

Examining factors influencing the development and evolution of self-efficacy and sense of belonging in a physics class

Danielle Maldonado[✉] and John Stewart^{✉*}

West Virginia University, Department of Physics and Astronomy, Morgantown West Virginia, 26506, USA



(Received 1 December 2025; accepted 29 June 2026; published 31 August 2026)

This study analyzed the relation between self-efficacy (SE) and sense of belonging (BL) in introductory physics courses by measuring these variables at two time points, early in the semester and midsemester. The study was performed in the introductory, calculus-based mechanics class at a large university in the United States and includes 1373 complete records. An extensive set of Observed Prior Causes (OPCs), including Force and Motion Conceptual Evaluation pretest scores, college and high school preparation variables, high school physics and math variables, personality variables, and demographics, were found to be predictive of SE or BL. Early in the semester, SE and BL were highly correlated ($r = 0.56$); after controlling for OPCs, $r = 0.37$ of this correlation remained. SE fully explained the relation of BL to midsemester test average, while OPCs fully explained the relation of SE to test average. Changes to SE and BL in response to course feedback in the form of test averages were also explored. Test average explained little of the relation between early and midsemester SE or BL. The analysis supported a model where test average predicted midsemester SE, which in turn predicted midsemester BL. Mediation analysis found both directed and correlational paths were important in explaining the complex relation of these variables. Some moderation effects were detected, but were generally small.

DOI: [10.1103/lkdr-cz49](https://doi.org/10.1103/lkdr-cz49)

I. INTRODUCTION

The American Institute of Physics recently released preliminary results from the 2023–2024 Nationwide Survey of High School Physics, which suggest that for the first time since 1987, enrollment in high school (HS) physics courses across the United States has decreased [1]. Retention in college Science, Technology, Engineering, and Mathematics (STEM) programs has long been a critical area of study to address expected shortfalls in the STEM workforce [2]. For physics programs, student retention can be critical to departmental survival as smaller programs are pressured to close due to low enrollment [3]. A study of two institutions with different levels of admissions selectivity found that, while there were many different points at which students are most at risk of leaving physics majors, students are at the highest risk of leaving college altogether during the spring semester of freshman year. Furthermore, precollege academic readiness variables, including being prepared to enroll in Calculus 1 in the first semester, high school grade point

average (GPA), and ACT composite scores, were highly predictive of persistence in physics degrees [4]. Instructional interventions and curriculum design also influence student engagement, attitudes toward physics, time to degree completion, and other factors related to physics retention [5–7].

Beyond academic preparation, there is often a social component to persistence in a given field. Network analysis has been utilized to show that social integration, particularly close connections with peers whom they reach out to often, can predict student persistence across introductory physics sequences [8]. Furthermore, students “in the middle of the pack” (earning a B, B–, or C+ as a final course grade) were most likely to persist through the course sequence if they developed strong connections with their peers outside the classroom [9].

Because retaining students in physics majors is essential to meet the demands of the workforce, it is important to investigate how academic performance and social support affect undergraduate physics students. Tinto’s Model of Departure [10] proposes that students’ decisions to persist throughout college are associated with many factors, including goal commitments, institutional experiences, student effort, educational outcomes, and academic and social integration. Tinto argues that students who are socially and academically involved put in more effort toward persistence, and that one form of integration is not more important than the other.

*Contact author: jcstewart1@mail.wvu.edu

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

For example, high-achieving students may still feel isolated without peer support, while strong social circles may offset the concern of poor academic performance [10].

While earlier versions of Tinto's Departure Model focus on academic retention from the perspective of the institution, more recent versions [11] examine the issue from the viewpoint of a student's motivation to persist. Students are not concerned with the benefits of high graduation rates to the institution but rather are motivated by a desire to earn a degree. To persist throughout several years of academic challenges, Tinto suggests that students' motivation is influenced by interactions between their academic goals, self-efficacy (SE), sense of belonging (BL), and perceptions of the curriculum. Self-efficacy informs a student's response to task management and goal orientation in the face of challenges. Meanwhile, a sense of belonging is important to academic persistence because it creates a sense of community for students to relate to when challenges arise. Students' perceptions of their daily interactions with instructors, peers, and the institution at large influence their willingness to engage in practices that lead to persistence. Physics self-efficacy and sense of belonging in physics are two important variables that can be used to measure students' beliefs about their performance in physics and how integrated they feel in their majors and departments.

A. Self-efficacy

Self-efficacy is defined by Bandura as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives" [12,13]. Physics self-efficacy helps students assess their perceived ability to complete assignments, lab work, exams, and other activities in the physics classroom. This assessment is formed from a variety of factors, including current and prior experiences in physics. Self-efficacy is commonly studied as a predictor of student academic achievement, retention, and persistence in the face of academic challenges [14–16]. For example, students with low self-efficacy are more likely to switch from correct to incorrect answers on physics assessments, especially as the difficulty of the problem increases [17]. Self-efficacy is also related to students' approach to academic tasks and skill development through the use of self-regulation [18–20].

Although high school boys and girls both experience lower levels of self-efficacy toward physics compared to other STEM and non-STEM activities, female students report drastically lower levels of physics self-efficacy than their male counterparts [21]. Even though the percentage of women enrolling in college physics majors has increased from around 15% to 25% in the past 30+ years [22,23], female students still consistently report lower levels of self-efficacy in introductory physics upon entry

into their programs and are more susceptible to negative shifts in self-efficacy over time [24,25]. However, regardless of gender, students often report lower levels of self-efficacy at the end of the academic semester compared to the beginning [26]. Contributions to gender differences in physics self-efficacy may include different sources of self-efficacy [12,27–29], variations in teaching and classroom styles [30,31], perceived recognition from instructors [32], personality [33,34], etc. Some studies have shown that gender differences persist even in classrooms where women outnumber men [35,36]. This suggests that gender differences in physics self-efficacy may not solely be rooted in representation but in deeper social stereotypes about gender minorities in the sciences and/or differences in motivational beliefs that contribute to students' self-efficacy.

B. Sense of belonging

Sense of belonging is a powerful social motivator [37]. Defined as a perception of being valued, accepted, and a legitimate member of a group, Lewis *et al.* outline four major factors influencing sense of belonging development: (i) quantity and qualities of peers and role models, (ii) ability stereotypes about people who succeed in STEM, (iii) direct interventions in STEM settings, and (iv) nonacademic outside influences [38]. Sense of belonging has both social and academic implications for students in STEM fields such as physics. For example, a student's sense of belonging in a classroom may affect their academic achievement by increasing or decreasing anxiety surrounding exam performance or changing a student's beliefs about their capacity to succeed. Mathematics anxiety [39,40], science anxiety [41–43], and stereotype threat [44,45] have all been studied as possible non-cognitive influences on performance and subsequently on belonging.

Gender and underrepresented minority and/or persons excluded based on ethnicity or race (PEER) status are often used to compare students' sense of belonging in physics courses. For example, female students are more likely to report lower levels of belonging in physics than their peers [46,47]. Walton *et al.* demonstrated how two different interventions were able to increase women's sense of belonging in engineering majors where women were largely underrepresented (enrollment less than 20%) [48]. Although many previous studies have highlighted strong positive correlations between sense of belonging and persistence [38,49], recent research is challenging this relationship among women in physics [50]. For women of color, having supportive peer relationships may be particularly important for their persistence in STEM [51]. Interviews of Black women with physics doctoral degrees highlighted feelings of isolation from social and study groups as a significant obstacle to success during their graduate programs [52].

Other factors, such as first-generation college student (FGCS) status and disability status, may also be significant predictors of sense of belonging in physics. One study interviewed 11 students with underrepresented identities (URIs) and identified themes of feeling a lack of belonging and community in their physics programs [53], and that students with more than one URI often reported that these feelings are compounded as they navigate multiple identities [53,54].

C. Research questions

This study explores the relationships between physics self-efficacy and sense of belonging in physics classes as they relate to an extensive set of cognitive and non-cognitive variables, called observed prior causes (OPCs) in this work. These include high school preparation, college academic performance, and noncognitive factors. It also examines the evolution of self-efficacy and sense of belonging during an introductory physics class and how class performance is associated with that evolution. This study seeks to answer the following research questions:

RQ1: What OPCs are most important in predicting early-semester physics self-efficacy and sense of belonging?

RQ2: How are early-semester physics self-efficacy and sense of belonging related? Which OPCs are important in explaining their relationship?

RQ3: How do self-efficacy and sense of belonging predict class performance, controlling for OPCs?

RQ4: How does academic performance in a physics class predict the evolution of self-efficacy and belonging during the class? How do mediating or moderating variables predict the evolution?

RQ5: How would the conclusions of this study change if some variables were not available?

The last research question is motivated by commentary from Weissman [55,56], who cautions that causal inferences drawn from path models like those presented in the current work are susceptible to the exclusion of unobserved variables.

D. Theoretical framework

Our hypotheses about the factors associated with the development and evolution of self-efficacy and sense of belonging in physics classes are informed by Bandura's [12] proposed four primary sources of self-efficacy: mastery experiences, vicarious experiences, social persuasion, and physiological states. Mastery experiences result from successful and failed attempts at performing a task. Vicarious experiences include observing other students successfully perform tasks and overcome challenges. Social persuasion includes receiving support from peers and instructors in their belief that they can complete a task successfully. Physiological states influence self-efficacy by informing students' beliefs in their ability to cope with

stressful situations, such as high-stakes performance assessments. Although there are multiple sources of self-efficacy, Bandura argues that mastery experiences are the most influential and authentic source of efficacy information because repeated feedback from success and failure builds resilience in the face of obstacles [16]. Course performance measurements, such as examination grades, should be examples of mastery experiences. However, Bandura further argues that performance measurements alone are not entirely indicative of one's capability. The impact of performance on self-efficacy is dependent on personal features such as prior beliefs about one's abilities, the perceived difficulty of a task, effort expenditure, aid and external circumstances, and other factors that affect how one interprets their performance on a given task. Therefore, mastery experiences encompass students' self-appraisal of their cognitive, behavioral, and self-regulatory skills when completing difficult tasks, not solely the completion of the task itself.

Theories of sources of belonging often focus on peer interactions and the number of those interactions [37]. Tinto's discussion of belonging as a predictor of motivation to persist highlights that while belonging can be influenced by representation in departments from faculty and staff and messaging from interactions in and out of the classroom, ultimately, sense of belonging is derived from students' perceptions of their interactions with others regardless of quantity. Even one negative interaction can be detrimental to a student's sense of belonging [11]. While the present study does not directly measure the quantity or quality of social interactions, it does include a measurement of personality; many facets of personality may relate to the types and frequency of social and academic interactions.

1. Hypotheses

Following Bandura's and Tinto's models, we hypothesize that exam grades, a mastery experience, are primarily related to changes in self-efficacy, and that midsemester changes in self-efficacy, in turn, predict midsemester changes in belonging. We hypothesize that feeling one belongs in a physics class and that one can complete the requirements of the class successfully should confer an advantage in examination grades, following studies that relate self-efficacy to academic achievement [57,58] and studies showing a relationship between sense of belonging and self-efficacy [59,60]. We statistically test the hypothesis that OPCs only affect midsemester variables through their actions on early-semester variables; i.e., early-semester observations of self-efficacy and belonging fully account for possible associations with presemester causes. While the authors could not find any prior work that supports this hypothesis, it has important consequences for future work relating self-efficacy to belonging. Finally, we assume that the effects of unmeasured variables, the

unobserved prior causes, act only through early-semester self-efficacy and belonging. No statistical support for this assumption is possible, but if the prior hypothesis is supported, it will suggest that this assumption is reasonable.

