroticism, and conscientiousness to grade. The regression coefficient β and its standard error (SE) are presented. Women are coded as zero, men as one. For indirect effects, the product of the path coefficients $\beta_i\beta_j$ is presented and the standard deviation (SD) of the product. Conscientiousness is abbreviated Cns, neuroticism Nrt, and self-efficacy SEF. Note that "a" denotes $p < 0.05$, "b" $p < 0.01$, and "c" $p < 0.001$. A Bonferroni correction was applied to the significance levels.

	Calculus 1A		Calculus 1B		Calculus 1		Physics 1		Physics 2	
	β	SE	β	SE	β	SE	β	SE	β	SE
Total and remaining effects										
Gender → Grade (I)	0.15	0.08	−0.03	0.09	−0.11	0.06	0.12	0.06	−0.10	0.06
$I = C' \times H + I'$										
Gender → Grade (I')	0.07	0.07	−0.07	0.09	-0.20^b	0.05	0.00	0.05	−0.17	0.06
Cns → Grade (F)	0.21^c	0.04	0.15^a	0.04	0.18^c	0.03	0.20^c	0.02	0.18^c	0.03
$F = D \times H + F'$										
Cns → Grade (F')	0.14^b	0.04	0.08	0.04	0.10^c	0.02	0.14^c	0.02	0.10^b	0.02
Nrt → Grade (G)	0.09	0.04	−0.01	0.05	0.10^b	0.05	0.06	0.03	0.02	0.03
$G = E \times H + G'$										
Nrt → Grade (G')	0.13^a	0.04	0.01	0.04	0.17^c	0.02	0.12^c	0.02	0.09^b	0.03
Direct effects										
SEF → Grade (H)	0.35^c	0.03	0.31^c	0.04	0.46^c	0.02	0.39^c	0.02	0.41^c	0.02
Indirect effects										
	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD	$\beta_i\beta_j$	SD
Gender → SEF → Grade ($C' \times H$)	0.09^c	0.02	0.03^c	0.02	0.10^c	0.03	0.12^c	0.02	0.12^c	0.02
% of total effect ($C' \times H/I$)	…		…		…		…		…	
Cns → SEF → Grade ($D \times H$)	0.06^c	0.01	0.07^c	0.02	0.08^c	0.01	0.06^c	0.01	0.06^c	0.01
% of total effect ($D \times H/F$)	32.1%		45.8%		44.7%		30.5%		44.1%	
Nrt → SEF → Grade ($E \times H$)	-0.03^c	0.01	-0.01^c	0.01	-0.07^c	0.01	-0.06^c	0.01	-0.05^c	0.01
% of total effect ($E \times H/G$)	…		…		-71.6%		…		…	
Correlations										
	r		r		r		r		r	
Correlation SEF and grade	0.36^c		0.32^c		0.43^c		0.39^c		0.40^c	

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

[2] A. M. Kelly, Social cognitive perspective of gender disparities in undergraduate physics, Phys. Rev. Phys. Educ. Res. **12**, 020116 (2016).

[3] Percent of physics bachelors earned by women, classes of 1976 to 2020 (American Institute of Physics, College Park, MD, 2020), https://www.aip.org/statistics/data-graphics/percent-physics-bachelor%E2%80%99s-earned-women-classes-1976-2020.

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

[5] R. Scherr, Never mind the gap: Gender-related research in Physical Review Physics Education Research, 2005–2016, Phys. Rev. Phys. Educ. Res. **12**, 020003 (2016).

[6] S. Salehi, E. Burkholder, G. P. Lepage, S. Pollock, and C. Wieman, Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics, Phys. Rev. Phys. Educ. Res. **15**, 020114 (2019).

[7] R. Henderson and J. Stewart, Racial and ethnic bias in the Force Concept Inventory, presented at PER Conf. 2017, Cincinnati, OH, 10.1119/perc.2017.pr.038.

[8] R. Henderson, C. Zabriskie, and J. Stewart, Rural and first generation performance differences on the Force and Motion Conceptual Evaluation, presented at PER Conf. 2018, Washington, DC, 10.1119/perc.2018.pr.Henderson.

[9] D. S. Hewagallage, J. Stewart, and R. Henderson, Differences in the predictive power of pretest scores of students underrepresented in physics, presented at PER Conf. 2019, Provo, UT, 10.1119/perc.2019.pr.Hewagallage.

[10] 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).