2. Path models

Path models encapsulate a series of regression models where the terminal node of a group of edges represents the dependent variable. Directed edges point from the parent node to the child node. All parent nodes of edges incident on the dependent variable are independent variables in the regression. Edges in the model can be used to express the causal assumption (hypothesis) that the parent node of the edge influences the child node of the edge [61,62]. Nodes not connected by an edge represent the causal assumption that the nodes are independent of each other, given the other edges in the model. The path model representing our hypotheses is shown in Fig. 1. Rectangular nodes represent measured variables and oval nodes represent unobserved variables.

It is important to stress that path models encapsulate theoretical assumptions about causal relationships. In general, statistical analyses cannot directly prove causal relationships. However, should the solid edges in Fig. 1 be statistically significant and the dashed edges nonsignificant, this will support (but not prove) our causal hypotheses. For an exploration of the use of path models in the context of physics education research (PER), see Adlakha and Kuo [63].

The path model proposed in Fig. 1 is based on the theoretical frameworks for self-efficacy and belonging. These frameworks are particularly important in proposing an early-semester correlational model between SE_E and BL_E and a midsemester directed model. This structure

conforms with our interpretation that test average is primarily a mastery experience, which Bandura's model relates to self-efficacy development, and that the early-semester model represents a long-term interplay between self-efficacy and belonging development. These structures cannot be established by statistical analysis. In fact, related path models have equivalent overall statistical properties. The proposed model, where test average predicts SE_M , which predicts BL_M , has the same overall statistical fit as an alternative model where test average predicts BL_M , which predicts SE_M , or a model where test average predicts both SE_M and BL_M while the two variables covary. While overall statistical fit is identical, each of these models would predict different values for regression coefficients and the significance of those coefficients.

II. METHODS

A. Sample

This study was performed with data collected from Spring 2017 to Spring 2023 at a large land-grant university in the Eastern United States. Semesters presented in an online format due to the COVID-19 pandemic were removed (Spring 2020–Spring 2021). The institution enrolled 24,000 students in Fall 2023. At that time, the university's general undergraduate population was 82% White, 6% reporting two or more races, 4% Hispanic or Latino, 3% Black or African American, 2% Asian, 2% U.S. nonresident, and other groups each with 1% or less [64].

The study was conducted in a calculus-based introductory mechanics course, generally taken by physical scientists and engineers. Students enrolled in this class were required to attend three 50-min lectures each week and one 3-h laboratory session. The class was designed and overseen by the same lead instructor throughout the data

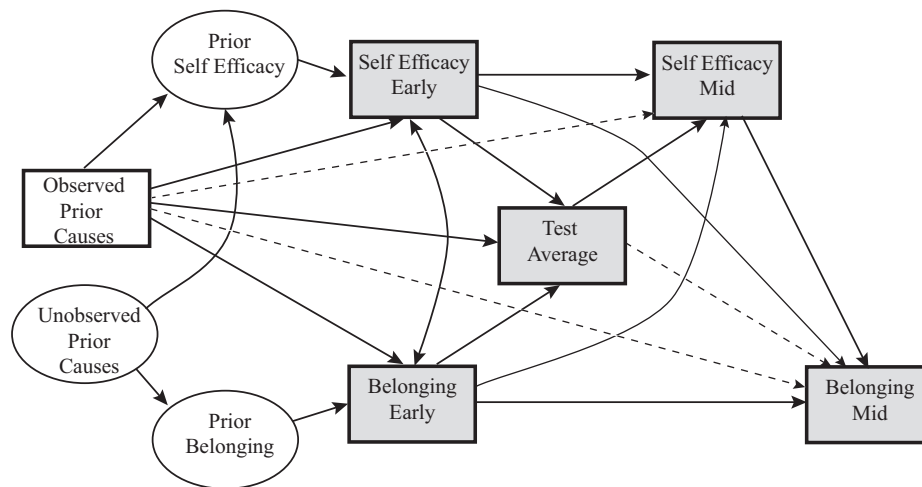


FIG. 1. Path model incorporating causal hypotheses. Rectangular nodes represent measured variables. Oval nodes represent unobserved (unmeasured) variables. Direct arrows represent regressions. Lines with arrowheads on both ends represent correlations. Dashed lines represent relationships hypothesized to be insignificant.

collection period. The instructor was knowledgeable about PER research-based techniques and implemented many of these techniques in the class. Many instructors taught lecture sections during the data collection period; however, all instructors implemented the pedagogical practices of the lead instructor, which included Peer Instruction [65]. The laboratory sessions implemented many research-based practices, including hands-on open inquiry activities, whiteboard activities, group problem-solving, and traditional experiments.

Students' general college performance data and demographic information were accessed from institutional records. Students without general high school information [high school GPA (HSGPA), ACT or SAT scores, etc.] or college information (college GPA, credits earned, etc.) were excluded from this study. High school preparation and noncognitive factors were measured by four surveys delivered to students electronically. Students who did not complete all four surveys were excluded from the analysis. Prior physics knowledge was measured with the Force and Motion Conceptual Evaluation (FMCE) [66,67] as a pretest, which was administered on the second day of class. Students who did not complete the pretest were also excluded from the analysis. Throughout the data collection period, the class enrolled 3235 students who completed the course for a letter grade. Removing students with incomplete data resulted in a dataset of 1373 records used in this study.

Students included in the analysis were higher performing and of a somewhat different demographic composition than the class as a whole. Overall, the class had a four-test average of 68.4% and a class GPA of 2.7 on a four-point scale and included 21.8% women, 9.3% PEER students, and 18.3% FGCS students. For the analysis sample, the four-test average was 71.3%, the class GPA was 3.1, and included 29.7% women, 6.6% PEER students, and 16.6% FGCS students. Students not included had a four-test average of 66.2%, a class GPA of 2.4, and included 14.9% women, 11.4% PEER students, and 19.7% FGCS students. The Supplemental Material [68] includes further discussion of the data filtering process. The differences resulted primarily from whether the student answered all four surveys. While somewhat different than the overall sample, the general demographic patterns remained with women, PEER, and FGCS students being substantially in the minority in the class. The class studied has also been examined in other PER studies [70–73].

B. Variables

Table I lists all the variables used in this study. Each variable is accompanied by an abbreviation, a description of the variable, and a symbol to denote if the variable is continuous (C) or dichotomous (D). All continuous variables were standardized by subtracting the mean and dividing by the standard deviation. All calculations were

performed with the “R” software system [74]. Basic descriptive statistics for all variables are presented in Supplemental Material [68]. Two categorical variables are used: HS physics (7 levels) and HS mathematics (4 levels). When these variables are used in a regression, one level must be selected as a reference level against which changes are measured. For HS physics, the reference level was not taking a physics class in high school, and for HS math, it was not taking calculus in high school.

C. Instruments

1. Self-efficacy

Self-efficacy was measured using items from the Self-Efficacy for Learning and Performance subscale from the Motivated Strategies for Learning Questionnaire (MSLQ) [75]. This widely used subscale [76] has strong validity and asks students to rate their agreement with various statements on a five-point Likert scale. For this study, three items were used to measure self-efficacy:

- 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.

The original eight-item scale was reduced to six items and specialized to specify self-efficacy in physics classes as part of a prior study [77]. Modifying the MSLQ by word substitution for specific domains has been employed in previous studies [78]. This study further reduced the scale to three items to produce early-semester surveys of acceptable length.

2. Sense of belonging

Students' sense of belonging in their physics class was measured using three items adapted from Good, Rattan, and Dweck's “Math Sense of Belonging” instrument [49]:

- I feel accepted when I am in physics classes and with students in my physics classes.
- I feel I fit in when I am in physics classes and with students in my physics classes.
- I feel respected when I am in physics classes and with students in my physics classes.

While both self-efficacy and sense of belonging subscales were modified from their original publications, each scale was revalidated using student interviews to verify that items were interpreted as intended, followed by an examination of responses to open-ended versions of the items. The internal reliability of each subscale after modification was high (Cronbach's alpha, $\alpha > 0.9$). Given the highly correlated relation of self-efficacy and sense of belonging, the reduction of the six-item scale used in prior works to three items raises concerns that the three-item

TABLE I. List of variables. Type indicates whether the variable is continuous (C) or dichotomous (D).

Panel	Abbreviation	Type	Description
Belonging	BelongEarly	C	Sense of belonging in physics class—early-semester
	BelongMid	C	Sense of belonging in physics class—midsemester
Self-efficacy	SelfEffEarly	C	Self-efficacy toward physics class—early-semester
	SelfEffMid	C	Self-efficacy toward physics class—midsemester
Pretest	Pretest	C	FMCE pretest percentage
Test average	TestAve	C	Test 1 and 2 average
Repeat	Repeat	D	Is the student repeating the class?
College	Complete	C	Percentage of college classes completed (before class)
	CGPA	C	College grade point average before class
	Credit	C	Credit hours completed before class
Math ready	MathReady	D	Was the student's first college mathematics class Calculus 1 or higher?
HS general	ACTM	C	ACT or SAT mathematics percentile score
	ACTV	C	ACT English or SAT verbal percentile score
	HSGPA	C	High school grade point average
AP general	AP.NMP	D	Does the student have AP credit excluding math and physics credit?
	AP.C.NMP	C	Number of nonmath or nonphysics classes with AP college credit
Transfer	TR.NMP	D	Does the student have transfer credit excluding math and physics credit?
	TR.C.NMP	C	How many nonmath and nonphysics transfer classes?
	TR.Phys	D	Does the student have transfer credit for physics?
	TR.Math	D	Does the student have transfer credit for math?
HS physics	HSP.NTake	D	High school physics not taken
	HSP.NAP.NA	D	High school physics class not AP—grade B, C, D
	HSP.NAP.A	D	High school physics class not AP—grade A
	HSP.APNP.NA	D	High school physics AP (test not passed)—grade B, C, D
	HSP.APNP.A	D	High school physics AP (test not passed)—grade A
	HSP.APP.NA	D	High school AP physics test passed—grade B, C, D
	HSP.APP.A	D	High school AP physics test passed—grade A
HS math	HSM.A	D	Was the grade in the student's most advanced high school math class an A?
	HSM.NCal	D	Was most advanced high school math class below calculus?
	HSM.NAP	D	Was most advanced high school math class calculus?
	HSM.APNP	D	Was most advanced high school math class AP calculus (test not passed)?
	HSM.APP	D	Was most advanced high school math class AP calculus (test passed)?
Personality	Cns	C	Personality facet—Conscientiousness
	Opn	C	Personality facet—Openness
	Nrt	C	Personality facet—Neuroticism
	Agr	C	Personality facet—Agreeableness
	Ext	C	Personality facet—Extraversion
Demographics	Gender	D	Does the student identify as female?
	FGCS	D	Is the student a first-generation college student?
	PEER	D	Does the student identify as a person excluded due to ethnicity or race (PEER)?

scale might no longer be able to resolve the constructs as separate entities. Factor analysis combining the two six-item scales measured midsemester showed excellent factor resolution between self-efficacy and sense of belonging items, with high loadings on different factors and with no cross-loadings at the 0.4 level. Confirmatory model fits were also good. A similar exploratory analysis with the three-item scales showed similar results. Items removed from the six-item scale were very highly correlated with those included, greater than 0.85 for both scales.

Both subscales were administered in a single survey given in the first full week of classes (Early), and in two separate surveys in the weeks following the second exam (Mid). Students received a small amount of course credit for their participation.

3. College preparation

College preparation variables were gathered from institutional records. This included the percent of college

classes completed before the class in which surveys were administered (Complete), the student's college GPA before the class (CGPA), the number of credit hours completed before the class (Credit), and if the student's first college math course was Calculus 1 or higher (MathReady). For most students in the class, Calculus 1 is the first mathematics course taken in their 4-year degree plans; as such, these students are on track to graduate in 4 years. Studies have shown that math-ready students tend to have superior academic outcomes in introductory physics [4,72,79].

4. High school preparation

HSGPA, ACT scores, and SAT scores were accessed from university records and converted to percentile scores using tables published by the testing companies. If ACT and SAT scores were both available, they were averaged. The ACT English score was used as students' verbal scores (ACTV). Many studies have explored the relationships between high school preparation and admission tests as predictors of college academic achievement [80–87]. Advanced Placement (AP General) and Transfer (Transfer General) data were taken from institutional records.

In a survey administered during the third week of class, students were asked in detail about the HS science and mathematics classes they had taken, the classification of these courses (e.g., normal, honors, and AP), and their class outcomes. The results of this survey were distilled into three variables: High School Math Last Grade, High School Math Last Class, and High School Physics. The HS Math variables contained information about the most advanced high school math class the student had taken: Not Calculus (HSM.NCal), Calculus Not AP (HSM.NAP), Calculus AP—Test Not Passed (HSM.APNP), and Calculus AP—Test Passed (HSM.APP), as well as if the student earned an A in the most advanced math class taken (HSM.A). HS Physics class information was summarized within a seven-level categorical variable, which captured both the student grade in the course (A or not A) and the type of course (no HS physics, non-AP, AP but AP exam not passed, and AP and AP exam passed). This seven-level variable was required to uniformly cover the dimensions of the kind of physics class and the performance in the class, while eliminating the interaction that students who did not take high school physics could not have a grade in high school physics. As shown in the descriptive table in Supplemental Material [68], this results in some levels with small occupation. The parameter estimates for these levels may be unreliable. For a more complete description of the conversion process from the raw survey responses into these variables, see Hewagallage *et al.* [72].

5. Personality

Personality was measured using the Big Five Inventory (BFI), administered in the first survey sent to students

early in the second week of class. The BFI characterizes personality with five facets: openness, conscientiousness, extraversion, agreeableness, and neuroticism. [88–90]. Openness describes the depth and complexity of one's mental and experiential life, their originality, and open-mindedness. Conscientiousness describes impulse control behaviors directed at goal-setting and task-completion, such as organizational skills and following rules. Extraversion encompasses traits such as sociability, assertiveness, and an energetic approach to the world. Agreeableness encompasses prosocial and communal behaviors, including altruism and trust. Finally, neuroticism refers to behaviors encompassing responses to stress and anxiety, such as nervousness. The instrument contains 44 items, each measured on a five-point Likert scale. The BFI has been used extensively in a wide variety of research [19,91,92], particularly as it relates to performance motivation [93].

Stajkovic *et al.* provide an excellent review of research relating BFI traits to self-efficacy and academic performance [94], noting that conscientiousness and neuroticism are the most consistent predictors of self-efficacy across the literature. They also present three models for the relationship between BFI traits, self-efficacy, ACT and GPA scores, and academic performance. By comparing harmonic means, they found that an *intrapersonal model*, where the effects of personality on academic performance were completely mediated by self-efficacy when controlling for past academic experiences, best explained the relationships between variables. While the links between personality and self-efficacy have been previously explored, the connections between personality and belonging are less clear. Furthermore, domain-specific research on these topics in physics is scarce. Using physics self-efficacy, physics belonging, and the BFI personality traits in a single study is a unique contribution to the literature.

III. RESULTS

SE and belonging (BL) were measured at two time points, early in the semester (SE_E and BL_E) and in the middle of the semester (SE_M and BL_M) after the grades of the second exam were available to students. Before entering a college physics class, students generally have some history of performance in high school mathematics and science classes. Most students in the class have also taken at least one college mathematics course. As such, they should have prior beliefs about their general ability to succeed in science and mathematics classes. As shown by the HS Physics variable (Table I, Supplemental Material [68]), 75% of students in this study have taken a physics class in high school and have had direct experiences with their academic achievement and feelings of belonging in that class. Early-semester measurements of SE and BL were taken in the second week of class before students received any feedback in the form of homework or test

grades. As such, early-semester measurements should be strongly related to prior feelings of SE and BL developed before the class and should also be informed by early-semester impressions of class requirements.

A. Correlation analysis

To explore RQ1, all variables in Table I (excluding TestAve, SE_E , BL_E , SE_M , and BL_M) were tested as possible OPCs to identify variables that were most important in predicting early-semester self-efficacy and belonging. To begin to understand the relationships between OPCs, SE_E , and BL_E , correlations between each variable with SE_E or BL_E were calculated and presented in Table I of Supplemental Material [68]. The strongest correlation among all variables was between SE_E and BL_E , $r = 0.56$. The OPCs most strongly correlated with SE_E included neuroticism ($r = -0.30$), openness ($r = 0.28$), conscientiousness ($r = 0.25$), and pretest ($r = 0.23$). OPCs most strongly correlated with BL_E included agreeableness ($r = 0.23$), neuroticism ($r = -0.23$), extraversion ($r = 0.23$), and conscientiousness ($r = 0.21$). The difference between the correlation of SE_E and BL_E and other correlations was dramatic. SE_E and BL_E explained 31% of the variance in each other, while the next most correlated variables only explained 9% of the variance in SE_E and 5% in BL_E .

A visualization of the correlation matrix is presented in Fig. 2 as a force-directed graph, which allows the identification of groups of related variables. More correlated variables are grouped closer together. Solid green lines represent positive correlations, while dashed red lines represent negative correlations. The thickness of each line indicates the magnitude of the correlation between two variables. This visualization shows that SE_E and BL_E , highlighted in yellow, are highly correlated with each other and closely related to the five personality variables. These seven variables form a relatively isolated cluster. SE_E is also correlated with some measures of prior physics performance. Academic variables are generally strongly correlated with each other and have weaker correlations with SE_E and BL_E . Demographic variables generally have a weak relation to SE and BL, except for gender and SE_E . An alternative visualization of the correlation matrix is presented in Supplemental Material [68] to better understand the magnitude of relationships between variables.

B. Variable importance

To explore how collections of similar variables predicted SE_E or BL_E , some variables were examined in groups referred to as “panels” in this work. Predictive power is a measure of a variable’s importance in a regression. Two different metrics were used to characterize how OPCs predict SE_E or BL_E as presented in Tables II and III of Supplemental Material [68]. First, R_f^2 represents

the fraction of variance explained when SE_E or BL_E is a dependent variable in a linear regression model and the only independent variables are contained within the panel. This provides a general measure of the relationship between the panel and the dependent variable, ignoring all other variables. The personality panel was the strongest predictor of both SE_E and BL_E when the only variables in the regression model were personality variables, $R_{SE,f}^2 = 0.17$ and $R_{BL,f}^2 = 0.14$.

The second metric, ΔR_l^2 , measures the change in R^2 when the paneled variables are the last added to a regression model containing all other OPCs. This measures the unique variance explained by the panel and not by other variables. Again, the personality panel was the most important when added last to the model, $\Delta R_{SE,l}^2 = 0.06$ and $\Delta R_{BL,l}^2 = 0.05$.

C. Covariance of self-efficacy and belonging

Early and midsemester self-efficacy and belonging both strongly covary ($r_E = 0.56$ and $r_M = 0.52$, respectively). These constructs likely coevolve over time. Some events may modify SE, which may in turn affect BL. Conversely, some events may modify BL, which may in turn affect SE. Other events may simultaneously impact both variables.

The most direct method of exploring these relationships is to calculate the amount of covariance between SE_E and BL_E explained by the paneled OPC variables. First, the total covariance between SE_E and BL_E without any other variables present, $\text{cov}_T(SE_E, BL_E)$, was calculated. Then, using structural equation modeling, the remaining covariance when a single panel is added to predict SE_E and BL_E , $\text{cov}_E(SE_E, BL_E)$, was calculated. This is analogous to the error variance in linear regression or the variance not explained by the model. In analogy to the calculation of R^2 in a regression analysis, the fraction of the first-in covariance $R_{CV,f}^2$ explained is calculated as

$$R_{CV,f}^2 = 1 - \frac{\text{cov}_E(SE_E, BL_E)}{\text{cov}_T(SE_E, BL_E)}. \quad (1)$$

A last-in $\Delta R_{CV,l}^2$ was also calculated as the change in $R_{CV,l}^2$ when the paneled variables are the last added to a model containing all other OPCs. The results of this calculation are shown in Tables II and III of Supplemental Material [68]. Once again, the personality panel stood out as explaining the most covariance, both when the personality variables were the only variables included in the model and when they were the last variables added.

D. Early-semester analysis

The analyses of the previous sections included all variables available to the study. This section examines

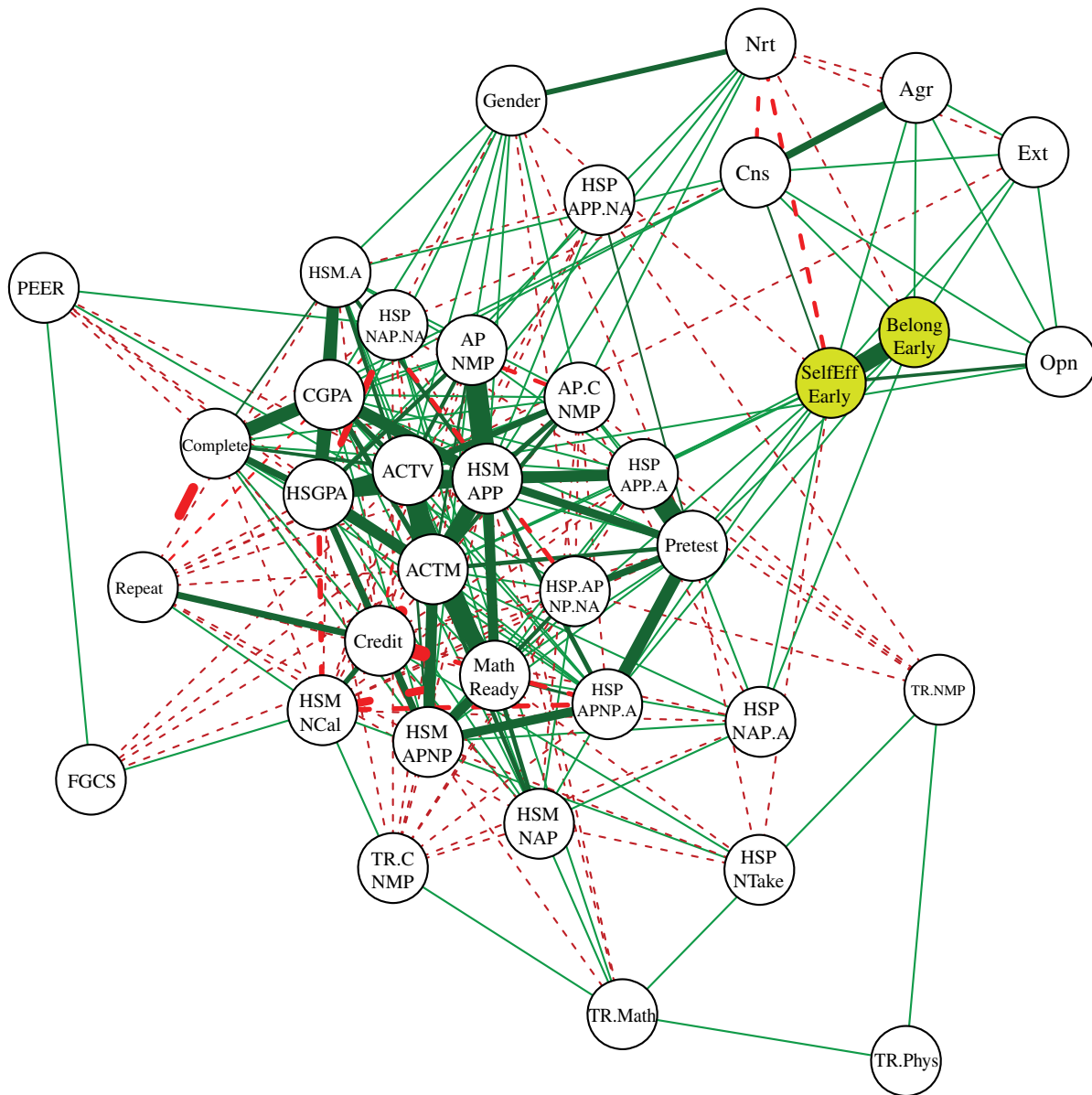


FIG. 2. Correlation matrix. Green (solid) lines represent positive correlations, while red (dashed) lines represent negative correlations. Thicker lines represent stronger correlations.

the combinations of variables that are most important for predicting SE and BL early in the semester. Regression analysis using large sets of related variables is susceptible to Type I error inflation and multicollinearity. To build and optimize regression models, many statistical tests are performed to identify significant regression coefficients, inflating the Type I error rate. To reduce the number of statistical tests, the models in this study were constructed to include only variables identified as substantial contributors by variable importance analysis and that have a theoretical basis for their inclusion. Furthermore, many of the academic OPCs are strongly correlated; including all OPCs in a regression model can distort the regression results of correlated pairs. Type I error inflation was

corrected by applying a Bonferroni correction to adjust significance thresholds, thereby correcting for the effect of performing multiple statistical tests.