[11] Z. Y. Kalender, E. Marshman, Christian 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).

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

[13] 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).

[14] 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).

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

[16] L. M. Larson, K. M. Pesch, S. Surapaneni, V. S. Bonitz, T. F. Wu, and J. D. Werbel, Predicting graduation: The role of mathematics/science self-efficacy, J. Career Assess. **23**, 399 (2015).

[17] S. Lau and R. W. Roeser, Cognitive abilities and motivational processes in high school students' situational engagement and achievement in science, Educ. Assess. **8**, 139 (2002).

[18] R. R. Bryan, S. M. Glynn, and J. M. Kittleson, Motivation, achievement, and advanced placement intent of high school students learning science, Sci. Educ. **95**, 1049 (2011).

[19] J. Dalgety and R. K. Coll, Exploring first-year science students' chemistry self-efficacy, Int. J. Sci. Math. Educ. **4**, 97 (2006).

[20] P. R. Aschbacher, E. Li, and E. J. Roth, Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine, J. Res. Sci. Teach. **47**, 564 (2010).

[21] D. A. Luzzo, P. Hasper, K. A. Albert, M. A. Bibby, and E. A. Martinelli Jr., Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students, J. Counsel. Psychol. **46**, 233 (1999).

[22] R. M. Marra, K. A. Rodgers, D. Shen, and B. Bogue, Women engineering students and self-efficacy: A multi-year, multi-institution study of women engineering student self-efficacy, J. Eng. Educ. **98**, 27 (2009).

[23] W. J. Hughes, Perceived gender interaction and course confidence among undergraduate science, mathematics, and technology majors, J. Women Minorities Sci. Eng. **6**, 1 (2000).

[24] C. Huang, Gender differences in academic self-efficacy: A meta-analysis, Eur. J. Psychol. Educ. **28**, 1 (2013).

[25] S. L. Eddy and S. E. Brownell, Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines, Phys. Rev. Phys. Educ. Res. **12**, 020106 (2016).

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

[27] F. Pajares, Gender and perceived self-efficacy in self-regulated learning, Theory Into Practice **41**, 116 (2002).

[28] X. Ma, A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics, J. Res. Math. Educ. **30**, 520 (1999).

[29] N. M. Else-Quest, J. S. Hyde, and M. C. Linn, Cross-national patterns of gender differences in mathematics: A meta-analysis, Psychol. Bull. **136**, 103 (2010).

[30] R. A. Alvaro, The effectiveness of a science therapy program upon science anxious undergraduates, Ph.D. thesis, Loyola University Chicago, 1978.

[31] J. V. Mallow, A science anxiety program, Am. J. Phys. **46**, 862 (1978).

[32] J. V. Mallow and S. L. Greenburg, Science anxiety: Causes and remedies, J. Coll. Sci. Teach. **11**, 356 (1982).

[33] J. Mallow, H. Kastrup, F. B. Bryant, N. Hislop, R. Shefner, and M. Udo, Science anxiety, science attitudes, and gender: Interviews from a binational study, J. Sci. Educ. Technol. **19**, 356 (2010).

[34] S. Srivastava, O. P. John, S. D. Gosling, and J. Potter, Development of personality in early and middle adulthood: Set like plaster or persistent change?, J. Pers. Soc. Psychol. **84**, 1041 (2003).

[35] N. Schmitt, The interaction of neuroticism and gender and its impact on self-efficacy and performance, Hum. Perform. **21**, 49 (2007).

[36] J. Stewart, G. L. Cochran, R. Henderson, C. Zabriskie, S. DeVore, P. Miller, G. Stewart, and L. Michaluk, Mediational effect of prior preparation on performance differences of students underrepresented in physics, Phys. Rev. Phys. Educ. Res. **17**, 010107 (2021).

[37] M. Richardson, C. Abraham, and R. Bond, Psychological correlates of university students' academic performance: A systematic review and meta-analysis, Psychol. Bull. **138**, 353 (2012).

[38] C. C. Cohen and N. Deterding, Widening the net: National estimates of gender disparities in engineering, J. Eng. Educ. **98**, 211 (2009).

[39] J. M. Braxton, W. R. Doyle, H. V. Hartley III, A. S. Hirschy, W. A. Jones, and M. K. McLendon, *Rethinking College Student Retention* (John Wiley & Sons, San Francisco, CA, 2013).