1. Variable selection

The variables included in the regression analyses were selected by examining the first-in and last-in paneled variable importance results and individual correlation and variable importance results, both shown in Supplemental Material [68]. To balance individual predictive power and collinearity, the selection process proceeded as follows. The Repeat, AP General, and Transfer groups increased R^2 by less than 0.005 in both the first-in and last-in analyses.

TABLE II. Early-semester regressions. β is the standardized regression coefficient and p the probability. DV indicates which variable is the dependent variable, while dashes ($\dots$) denote variables intentionally excluded from the model. The variance explained by each model is measured by R^2 . Italicized entries denote values that are not statistically significant but are included as part of a categorical paneled variable. Only significant variables are reported. The Bonferroni-corrected α_C thresholds for SE_E and BL_E are $\alpha = 0.05$ becomes $\alpha_C(0.05) = 0.0028$, $\alpha = 0.01$ becomes $\alpha_C(0.01) = 0.00056$, and $\alpha = 0.001$ becomes $\alpha_C(0.001) = 0.00006$.

Panel	Variable	Self-efficacy		Sense of belonging	
		β_{SE_E}	p_{SE_E}	β_{BL_E}	p_{BL_E}
Belonging	BelongEarly	$\dots$	$\dots$	DV	DV
Self-efficacy	SelfEffEarly	DV	DV	$\dots$	$\dots$
Pretest	Pretest	0.12	0.00002	0.12	0.00002
College	CGPA				
HS general	ACTM	0.11	0.00002		
HS physics	HSP.NAP.NA	<i>0.10</i>	<i>0.20134</i>	<i>0.13</i>	<i>0.11818</i>
	HSP.NAP.A	0.27	0.00002	0.22	0.00101
	HSP.APNP.NA	<i>0.03</i>	<i>0.77089</i>	<i>0.05</i>	<i>0.61983</i>
	HSP.APNP.A	0.34	0.00005	<i>0.14</i>	<i>0.10389</i>
	HSP.APP.NA	<i>0.55</i>	<i>0.05244</i>	<i>0.36</i>	<i>0.22622</i>
	HSP.APP.A	<i>0.47</i>	<i>0.00699</i>	<i>0.38</i>	<i>0.03746</i>
HS math	HSM.A				
Personality	Cns	0.18	0.00000	0.09	0.00239
	Opn	0.22	0.00000	0.13	0.00000
	Nrt	-0.18	0.00000	-0.09	0.00112
	Agr			0.15	0.00000
	Ext			0.15	0.00000
Demographics	Gender	-0.27	0.00000		
	FGCS				
	PEER				
	R^2	0.25	0.00000	0.17	0.00000

As such, these variables were removed from the analysis. Examining first-in $R^2_{SE,f}$ and $R^2_{BL,f}$, the College, Math-Ready, and HS Math panels each increased R^2 by approximately 0.01. Important collinearity relationships exist between these groups, some of which are a result of the math placement process at the institution studied—the math placement process uses ACTM scores as one of the metrics considered. Examining the HS General panel, ACTM, ACTV, and HSGPA were all strongly correlated, which is supported by Fig. 2, as well as Fig. 2 in Supplemental Material [68]. However, ACTM explained more variance in SE_E or BL_E scores than ACTV or HSGPA (Table II, Supplemental Material [68]). As such, ACTM was retained, and ACTV and HSGPA were eliminated to reduce multicollinearity and preserve statistical power.

The College panel (Complete, CGPA, and Credit) contains three highly correlated variables with theoretical dependencies. Complete and CGPA are strongly correlated, as students who have successfully completed more courses before the class studied have higher CGPAs. However, CGPA has been used in many retention studies and explained more first-in $R^2_{SE,f}$ and covariance (Table II,

Supplemental Material [68]). Therefore, CGPA was retained, and Complete was removed.

While MathReady is not in the HS General panel, it is highly correlated with Credit. Students who are not ready to take Calculus 1 upon entering university must take more math classes before enrolling in the course studied, thus increasing the number of credit hours completed before the class. MathReady explained more variance than Credit and was (temporarily) retained while Credit was removed. MathReady is also highly correlated with ACTM because ACTM is one of the primary factors used to determine students' initial math placement. However, ACTM explained substantially more variance than MathReady, so ACTM was retained and MathReady was eliminated.

The last panel analyzed for multicollinearity was HS Math. While ACTM is highly correlated with the type of the last math class taken in high school, it is less related to the grade in that class. Therefore, we retained the variable containing students' last high school math grade (HSM.A) but removed the type of math class taken. We note that many institutions have stopped requiring ACTM scores for admission. Unfortunately, the more universally available

TABLE III. Midsemester regressions. β is the standardized regression coefficient and p the uncorrected probability that β was observed by chance. DV indicates which variable is the dependent variable, while dashes ($\cdots$) denote variables purposely excluded from the model. Variables with no entry were not significant. The variance explained by each model is measured by R^2 . Italicized entries denote values that are not statistically significant but are included as part of a panel or a significant moderator. The 0.05 Bonferroni-corrected significance level was $\alpha_C(0.05) = 0.00250$ for test average, $\alpha_C(0.05) = 0.00185$ for SE_M , and $\alpha_C(0.05) = 0.00192$ for BL_M .

Panel	Variable	Test average		Self-efficacy		Sense of belonging	
		β_{TA}	p_{TA}	β_{SE_M}	p_{SE_M}	β_{BLM}	p_{BLM}
Belonging	BelongEarly			<i>0.06</i>	<i>0.02055</i>	0.46	0.00000
	BelongMid	$\cdots$	$\cdots$	$\cdots$	$\cdots$	DV	DV
Self-efficacy	SelfEffEarly			0.36	0.00000	<i>-0.09</i>	<i>0.00605</i>
	SelfEffMid	$\cdots$	$\cdots$	DV	DV	0.36	0.00000
Test average	TestAve	DV	DV	0.35	0.00000		
Pretest	Pretest	0.26	0.00000				
College	CGPA	0.33	0.00000	-0.10	0.00013		
HS general	ACTM	0.31	0.00000				
HS physics	HSP.NAP.NA	<i>0.08</i>	<i>0.22084</i>				
	HSP.NAP.A	<i>0.15</i>	<i>0.00372</i>				
	HSP.APNP.NA	0.27	0.00041				
	HSP.APNP.A	0.25	0.00023				
	HSP.APP.NA	<i>0.57</i>	<i>0.01356</i>				
	HSP.APP.A	<i>0.21</i>	<i>0.14738</i>				
HS math	HSM.A					0.09	0.05632
Personality	Agr						
	Cns						
	Opn						
	Nrt			-0.08	0.00096		
	Ext						
Demographics	Gender	-0.16	0.00077				
	FGCS					-0.02	0.77494
	PEER						
Interactions	BelongEarly:HSM.A	$\cdots$	$\cdots$			-0.15	0.00078
	BelongEarly:Nrt	$\cdots$	$\cdots$	0.08	0.00023		
	SelfEffEarly:CGPA	$\cdots$	$\cdots$	0.07	0.00118		
	SelfEffEarly:FGCS	$\cdots$	$\cdots$			0.19	0.00160
	R^2	0.51	0.00000	0.40	0.00000	0.40	0.00000

HSGPA is a far worse predictor of self-efficacy or belonging than the standardized test scores in the analysis above. For this reason, we chose to continue to include ACTM in this study for its predictive power.

Ultimately, 18 variables were chosen to be included in the following regression analyses. These variables are shown in Tables II and III.

2. Early-semester regressions

To continue to address RQ1, using the reduced set of 18 independent variables, multiple linear regression was performed with SE_E and BL_E as the dependent variables to understand how OPCs predict SE_E and BL_E . SE_E and BL_E were not included as independent variables. These results are presented in Table II, where standardized regression coefficients, β , are reported for each variable

as β_{SE_E} and β_{BL_E} . Significance thresholds were determined by a Bonferroni correction based on the 18 independent variables. A Bonferroni correction divides the significance threshold (the probability at which a result is considered significant), α , by the number of tests performed (the number of independent variables) to produce a corrected significance threshold, α_C . In this case, $\alpha = 0.05$ becomes $\alpha_C(0.05) = 0.0028$, $\alpha = 0.01$ becomes $\alpha_C(0.01) = 0.00056$, and $\alpha = 0.001$ becomes $\alpha_C(0.001) = 0.00006$ where the uncorrected α value is included in parentheses. The p values reported in Tables II and III have not been corrected. Only significant coefficients are reported.

For continuous independent variables, the regression coefficient is related to the correlation coefficient. Cohen's criteria for correlation coefficients define $r \geq 0.1$ as a small effect, $r \geq 0.3$ as a medium effect, and $r \geq 0.5$ as a

large effect [95]. For dichotomous independent variables, standardized regression coefficients measure the change in standard deviation units between two levels of the variable. The coefficient is related to Cohen's d , where $d \geq 0.2$ is a small effect, $d \geq 0.5$ is a medium effect, and $d \geq 0.8$ is a large effect [95]. Cohen's criteria are general rules of thumb; Supplemental Material [68] explores an alternative criterion for the test average variable producing cutoffs that are similar but somewhat lower than Cohen's criteria. As such, the reader is cautioned that Cohen's criteria may not be accurate for the variables in this study.

Seven variables, including the categorical HS Physics variable, were significant predictors of SE_E ; all were small effects. The strongest predictors were HS physics, gender, conscientiousness, openness, and neuroticism, while pretest and ACTM were less important. Any HS Physics variable in which the student received an A, regardless of whether the class was AP or if the AP Exam was passed, was associated with higher levels of SE_E ; not receiving an A in the course did not increase SE_E over the reference level of not having a high school physics class. Note that all variables representing levels of a categorical variable are retained if one level is significant. The model explained 25% of the variance in SE_E . The -0.27 standard deviation lower SE_E for women, a small effect, is consistent with other research showing that women have lower self-efficacy in some science classes than men [21,24,25,28,29,31,33].

BL_E models were similar, but with some differences. Seven variables were significant predictors of BL_E . Pretest scores and HS physics were significant, but ACTM was not. Additionally, only one level of the HS Physics variable was significant, HSP.NAP.A. This is consistent with the variable importance results, where HS Physics was substantially more predictive of SE_E than BL_E . All personality variables were significant predictors of BL_E . Furthermore, gender was not a significant predictor of BL_E , despite being one of the strongest predictors of SE_E . This model explained 17% of the variance in BL_E .

In summary, using a reduced set of variables, prior academic experiences relating to physics and general mathematics preparation, as well as students' gender and personality, were significant predictors of SE_E . Academic factors such as prior experiences in physics were less predictive of BL_E compared to SE_E , while more personality facets were important.

3. Combined models

To address RQ2, structural equation modeling was used to fit the combined model predicting both SE_E and BL_E from the OPC variables. The optimal models in Table II were used to predict SE_E and BL_E . The model had excellent model fit, with Comparative Fit Index(CFI) = 0.995, Tucker – Lewis Index(TLI) = 0.962, Root Mean Square Error of Approximation(RMSEA) = 0.031, and

Standardized Root Mean Square Residual(SRMR) = 0.006. The combined model left an unexplained covariance of 0.37 between SE_E and BL_E ; the total covariance (correlation) between the variables was 0.56. As such, while the OPC variables explain a substantial amount of the relation between SE_E and BL_E , the majority of the relation is explained by unobserved variables.

E. Midsemester analysis

To address RQ3 and RQ4, multiple linear regression analyses were performed on variables measured in the middle of the semester— SE_M , BL_M , and test average—to explore how self-efficacy and belonging evolve throughout a physics class. The average of the first two in-semester test scores (TestAve) was included in these models. Both exam scores were available to students before midsemester variables were measured. Midsemester regression analyses are shown in Table III and summarized in Fig. 3(a). The path models use a notation presenting the regression coefficient resulting from a regression using only the variable at the start of the edge to predict the variable at the end of the edge before the arrow ($\rightarrow$) and the full regression model implied by the path model (Table III) after the arrow. This allows the determination of how other variables in the model change the relationship between the two variables. Figure 3(b) shows the resulting models if the OPC variables are excluded from the analysis. Figure 3(b) is used to address RQ5 and to perform the mediation analysis, as the moderating variables in Fig. 3(a) make the mediation analysis impossible.

1. Test average analysis

To explore RQ3, linear regression coefficients for the significant predictors of TestAve from the reduced set of 18 variables, β_{TA} , are shown in Table III. Both SE_E and BL_E were included as independent variables in this model; however, neither was a significant predictor of TestAve when controlling for OPCs. All significant variables were academic measures, with small to medium effects. Personality was also not important in predicting TestAve, controlling for SE_E and BL_E . Gender was the only significant demographic variable, with a less-than-small effect size. As such, our hypothesis that SE_E and BL_E predict TestAve was incorrect. Instead, the relation of SE_E and BL_E to TestAve was fully explained by academic OPC variables.

2. Moderation analysis

RQ4 examines the relationship of test average to midsemester self-efficacy and belonging development and explores possible mediators and moderators. A brief review of mediation and moderation analysis is included in Supplemental Material [68]. For midsemester models, interactions between OPCs and SE_E , BL_E , SE_M , and TestAve were tested. Interactions represent moderating

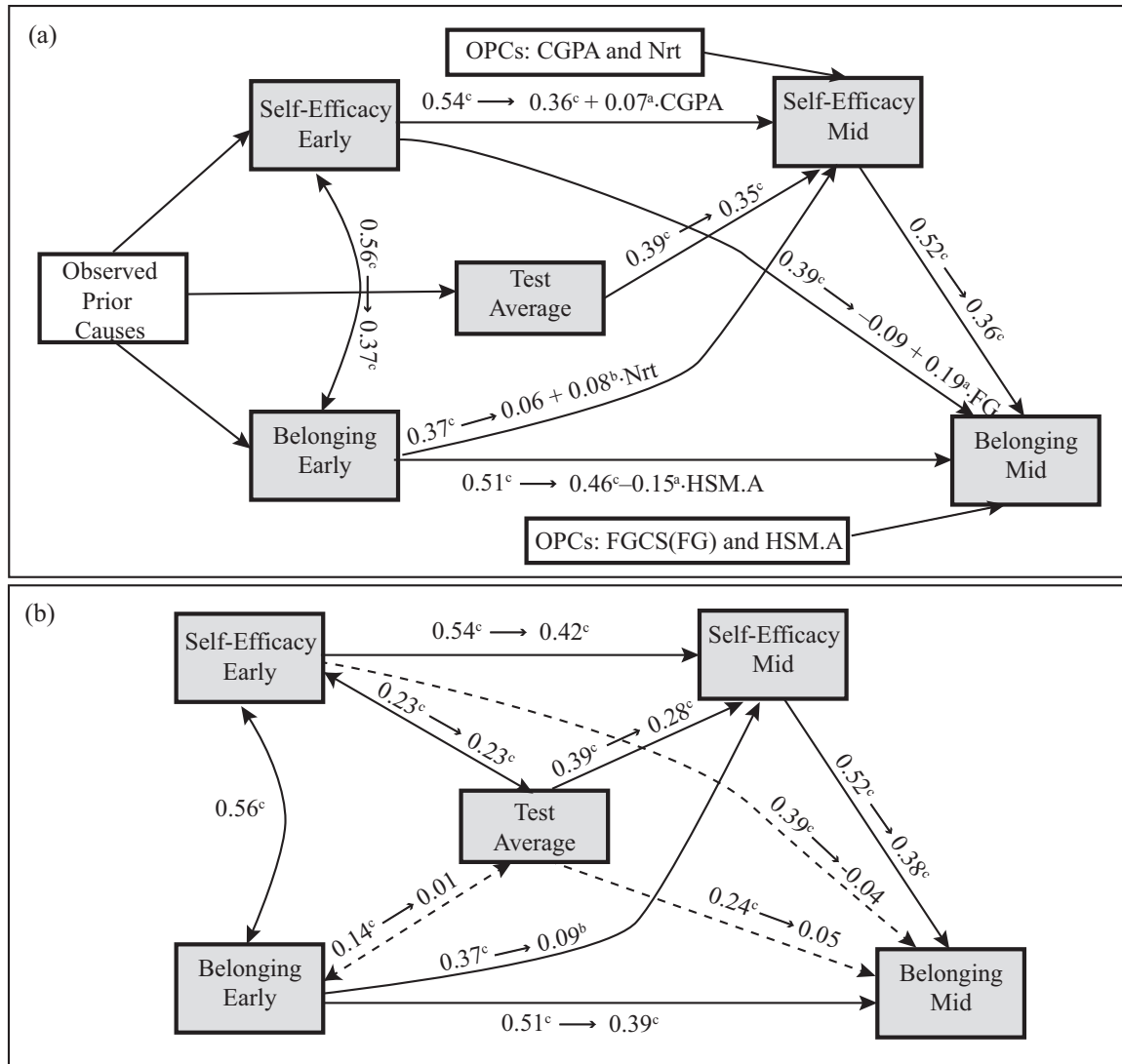


FIG. 3. Path model for midsemester models. (a) With OPCs and (b) without OPCs. Dashed lines represent nonsignificant direct effects. (b) Simplified model excluding OPCs and moderators. In all cases, the number to the left of the arrow represents the total effect, or the regression coefficient resulting from using only the variables at the head and tail of the edge in the model. The number to the right of the arrow is the regression coefficient from the full model. Nonsignificant edges are excluded from (a) for readability. The Bonferroni-corrected significance level of the coefficients is represented by superscripts: *a* represents $p < \alpha_c(0.05)$, *b* represents $p < \alpha_c(0.01)$, and *c* represents $p < \alpha_c(0.001)$.

effects where one variable changes the relation between two other variables and appear in regression models as products of independent variables. Ideally, fully saturated models containing all possible interactions should be tested first, followed by the removal of nonsignificant higher-level interactions until all remaining interaction terms are significant [96]. In our models, SE_E , BL_E , SE_M , and $TestAve$ were selected to test as potential moderators. Testing all combinations of these variables with all 18 OPC variables was impractical. Instead, each interaction term was individually tested in a regression containing the 18 OPC variables, SE_E , BL_E , and $TestAve$. SE_M interactions were also tested when BL_M was the dependent variable.

Significant interactions were determined using a Bonferroni correction. All significant interactions were identified individually, then added collectively to a single model, and nonsignificant interactions were removed. Remaining significant terms are reported in Table III. Note that, even if an independent variable was not significant on its own, it was reported if it was part of a significant interaction.

Interactions are presented in Fig. 3(a) by adding the product of the β coefficient multiplied by the moderating variable. For example, the interaction $SE_E \times CGPA$, where $\beta = 0.07$, is significant in the regression with SE_M as the dependent variable. As such, the edge between SE_E and

SE_M includes the term $0.07 \times CGPA$. OPC variables significant in midsemester models are shown in the rectangles labeled “OPCs” with edges to the variables they predict.

3. Midsemester self-efficacy regression

Unlike the early-semester models, very few variables were significant predictors of SE_M . SE_E and TestAve were the strongest predictors of SE_M ($\beta_{SE_M} = 0.36$ and 0.35 , respectively, medium effects). Other variables (CGPA and Nrt) were less predictive of SE_M . The relationship between SE_E and SE_M was moderated by CGPA; the negative relation of CGPA to SE_M was reduced for students with higher college GPAs. Reasons for the negative relation of CGPA to SE_M are unclear. Neuroticism moderated the relationship between BL_E and SE_M . While high levels of Nrt were related to lower SE_M scores, increases in BL_E predicted stronger increases in SE_M for students with higher-than-average Nrt scores. This suggests that having high levels of BL_E may insulate stress-prone students from midsemester decreases in self-efficacy. Overall, the model explained 40% of the variance in students' SE_M .

4. Midsemester belonging regression

The final regression in this study examined the relation of all regression variables to BL_M . SE_M and BL_E were the only variables that significantly predicted BL_M . Two interactions were also significant. FGCS moderated the relationship between SE_E and BL_M . For first-generation college students, higher levels of SE_E were associated with higher levels of BL_M , controlling for other factors. HSM.A moderated the relationship between BL_E and BL_M ; students who did not receive an A in their last high school math class had a stronger relationship between their BL_E and BL_M scores. The final model explained 40% of the variance in BL_M .

In summary, overall, few OPCs were significant in the midsemester SE_M and BL_M models. SE_E and TestAve explained most of the variance in SE_M ($R^2 = 0.36$), while BL_E and SE_M were the only significant predictors of BL_M . Some interactions were significant, but their coefficients were less than a small effect. The models accounted for 40% of the variance explained in both SE_M and BL_M .

5. Mediation

While path models express the complex theoretical assumptions of a research study, mediation provides a framework for quantitatively discussing those relations. The Supplemental Material [68] provides a general introduction to mediation analysis, which has its own language. Mediation partitions the “total effect” relating two variables into a “direct effect” and a set of “indirect effects.” The use of the term “effect” is traditional in mediation analysis and should not be interpreted as a causal claim.

The combination of moderators and paths through the OPC variables significant in midsemester models makes a full quantitative mediation analysis impossible. However, an approximate mediation analysis is possible using Fig. 3(b) by recognizing that the moderating effects are generally small and that the overall effect of the OPCs on early-semester variables is captured in the early-semester edges in Fig. 3(b). Figure 3(b) contains some correlational paths that must be included to carry out the mediation analysis. These pathways represent confounding effects that modify the causal interpretation of the overall regression coefficient. Pathways that traverse a correlation are called noncausal paths in this work.