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

[41] A. Bandura, Self-efficacy, in *Encyclopedia of Human Behavior*, Vol. 4, edited by V. S. Ramachandran (Academic Press, San Diego, CA, 1994), p. 71.

[42] D. H. Schunk and P. Ertmer, Self-regulation and academic learning: Self-efficacy enhancing interventions, in *Handbook of Self-Regulation*, edited by M. Boekaerts, P. R. Pintrich, and M. Zeidner (Elsevier, San Diego, CA, 2000), p. 631.

[43] S. Andrew, Self-efficacy as a predictor of academic performance in science, J. Adv. Nurs. **27**, 596 (1998).

[44] J. Pietsch, R. Walker, and E. Chapman, The relationship among self-concept, self-efficacy, and performance in

mathematics during secondary school, J. Educ. Psychol. **95**, 589 (2003).

[45] N. E. Betz and G. Hackett, The relationship of career-related self-efficacy expectations to perceived career options in college women and men, J. Counsel. Psychol. **28**, 399 (1981).

[46] R. W. Lent, S. D. Brown, and K. C. Larkin, Self-efficacy in the prediction of academic performance and perceived career options, J. Counsel. Psychol. **33**, 265 (1986).

[47] R. W. Lent, S. D. Brown, and K. C. Larkin, Comparison of three theoretically derived variables in predicting career and academic behavior: Self-efficacy, interest congruence, and consequence thinking, J. Counsel. Psychol. **34**, 293 (1987).

[48] D. Voyer and S. D. Voyer, Gender differences in scholastic achievement: A meta-analysis, Psychol. Bull. **140**, 1174 (2014).

[49] M. E. Junge and B. J. Dretzke, Mathematical self-efficacy gender differences in gifted/talented adolescents, Gifted Child Quart. **39**, 22 (1995).

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

[51] V. Sawtelle, E. Brewe, R. M. Goertzen, and L. H. Kramer, Identifying events that impact self-efficacy in physics learning, Phys. Rev. Phys. Educ. Res. **8**, 020111 (2012).

[52] J. M. Nissen, Gender differences in self-efficacy states in high school physics, Phys. Rev. Phys. Educ. Res. **15**, 013102 (2019).

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

[54] 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. (formerly CAL-laborate International) **19**, 1 (2011), https://openjournals.library.usyd.edu.au/index.php/CAL/article/view/4770.

[55] L. E. Kost-Smith, Characterizing, modeling, and addressing gender disparities in introductory college physics, Ph.D. thesis, University of Colorado at Boulder, 2011.

[56] R. Dou, E. Brewe, J. P. Zwolak, G. Potvin, E. A. Williams, and L. H. Kramer, Beyond performance metrics: Examining a decrease in students physics self-efficacy through a social networks lens, Phys. Rev. Phys. Educ. Res. **12**, 020124 (2016).

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

[58] V. Sawtelle, E. Brewe, and L. H. Kramer, Positive impacts of modeling instruction on self-efficacy, AIP Conf. Proc. **1289**, 289 (2010).

[59] K. Miller, J. Schell, A. Ho, B. Lukoff, and E. Mazur, Response switching and self-efficacy in Peer Instruction classrooms, Phys. Rev. Phys. Educ. Res. **11**, 010104 (2015).

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

[61] J. M. Hall and M. K. Ponton, Mathematics self-efficacy of college freshman, J. Devel. Educ. **28**, 26 (2005), https://eric.ed.gov/?id=EJ718579.

[62] M. L. Peters, Examining the relationships among classroom climate, self-efficacy, and achievement in undergraduate mathematics: A multi-level analysis, Int. J. Sci. Math. Educ. **11**, 459 (2013).

[63] M. L. Bernacki, T. J. Nokes-Malach, and V. Aleven, Examining self-efficacy during learning: Variability and relations to behavior, performance, and learning, Metacogn. Learn. **10**, 99 (2015).

[64] H. Hartman and M. Hartman, Do gender differences in undergraduate engineering orientations persist when major is controlled?, Int. J. Gender Sci. Tech. **1**, 61 (2009), http://genderandset.open.ac.uk/index.php/genderandset/article/view/56.

[65] M. Micari, P. Pazos, and M. J. Z. Hartmann, A matter of confidence: Gender differences in attitudes toward engaging in lab and course work in undergraduate engineering, J. Women Minorities Sci. Eng. **13**, 279 (2007).