In previous sections, structural equation modeling showed that OPC variables left an unexplained covariance of 0.37 between SE_E and BL_E . As such, some unobserved prior causes likely predict both SE_E and BL_E and explain the remaining covariance. The moderating and OPC variables in the midsemester models had small or less-than-small effects, causing little change in the path coefficients from Fig. 3(b). The overall effect of both unobserved and observed variables is captured by two early-semester edges in Fig. 3(b). First, the OPC variables themselves reduce the relations between SE_E and TestAve (TA) from 0.23 to 0. As such, the sum of the paths from SE_E to OPCs to TestAve must be 0.23, producing an effective correlation of 0.23 between SE_E and TestAve. As such, we have replaced the directed path in Fig. 3(b) from SE_E to TestAve with the correlation $SE_E \leftrightarrow TA$ (similarly for the path from $BL_E \leftrightarrow TA$). Furthermore, the sum of paths from SE_E to BL_E through both observed and unobserved prior causes must equal the overall correlation between these variables. Traversing the correlation between SE_E and BL_E should be understood to represent all paths through the prior causes. With this analysis, an approximate mediation analysis using Fig. 3(b) is possible.

In Fig. 1, we hypothesized that the relation of OPCs to TestAve would be partially mediated by both SE_E and BL_E . This hypothesis was incorrect; none of the relations of the OPCs to TestAve acted through the paths $OPC \rightarrow SE_E \rightarrow TA$ or $OPC \rightarrow BL_E \rightarrow TA$. The paths from SE_E or BL_E to TestAve were not significant; therefore, there was no mediation.

Likewise, TestAve did not mediate the relationship between SE_E and SE_M because the path from SE_E to TestAve was not significant; however, this was because the OPC variables fully explained the relation between SE_E and TestAve. With this observation, the reduction of the total effect of 0.54 to a direct effect of 0.42 can be understood as the result of the noncausal path $SE_E \leftrightarrow TA \rightarrow SE_M$ ($0.23 \times 0.28 = 0.06$) and the noncausal path $SE_E \leftrightarrow BL_E \rightarrow SE_M$ ($0.56 \times 0.09 = 0.05$). As such, all of the reduction of the total effect results from noncausal paths.

The relation of TestAve to BL_M was completely mediated, wherein the total effect of TestAve on BL_M (0.24) was reduced to a nonsignificant direct effect of 0.05. Multiplying the effects on the path $TA \rightarrow SE_M \rightarrow BL_M$ shows that $0.28 \times 0.38 = 0.11$ of the 0.24 total effect came from this path. The remaining effect results from noncausal paths through the correlation of SE_E and TestAve caused by the OPC variables, with 0.06 of the effect explained by the path $TA \leftrightarrow SE_E \leftrightarrow BL_E \rightarrow BL_M$ and 0.04 explained by the path $TA \leftrightarrow SE_E \rightarrow SE_M \rightarrow BL_M$. Other paths explained little of the total effect. The paths through directed edges explained the majority of the relation between TestAve and BL_M , but noncausal paths were also important.

The total effect of BL_E on BL_M was 0.51, which was reduced to a direct effect of 0.39 with the addition of indirect paths. As such, little of the effect of BL_E on BL_M acted through indirect paths. The directed path $BL_E \rightarrow SE_M \rightarrow BL_M$ accounted for $0.09 \times 0.38 = 0.03$ of this effect. Noncausal paths through the correlation between SE_E and BL_E account for the rest of the reduction in total effect. The path $BL_E \leftrightarrow SE_E \rightarrow SE_M \rightarrow BL_M$ was the most important path of the noncausal relations, explaining $0.56 \times 0.42 \times 0.38 = 0.09$ of the effect. The path $BL_E \leftrightarrow SE_E \rightarrow BL_M$ accounted for an additional $0.56 \times (-0.04) = -0.02$ of the relation. Other paths contributed less than 0.01 each to the indirect effect. As such, most of the reduction of the total effect results from the correlation of SE_E and BL_E .

The last mediating structure explores the effect of SE_E on BL_M . This relationship is completely mediated by indirect paths, reducing the total effect from 0.39 to a nonsignificant direct effect of -0.04 . The path $SE_E \rightarrow SE_M \rightarrow BL_M$ accounts for $0.42 \times 0.38 = 0.16$ of the total effect. Other paths through correlations account for the remaining reduction of the total effect. The path through $SE_E \leftrightarrow BL_E \rightarrow BL_M$ accounts for $0.56 \times 0.49 = 0.22$ of the total effect. The paths $SE_E \leftrightarrow TA \rightarrow SE_M \rightarrow BL_M$ and $SE_E \leftrightarrow BL_E \rightarrow SE_M \rightarrow BL_M$ each account for 0.02 of the total effect. Most of the relationship between SE_E and BL_M is explained by noncausal paths traversing a correlation.

In summary, the total effect of $BL_E \rightarrow BL_M$ and $SE_E \rightarrow SE_M$ was somewhat, but not greatly, reduced by the mediating effects of other variables in the model. Most of the changes in midsemester SE_M or BL_M were predicted by students' self-efficacy or belonging early in the class; academic feedback in the form of test averages changed these variables relatively little. Both directional and correlational pathways were important for explaining the reduction in the total effect. Should the causal structure implied by Fig. 3(a) be validated by future research, the above implies that a substantial portion of the effects of the model's variables act through noncausal correlational paths.

IV. DISCUSSION

This study investigated five research questions, which are discussed in the order proposed.

RQ1: What OPCs are most important in predicting early-semester physics self-efficacy and sense of belonging? SE_E was significantly predicted by high school preparation variables, prior physics experience, some personality traits, and gender ($R^2 = 0.25$). This supports prior research linking self-efficacy to academic success [15], personality [33,71], and gender [12,21]. Most predictors had small effect sizes, except for variables related to taking AP physics. Conscientiousness and neuroticism were significant predictors of self-efficacy, in agreement with prior research relating these variables [94]. The significance of openness is a new contribution to understanding personality and self-efficacy. This may relate to one's willingness to engage in challenging tasks, which can manifest in students' self-efficacy via their beliefs in their capacity to succeed. BL_E was significantly predicted by a similar set of variables ($R^2 = 0.17$), although gender and ACTM were not significant, and additional personality variables were. All significant predictors had small effect sizes.

RQ2: How are early-semester physics self-efficacy and sense of belonging related? Which OPCs are important in explaining their relationship? SE_E and BL_E were strongly correlated ($r = 0.56$, a large effect size). This correlation was much larger than the correlations between SE_E and BL_E and any other OPC. Variable importance analysis showed 24% of the covariance between SE_E and BL_E and was explained by the personality panel (Table III in Supplemental Material [68]), which was much larger than that explained by other groups of variables. Considering the extensive list of variables measuring cognitive skills in this study, including high school physics and math backgrounds, physics pretests, high school and college GPAs, math readiness, and standardized test scores, we argue that most of the relevant cognitive factors, like problem-solving, information processing, and critical thinking, are represented within the OPCs, although we cannot exclude the possibility of missing cognitive constructs. The covariance explained by the personality panel suggests that noncognitive factors are more important to understanding the relationship between self-efficacy and belonging than cognitive factors. Students' susceptibility to math anxiety, science anxiety, or stereotype threat [39–45] may be related to their personality, which in turn relates to self-efficacy and belonging. This potential importance of noncognitive variables is also supported by the 0.37 covariance unexplained by the effect of the OPC variables on both SE_E and BL_E , assuming the OPCs capture the relevant cognitive effects.

The analysis provides new and interesting insight into the relationship between personality and belonging. Conscientiousness, openness, and neuroticism significantly

predicted SE_E and BL_E ; however, the traits associated with social behaviors—agreeableness and extraversion—were uniquely significant to belonging. This informs the understanding of the relationship between self-efficacy and belonging. While self-efficacy is associated with traits that reflect on one's goal-attainment behaviors, like task management and regulating stress, belonging is closely tied to traits that describe students' tendencies to engage in social settings and community-building. Although self-efficacy and belonging are highly correlated, self-efficacy appears to act more as a cognitive measurement, while belonging presents as a more social variable. This adds to the theoretical discussion of which prior causes may be better predictors for each variable.

RQ3: How do self-efficacy and sense of belonging predict class performance, controlling for OPCs? While SE_E and BL_E both individually significantly predict TestAve without the OPCs, as shown by the total effects in Fig. 3(b), BL_E did not significantly predict TestAve when controlling for SE_E . This informs the interpretation of prior work that uses belonging to predict course grades without controlling for self-efficacy [46,97]. Exam scores are generally a very important component in calculating course grades. Furthermore, neither SE_E nor BL_E was a significant predictor of TestAve when controlling for OPCs. Every OPC that significantly predicted TestAve was an academic variable, except for gender (a small effect). As such, studies using self-efficacy or belonging to predict course grades should include a robust set of academic control variables. Failing to control for these factors may lead to conclusions that overstate the relationship between self-efficacy or belonging and academic performance. It is important to note that while OPCs fully explain the relation between self-efficacy and belonging on course performance, this is likely because the academic effects of these two variables have already modified prior academic performance.

RQ4: How does academic performance in a physics class predict the evolution of self-efficacy and belonging during the class? How do mediating or moderating variables predict the evolution? TestAve significantly predicted SE_M , controlling for OPCs, a medium effect. As hypothesized, TestAve did not significantly predict BL_M , controlling for OPCs. Instead, the relation of TestAve to BL_M resulted from its association with SE_M . This supports a model where academic feedback through mastery experiences, such as physics grades, is more closely related to self-efficacy development than to making students feel like they belong in physics classrooms.

Few variables were significant predictors of SE_M controlling for SE_E : TestAve, CGPA, and Nrt. CGPA moderated the relation of SE_E to SE_M . The relation of CGPA to SE_M was significant, negative, and a small effect, but this relation was largely eliminated for students with higher SE_E . This suggests that CGPA applies a negative correction

to the SE_M one would predict from SE_E and TestAve alone, but only for lower-achieving students. Although the relation of BL_E to SE_M was not significant, a significant moderation with Nrt was measured. Students with higher levels of Nrt were more likely to have higher SE_M if they also had higher BL_E . This moderating effect suggests that developing social support systems early on in physics classrooms may shield students who are more susceptible to stress from its negative effects on their self-efficacy as they progress throughout the semester and reaffirms the significance of this personality trait in relation to self-efficacy research [34,93]. This result is also interesting when we interpret Nrt as a measure of students' physiological state, i.e., their ability to cope with stressful situations, which is one of Bandura's sources of self-efficacy [16].

No OPCs significantly predicted BL_M controlling for early-semester variables. BL_M was only significantly predicted by SE_M and BL_E . As hypothesized, TestAve was not a significant predictor of BL_M , controlling for SE_M . However, the relation of SE_E and BL_E to BL_M was moderated by both FGCS and HS Math grades. The relation of SE_E and BL_M was significant and positive for FGCS students but not significant for continuing generation students. Although FGCS students are typically less academically prepared for college courses [73,98], their confidence in performing well in the class is more important for BL_M than for continuing generation students. There was also a small, nonsignificant, positive relation between earning an A in the student's last high school mathematics class and BL_M . This effect was moderated and became negative for students with higher BL_E . Students who did not earn an A in their last high school math class were more sensitive to changes in their sense of belonging, suggesting that prior math experiences are important to belonging development in college physics courses.

Mediation analysis revealed that only a small portion of the total effect of SE_E on SE_M or BL_E on BL_M was explained by indirect paths. Using Fig. 3(b), much of the reduction of the total effect of SE_E on SE_M resulted from the correlation of SE_E and TestAve produced by the OPC variables. The mediating path through SE_M explained the majority of the relation between TestAve and BL_M . However, correlational noncausal paths were also important; most of the reduction of the total effect of BL_E on BL_M resulted from noncausal paths. As such, when examining the complex relations between these variables, it is crucial to include correlations between variables.