[66] E. Cech, B. Rubineau, S. Silbey, and C. Seron, Professional role confidence and gendered persistence in engineering, Am. Sociol. Rev. **76**, 641 (2011).

[67] M. Besterfield-Sacre, M. Moreno, L. J. Shuman, and C. J. Atman, Gender and ethnicity differences in freshmen engineering student attitudes: A cross-institutional study, J. Eng. Educ. **90**, 477 (2001).

[68] J. P. Concannon and L. H. Barrow, A reanalysis of engineering majors self-efficacy beliefs, J. Sci. Educ. Technol. **21**, 742 (2012).

[69] C. M. Jagacinski, Women engineering students: Competence perceptions and achievement goals in the freshman engineering course, Sex Roles **69**, 644 (2013).

[70] C. M. Vogt, D. Hocevar, and L. S. Hagedorn, A social cognitive construct validation: Determining women's and men's success in engineering programs, J. High. Educ. **78**, 337 (2007).

[71] R. W. Lent, S. D. Brown, H. Sheu, J. Schmidt, B. R. Brenner, C. S. Gloster, G. Wilkins, L. C. Schmidt, H. Lyons, and D. Treistman, Social cognitive predictors of academic interests and goals in engineering: Utility for women and students at historically Black universities, J. Counsel. Psychol. **52**, 84 (2005).

[72] M. A. Hutchison, D. K. Follman, M. B. Sumpter, and M. George, Factors influencing the self-efficacy beliefs of first-year engineering students, J. Eng. Educ. **95**, 39 (2006).

[73] R. W. Lent, H. B. Sheu, D. Singley, J. A. Schmidt, L. C. Schmidt, and C. S. Gloster, Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students, J. Vocat. Behav. **73**, 328 (2008).

[74] A. Uitto, Interest, attitudes and self-efficacy beliefs explaining upper-secondary school students' orientation towards biology-related careers, Int. J. Sci. Math. Educ. **12**, 1425 (2014).

[75] L. Ainscough, E. Foulis, K. Colthorpe, K. Zimbardi, M. Robertson-Dean, P. Chunduri, and L. Lluka, Changes in biology self-efficacy during a first-year university course, CBE Life Sci. Educ. **15**, 1 (2016).

[76] S. M. Villafañe, C. A. Garcia, and J. E. Lewis, Exploring diverse students' trends in chemistry self-efficacy throughout a semester of college-level preparatory chemistry, Chem. Educ. Res. Pract. **15**, 114 (2014).
[77] R. R. McCrae and P. T. Costa, Validation of the five-factor model of personality across instruments and observers, J. Pers. Soc. Psychol. **52**, 81 (1987).
[78] R. R. McCrae and P. T. Costa, More reasons to adopt the five-factor model, Am. Psychol. **44**, 451 (1989).
[79] R. R. McCrae and P. T. Costa, Rotation to maximize the construct validity of factors in the NEO Personality Inventory, Multivar. Behav. Res. **24**, 107 (1989).
[80] L. R. Goldberg, The development of markers for the Big-Five factor structure, Psychol. Assess. **4**, 26 (1992).
[81] O. P. John, E. M. Donahue, and R. L. Kentle, *The Big Five Inventory versions 4a and 54* (Institute of Personality and Social Research, University of California, Berkeley, CA, 1991).
[82] A. E. Poropat, A meta-analysis of the five-factor model of personality and academic performance, Psychol. Bull. **135**, 322 (2009).
[83] N. S. Cole, *The ETS Gender Study: How Females and Males Perform in Educational Settings* (Educational Testing Service, Princeton, NJ, 1997).
[84] G. V. Caprara, M. Vecchione, G. Alessandri, M. Gerbino, and C. Barbaranelli, The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study, Br. J. Educ. Psychol. **81**, 78 (2011).
[85] K. D. Multon, S. D. Brown, and W. Robert, Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation, J. Counsel. Psychol. **38**, 30 (1991).
[86] T. A. Judge and R. Ilies, Relationship of personality to performance motivation: A meta-analytic review, J. Appl. Psychol. **87**, 797 (2002).
[87] I. Sanchez-Cardona, R. Rodriguez-Montalbán, E. Acevedo-Soto, K. N. Lugo, F. Torres-Oquendo, and J. Toro-Alfonso, Self-efficacy and openness to experience as antecedent of study engagement: An exploratory analysis, Procedia Soc. Behav. Sci. **46**, 2163 (2012).
[88] G. V. Caprara, G. Alessandri, L. D. Giunta, L. Panerai, and N. Eisenberg, The contribution of agreeableness and self-efficacy beliefs to prosociality, Eur. J. Pers. **24**, 36 (2010).
[89] G. Chen, W. J. Casper, and J. M. Cortina, The roles of self-efficacy and task complexity in the relationships among cognitive ability, conscientiousness, and work-related performance: A meta-analytic examination, Hum. Perform. **14**, 209 (2001).
[90] L. Dörrenbächer and F. Perels, Self-regulated learning profiles in college students: Their relationship to achievement, personality, and the effectiveness of an intervention to foster self-regulated learning, Learn. Individ. Diffe. **51**, 229 (2016).
[91] M. K. Udo, G. P. Ramsey, and J. V. Mallow, Science anxiety and gender in students taking general education science courses, J. Sci. Educ. Technol. **13**, 435 (2004).
[92] D. Hestenes, M. Wells, and G. Swackhamer, Force concept inventory, Phys. Teach. **30**, 141 (1992).
[93] K. Williams, Understanding communication anxiety and gender in physics, J. Coll. Sci. Teach. **30**, 232 (2000), http://www.jstor.com/stable/42991212.
[94] N. Hall and D. Webb, Instructor's support of student autonomy in an introductory physics course, Phys. Rev. Phys. Educ. Res. **10**, 020116 (2014).
[95] US News and World Report: Education, https://premium.usnews.com/best-colleges.
[96] W. K. Hofstee, B. de Raad, and L. R. Goldberg, Integration of the Big Five and circumplex approaches to trait structure, J. Pers. Soc. Psychol. **63**, 146 (1992).
[97] O. P. John, L. P. Naumann, and C. J. Soto, Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues, in *Handbook of Personality: Theory and Research* (Guilford Press, New York, NY, 2008), p. 114.
[98] D. Van der Linden, J. te Nijenhuis, and A. B. Bakker, The general factor of personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study, J. Res. Pers. **44**, 315 (2010).
[99] P.R. Pintrich, D.A.F. Smith, T. Garcia, and W.J. Mckeachie, Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ), Educ. Psychol. Meas. **53**, 801 (1993).
[100] J. Stewart, R. Henderson, L. Michaluk, J. Deshler, E. Fuller, and K. Rambo-Hernandez, Using the social cognitive theory framework to chart gender differences in the developmental trajectory of STEM self-efficacy in science and engineering students, J. Sci. Educ. Technol. **29**, 758 (2020).
[101] 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).
[102] R. M. Baron and D. A. Kenny, The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, J. Pers. Soc. Psychol. **51**, 1173 (1986).
[103] S. Cwik and C. Singh, Damage caused by societal stereotypes: Women have lower physics self-efficacy controlling for grade even in courses in which they outnumber men, Phys. Rev. Phys. Educ. Res. **17**, 020138 (2021).
[104] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1986).
[105] T. L. Rosenthal and B. J. Zimmerman, *Social Learning and Cognition* (Academic Press, New York, NY, 1978).
[106] D. H. Schunk and C. W. Swartz, Goals and progress feedback: Effects on self-efficacy and writing achievement, Contemp. Educ. Psychol. **18**, 337 (1993).
[107] D. J. Follmer, S. E. Zappe, E. D. Gomez, E. W. Gomez, and S. E. Haydt, Changing the conversation: Impact of a seminar-based classroom innovation on student perceptions of engineering, Int. J. Eng. Educ. **33**, 519 (2017), https://www.ijee.ie/contents/c330217A.html.
[108] D. Huffman, Effect of explicit problem solving instruction on high school students' problem solving

performance and conceptual understanding of physics, J. Res. Sci. Teach. **34**, 551 (1997).

[109] E. Scanlon, J. Schreffler, W. James, E. Vasquez, and J. J. Chini, Postsecondary physics curricula and universal design for learning: Planning for diverse learners, Phys. Rev. Phys. Educ. Res. **14**, 020101 (2018).

[110] E. Scanlon, T. Legron-Rodriguez, J. Schreffler, E. Ibadlit, E. Vasquez, and J. J. Chini, Postsecondary chemistry curricula and universal design for learning: Planning for variations in learners abilities, needs, and interests, Chem. Educ. Res. Pract. **19**, 1216 (2018).