RQ5: How would the conclusions of this study change if some variables were not available? Weissman [55,56] warned of the effect of missing variable bias in PER studies. The current work can provide important context for the potential effects of missing variables. Consider the change of the early-semester models if some combination of OPCs, SE_E , or BL_E were unavailable. In the absence of OPCs and SE_E , one might conclude that BL_E directly predicted

TestAve, indicated by the total effect in Fig. 3(b). This conclusion would suggest that an early-semester belonging intervention could improve academic achievement. Once SE_E is added to the model, the predictive relationship between BL_E and TestAve disappears; all of the relation of belonging to TestAve acts through the correlation between SE_E and BL_E . One would then conclude that the significant prediction of TestAve comes from SE_E instead. Similarly, the addition of OPCs to the model [Fig. 3(a)], particularly academic variables, also renders this relationship nonsignificant. Omitting prior academic variables would lead to very different conclusions about the relationships between SE_E , BL_E , and TestAve.

OPCs and moderators only minimally altered the results of midsemester regressions. Direct effects on SE_M and BL_M were reduced on average by 0.05 standard deviations. Additionally, controlling for SE_E and BL_E removed most of the association of OPCs with the midsemester variables. Overall, the conclusions of the midsemester analyses would change very little if the OPCs were not available, which is quite different from the conclusions of the early-semester models.

A. Hypotheses revisited

This work began by making a set of causal hypotheses, captured in the path model in Fig. 1. Some of these hypotheses were supported (but not proven) by the analysis, while some were not. The hypothesis that early-semester variables (SE_E and BL_E) control for associations between OPCs and midsemester variables was supported; few OPCs were significant predictors of SE_M or BL_M when controlling for SE_E and BL_E . Those that were significant had fairly small coefficients. The hypothesis that SE_E and BL_E mediated the relation of OPCs to TestAve was incorrect, as was the hypothesis that TestAve mediated the relation of SE_E to SE_M . Because TestAve was not significantly related to BL_M when controlling for SE_M , our hypothesis that TestAve is primarily associated with BL_M through its effect on SE_M was supported.

V. LIMITATIONS

It is important to remember that the causal hypotheses of the previous section are the theoretical hypotheses of the research team. No correlational study can prove these hypotheses, but they can provide support.

Although the set of variables used in this study was extensive, it was not complete. Additional noncognitive variables theoretically related to belonging or self-efficacy could be important, providing insight into the social contributions to self-efficacy or belonging development. For example, Zeldin and Pajares suggest that women rely more on sources of vicarious experiences and social persuasion than mastery experiences when developing their physics self-efficacy [28,29], particularly in STEM

fields where women are underrepresented. In the studies above, interviews with predominantly Caucasian men and women reflected on the participants' choices to pursue careers in STEM fields with less representation of women. While men frequently attributed their confidence in pursuing a STEM career to their cognitive skills and past achievements, women overwhelmingly reported that encouragement from family members, teachers, and (less frequently) peers and supervisors helped them build a foundation of resilience when facing challenges along their academic trajectories. However, Zeldin and Pajares note that while their findings support Bandura's emphasis on mastery experiences [16], Bandura does not ascribe motivational beliefs by gender [99].

This study was conducted at a single institution. Considering the importance of high school physics and math experiences, this work should be replicated at institutions with a range of incoming students' levels of preparation and demographic composition to determine if the results are generalizable.

Although the sample in this study was large ($n = 1,373$), it only represents about 42% of students who were enrolled in the course over the period studied. In survey methods research, nonresponse rates offer interesting insight into the noncognitive skills of survey takers. For example, some studies have shown positive correlations between survey response rate and academic attainment [100,101], although this result diminishes when controlling for cognitive abilities [101]. This was also observed in the current study. Item response rates may be related to student-task orientation, motivation, or even conscientiousness, which can be predictive of students' self-efficacy and belonging. While the demographic composition of the sample is somewhat different than the general population of the class, the primary difference between the class and sample population is that students in the sample were substantially higher performing. This may mean that the results of this study do not fully generalize to weaker-performing or less engaged students.

VI. IMPLICATIONS

Studies that fail to account for the relationship between prior experiences and in-semester academic performance may draw incorrect conclusions about the relationships between self-efficacy, belonging, and other variables. Mediation analysis also illustrates the importance of utilizing a complete set of variables. For some variables, a substantial part of the mediation acted through paths containing correlations. If correlating variables were not present, one might overstate the effects of available variables.

This work used path models to encapsulate the causal hypotheses of the research team. These models allow more sophisticated conclusions when applying the causal reasoning framework pioneered by Pearl [61,62]. Controlling

for SE_E , BL_E , and TestAve largely eliminated the effect of OPCs on the models predicting SE_M and BL_M . As such, Pearl's framework would suggest that conditioning on these variables will make SE_M and BL_M conditionally independent from OPCs. Furthermore, Pearl's framework allows us to estimate the effect of intervening on a variable, perhaps SE_M , if its parent variables are observed.

Self-efficacy and belonging did not significantly predict course performance when controlling for OPCs, suggesting that interventions to change early-semester physics self-efficacy or belonging may not have a direct effect on course grades because of confounding prior experiences. However, they may still be useful in increasing student retention [14,15]. As outlined by Tinto, academic and social integration, as well as self-efficacy and belonging, are important factors that ultimately relate to persistence in physics classes and college overall. Although they are highly correlated, treating self-efficacy as an academic variable and belonging as a social variable may set researchers on the right path when developing interventions to improve these measures in the physics classroom. By understanding how self-efficacy and belonging develop over time, we can strive to improve students' beliefs in their capacity to succeed and bolster their surrounding social supports, ultimately aiming to increase physics student retention.

VII. CONCLUSIONS AND FUTURE WORK

Self-efficacy and sense of belonging are highly correlated variables ($r = 0.56$), much more correlated than with any other variable in the study ($r = 0.30$ at maximum). Personality variables were the most important for explaining the relation of self-efficacy and belonging and were substantially more important than prior academic achievement. Controlling for high school and college academic factors completely explained the relation of early-semester self-efficacy and belonging to test average. Midsemester self-efficacy and belonging were predicted by test average; however, relatively little of the total effect of either variable was explained by controlling for test average. The addition of cognitive and noncognitive control variables substantially changed early-semester models, indicating that missing variables could greatly change the

conclusions of studies using early-semester variables. However, cognitive and noncognitive control variables had little effect on midsemester models when controlling for early-semester self-efficacy and belonging.

The results in this paper supported the role of mastery experiences in the development of self-efficacy, examined a model where changes in self-efficacy due to mastery experiences were related to changes in belonging, and showed that OPC variables completely explained the relation of early-semester self-efficacy and belonging to test average. The paper also demonstrated the importance of noncognitive variables—particularly personality—in explaining the relation between self-efficacy and belonging. Additional research is needed to fully understand the relationship between these variables. While the interactions between personality and self-efficacy are well explored, further work could be done to relate students' sense of belonging to different personality traits.

Early-semester variables retained substantial amounts of unexplained variance; as such, it is important to explore alternative measures of prior student experience and how they may shed light on the relationship between self-efficacy and belonging in physics classrooms. Robust scales that measure other theorized sources of self-efficacy and belonging, such as the quality of student interactions and dimensions of social persuasion, should be developed. Likewise, studies with time resolution greater than two time points are needed to better establish the magnitude and direction of changes in self-efficacy and belonging across multiple classes and stages of a student's academic career.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under Grants No. ECR-1561517, No. DUE-1833694, and No. HRD-1834569.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain sensitive personal information. The data are available from the authors upon reasonable request.

-
- [1] S. White, R. Y. Chu, and J. Oman, *Has the Number of Students Taking Physics in US High Schools Declined?* (American Institute of Physics, College Park, MD, 2025).
 - [2] President's Council of Advisors on Science and Technology, Report to the president. Engage to excel: Producing one million additional college graduates with

degrees in science, technology, engineering, and mathematics, Executive Office of the President: Washington, DC, 2012.

- [3] T. Hodapp, The back page: The economics of education: Closing undergraduate physics programs, APS News **20**, 8 (2011).

- [4] J. Stewart, J. Hansen, and E. Burkholder, Visualizing and predicting the path to an undergraduate physics degree at two different institutions, *Phys. Rev. Phys. Educ. Res.* **18**, 020117 (2022).
- [5] J.D. Hansen, Methods for analyzing physics student retention and physics curricula, Ph.D. thesis, West Virginia University, 2023.
- [6] Z. Topdemir, Pedagogical effects on student learning, attitude, and retention of physics majors, Ph.D. thesis, Georgia State University, 2020.
- [7] D. Maronde, The Effectiveness of Teaching Methods Designed to Improve Student Engagement and Retention of Physics Subject Matter for Both Science and Non-Science Majors, Ph.D. thesis, University of Central Florida, 2011.
- [8] J.P. Zwolak, R. Dou, E. A. Williams, and E. Brewe, Students' network integration as a predictor of persistence in introductory physics courses, *Phys. Rev. Phys. Educ. Res.* **13**, 010113 (2017).
- [9] J.P. Zwolak, M. Zwolak, and E. Brewe, Educational commitment and social networking: The power of informal networks, *Phys. Rev. Phys. Educ. Res.* **14**, 010131 (2018).
- [10] V. Tinto, Classrooms as communities: Exploring the educational character of student persistence, *J. High. Educ.* **68**, 599 (1997).
- [11] V. Tinto, Through the eyes of students, *J. Coll. Stud. Retent-Res. Theor. Pract.* **19**, 254 (2017).
- [12] A. Bandura, Self-efficacy: Toward a unifying theory of behavioral change, *Psychol. Rev.* **84**, 191 (1977).
- [13] A. Bandura, *Self-Efficacy* (Academic Press, New York, 1994), Vol. 4, p. 71.
- [14] 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).
- [15] 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).
- [16] A. Bandura, *Self-Efficacy: The Exercise of Control* (Freeman, New York, NY, 1997), Vol. 11.
- [17] 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).
- [18] D. Maldonado and J. Stewart, Adding self-regulated learning instruction to an introductory physics class, in *PER Conference Proceedings, Sacramento, CA, PER Conference* (American Institute of Physics, College Park, MD, 2023), p. 199.
- [19] D. Maldonado and J. Stewart, Comparing the evolution of self-regulated learning behaviors to academic performance, personality, and self-efficacy in the introductory physics classroom, in *PER Conference Proceedings, Boston, MA, PER Conference* (American Institute of Physics, College Park, MD, 2024), p. 256.
- [20] F. Pajares, Gender and perceived self-efficacy in self-regulated learning, *Theory Pract.* **41**, 116 (2002).
- [21] J. M. Nissen, Gender differences in self-efficacy states in high school physics, *Phys. Rev. Phys. Educ. Res.* **15**, 013102 (2019).
- [22] A. M. Porter, *How Women Persist in Undergraduate Physics* (American Institute of Physics, College Park, MD, 2025).
- [23] P. Mulvey and S. Nicholson, *Percent of Bachelor's, Master's and Doctorates in Physics Earned by Women, Classes 1991 Through 2023* (American Institute of Physics, College Park, MD, 2025).
- [24] J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020105 (2016).
- [25] R. Henderson, V. Sawtelle, and J. M. Nissen, Gender & self-efficacy: A call to physics educators, *Phys. Teach.* **58**, 345 (2020).
- [26] 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).
- [27] S. L. Anderson and N. E. Betz, Sources of social self-efficacy expectations: Their measurement and relation to career development, *J. Vocat. Behav.* **58**, 98 (2001).
- [28] A. L. Zeldin and F. Pajares, Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers, *Am. Educ. Res. J.* **37**, 215 (2000).
- [29] A. L. Zeldin, S. L. Britner, and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *J. Res. Sci. Teach.* **45**, 1036 (2008).
- [30] V. Sawtelle, E. Brewe, and L. H. Kramer, Positive impacts of modeling instruction on self-efficacy, *AIP Conf. Proc.* **1289**, 289 (2010).
- [31] T. Espinosa, K. Miller, I. Araujo, and E. Mazur, Reducing the gender gap in students' physics self-efficacy in a team-and project-based introductory physics class, *Phys. Rev. Phys. Educ. Res.* **15**, 010132 (2019).
- [32] Y. Li and C. Singh, Impact of perceived recognition by physics instructors on women's self-efficacy and interest, *Phys. Rev. Phys. Educ. Res.* **19**, 020125 (2023).
- [33] R. Henderson, D. Hewagallage, J. Follmer, L. Michaluk, J. Deshler, E. Fuller, and J. Stewart, Mediating role of personality in the relation of gender to self-efficacy in physics and mathematics, *Phys. Rev. Phys. Educ. Res.* **18**, 010143 (2022).
- [34] N. Schmitt, The interaction of neuroticism and gender and its impact on self-efficacy and performance, *Hum. Perform.* **21**, 49 (2007).
- [35] 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).
- [36] S. Cwik and C. Singh, Gender differences in students' self-efficacy in introductory physics courses in which women outnumber men predict their grade, *Phys. Rev. Phys. Educ. Res.* **18**, 020142 (2022).
- [37] R. F. Baumeister and M. R. Leary, The need to belong: Desire for interpersonal attachments as a fundamental human motivation, *Interpers. Dev.*, 57 (2017).

- [38] K. L. Lewis, J. G. Stout, S. J. Pollock, N. D. Finkelstein, and T. A. Ito, Fitting in or opting out: A review of key social-psychological factors influencing a sense of belonging for women in physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020110 (2016).
- [39] X. Ma, A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics, *J. Res. Math. Educ.* **30**, 520 (1999).
- [40] 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).
- [41] J. V. Mallow and S. L. Greenburg, Science anxiety: Causes and remedies, *J. Coll. Sci. Teach.* **11**, 356 (1982).
- [42] 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).
- [43] 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).
- [44] J. R. Shapiro and A. M. Williams, The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields, *Sex Roles* **66**, 175 (2012).
- [45] J. G. Stout, T. A. Ito, N. D. Finkelstein, and S. J. Pollock, How a gender gap in belonging contributes to the gender gap in physics participation, *AIP Conf. Proc.* **1513**, 402 (2013).
- [46] Y. Li and C. Singh, Sense of belonging is an important predictor of introductory physics students' academic performance, *Phys. Rev. Phys. Educ. Res.* **19**, 020137 (2023).
- [47] K. Rainey, M. Dancy, R. Mickelson, E. Stearns, and S. Moller, Race and gender differences in how sense of belonging influences decisions to major in STEM, *Int. J. STEM Educ.* **5**, 1 (2018).
- [48] G. M. Walton, C. Logel, J. M. Peach, S. J. Spencer, and M. P. Zanna, Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering, *J. Educ. Psychol.* **107**, 468 (2015).
- [49] C. Good, A. Rattan, and C. S. Dweck, Why do women opt out? Sense of belonging and women's representation in mathematics, *J. Pers. Soc. Psychol.* **102**, 700 (2012).
- [50] M. Franklin and E. Brewe, What correlates with persistence of women in physics?, *Phys. Rev. Phys. Educ. Res.* **21**, 010115 (2025).
- [51] L. Espinosa, Pipelines and pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence, *Harv. Educ. Rev.* **81**, 209 (2011).
- [52] K. Rosa and F. M. Mensah, Educational pathways of Black women physicists: Stories of experiencing and overcoming obstacles in life, *Phys. Rev. Phys. Educ. Res.* **12**, 020113 (2016).
- [53] D. K. Kehlbeck, K. Piatek-Jimenez, and C. Medina Medina, Undergraduate physics students' experiences: Exploring the impact of underrepresented identities and intersectionality, *Phys. Rev. Phys. Educ. Res.* **20**, 020120 (2024).
- [54] R. A. Miller and M. Downey, Examining the STEM climate for queer students with disabilities, *J. Postsecond. Educ. Di.* **33**, 169 (2020).
- [55] M. B. Weissman, Methods of causal inference in physics education research, *arXiv:2110.04266*.
- [56] M. B. Weissman, Policy recommendations from causal inference in physics education research, *Phys. Rev. Phys. Educ. Res.* **17**, 020118 (2021).
- [57] T. Honicke and J. Broadbent, The influence of academic self-efficacy on academic performance: A systematic review, *Educ. Res. Rev.* **17**, 63 (2016).
- [58] E. L. Usher and F. Pajares, Sources of self-efficacy in school: Critical review of the literature and future directions, *Rev. Educ. Res.* **78**, 751 (2008).
- [59] S. Zumbrunn, C. McKim, E. Buhs, and L. R. Hawley, Support, belonging, motivation, and engagement in the college classroom: A mixed method study, *Instr. Sci.* **42**, 661 (2014).
- [60] T. M. Freeman, L. H. Anderman, and J. M. Jensen, Sense of belonging in college freshmen at the classroom and campus levels, *J. Exp. Educ.* **75**, 203 (2007).
- [61] J. Pearl, *Causality* (Cambridge University Press, Cambridge, United Kingdom, 2009).
- [62] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer* (John Wiley & Sons, Hoboken, NJ, 2016).
- [63] V. Adlakha and E. Kuo, Critical issues in statistical causal inference for observational physics education research, *Phys. Rev. Phys. Educ. Res.* **19**, 020160 (2023).
- [64] National Center for Education Statistics, <https://nces.ed.gov/collegenavigator>. Accessed 2/10/2025.
- [65] E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1997).
- [66] R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- [67] R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, Comparing the force and motion conceptual evaluation and the force concept inventory, *Phys. Rev. Phys. Educ. Res.* **5**, 010105 (2009).
- [68] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/lkdr-cz49> which includes Ref. [69], for descriptive statistics, correlation analysis, variable selection, and interaction plots.
- [69] 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).
- [70] J. Stewart, G. L. Cochran, R. Henderson, C. Zabriskie, S. DeVore, P. Miller, G. Stewart, and L. Michaluk, Mediation effect of prior preparation on performance differences of students underrepresented in physics, *Phys. Rev. Phys. Educ. Res.* **17**, 010107 (2021).
- [71] J. Stewart and D. Hewagallage, The relation of personality, gender, and achievement in science classes, in *AERA Conference Proceedings, Toronto, Ontario* (AERA, New York, NY, 2019).

- [72] D. Hewagallage, E. Christman, and J. Stewart, Examining the relation of high school preparation and college achievement to conceptual understanding, *Phys. Rev. Phys. Educ. Res.* **18**, 010149 (2022).
- [73] R. Henderson, C. Zabriskie, and J. Stewart, Rural and first generation performance differences on the Force and Motion Conceptual Evaluation, in *PER Conference Proceedings, Washington, DC* (AIP, College Park, MD, 2018).
- [74] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [75] P. R. Pintrich, D. A. F. Smith, T. Garcia, and W. J. McKeachie, Reliability and redictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ), *Educ. Psychol. Meas.* **53**, 801 (1993).
- [76] T. G. Duncan and W. J. McKeachie, The making of the Motivated Strategies for Learning Questionnaire, *Educ. Psych.* **40**, 117 (2005).
- [77] 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. Tech.* **29**, 758 (2020).
- [78] 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).
- [79] C. Zabriskie, J. Yang, S. DeVore, and J. Stewart, Using machine learning to predict physics course outcomes, *Phys. Rev. Phys. Educ. Res.* **15**, 020120 (2019).
- [80] P. A. Westrick, J. P. Marini, L. Young, H. Ng, D. Shmueli, and E. J. Shaw, *Validity of the SAT for Predicting First-Year Grades and Retention to the Second Year* (College Board, New York, NY, 2019).
- [81] P. A. Westrick, H. Le, S. B. Robbins, J. M. R. Radunzel, and F. L. Schmidt, College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES, *Educ. Assess.* **20**, 23 (2015).
- [82] J. Allen, *Updating the ACT College Readiness Benchmarks*, ACT Research Report Series 2013 (6) (ACT, Inc., Iowa City, IA, 2013).
- [83] W. J. Camara and G. Echternacht, The SAT I and high school grades: Utility in predicting success in college, *Res. Not.* **10** (2000).
- [84] B. M. Galla, E. P. Shulman, B. D. Plummer, M. Gardner, S. J. Hutt, J. P. Goyer, S. K. D'Mello, A. S. Finn, and A. L. Duckworth, Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability, *Am. Educ. Res. J.* **56**, 2077 (2019).
- [85] W. G. Bowen, W. M. Chingos, and M. S. McPherson, Test scores and high school grades as predictors, in *Crossing the Finish Line* (Princeton University Press, Princeton, NJ, 2009), p. 112.
- [86] S. L. Geiser and R. Studley, UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California, *Educ. Assess.* **8**, 1 (2002).
- [87] S. Geiser and M. V. Santelices, *Validity of High-School Grades in Predicting Student Success beyond the Freshman Year: High-School Record vs Standardized Tests as Indicators of Four-Year College Outcomes*, Research & Occasional Paper Series: CSHE. 6.07 (Center for Studies in Higher Education, University of California - Berkeley, Berkeley, CA, 2007).
- [88] L. R. Goldberg, The development of markers for the Big-Five factor structure, *Psychol. Assess.* **4**, 26 (1992).
- [89] 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, Berkeley, CA, 1991).
- [90] 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* (The Guilford Press, New York, NY, 2008), p. 114.
- [91] 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).
- [92] A. E. Poropat, A meta-analysis of the five-factor model of personality and academic performance, *Psychol. Bull.* **135**, 322 (2009).
- [93] T. A. Judge and R. Ilies, Relationship of personality to performance motivation: A meta-analytic review, *J. Appl. Psychol.* **87**, 797 (2002).
- [94] A. D. Stajkovic, A. Bandura, E. A. Locke, D. Lee, and K. Sergeant, Test of three conceptual models of influence of the big five personality traits and self-efficacy on academic performance: A meta-analytic path-analysis, *Pers. Individ. Differ.* **120**, 238 (2018).
- [95] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Academic Press, New York, NY, 1977).
- [96] B. T. West, K. B. Welch, and A. T. Galecki, *Linear Mixed Models: A Practical Guide Using Statistical Software* (Chapman and Hall - CRC Press, Boca Raton, FL, USA, 2022).
- [97] S. Cwik and C. Singh, Students' sense of belonging in introductory physics course for bioscience majors predicts their grade, *Phys. Rev. Phys. Educ. Res.* **18**, 010139 (2022).
- [98] M. C. Atherton, Academic preparedness of first-generation college students: Different perspectives, *J. Coll. Stud. Dev.* **55**, 824 (2014).
- [99] K. Bussey and A. Bandura, Social cognitive theory of gender development and differentiation, *Psychol. Rev.* **106**, 676 (1999).
- [100] T. W. Smith, Educated don't knows: An analysis of the relationship between education and item nonresponse, *Polit. Methodol.* **8**, 47 (1982).
- [101] C. Hitt, J. Trivitt, and A. Cheng, When you say nothing at all: The predictive power of student effort on surveys, *Econ. Educ. Rev.* **52**, 105 (2016).