Bioscience students' internalized mindsets predict grades and reveal gender inequities in physics courses

Alysa Malespina,[1,*] Christian D. Schunn,[2] and Chandralekha Singh[1]

[1]Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania USA, 15260
[2]Learning, Research, and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania USA, 15260

(Received 29 November 2022; accepted 27 January 2023; published 28 September 2023)

Students' motivational beliefs, such as disciplinary intelligence mindsets, can influence their physics performance and persistence. Intelligence mindset beliefs have long been argued to fall along of continuum between fixed and growth mindsets. Those with fixed physics mindsets believe that ability in physics is innate and unchangable, while those with growth mindset believe that ability in physics can be developed with effort. More recent research with physical science and engineering majors suggests these are somewhat separable beliefs, with some students believing aspects of both fixed and growth mindsets, and that students can hold different beliefs about typical other students versus beliefs about themselves (e.g., others could improve through effort but they themselves could not). In this study, 419 students in physics 1 for students pursuing bioscience majors took pre- and post-physics mindset surveys. We investigated whether the physics mindset views of students pursuing bioscience or health-related majors were separable into more nuanced dimensions, if the means and distribution of these views varied by gender or sex and over time, and if any of these views predicted course grade. Replicating prior findings with physical science and engineering majors, we found that intelligence mindsets can be divided into four separable but correlated constructs: my ability, my growth, others' ability, and others' growth. Further, in this bioscience or health-related majors group, the "ability" beliefs grew stronger and the "growth" beliefs became weaker over time. These shifts were particularly strong for women. The changes in beliefs were also stronger for "my" beliefs than "others" beliefs for both men and women Unfortunately, my ability and my growth scores were also the strongest predictors of course grades above and beyond academic preparation differences as assessed by high school GPA and SAT/ACT math scores. These findings have implications for eliminating classroom inequities, such as through the development of new mindset interventions.

DOI: 10.1103/PhysRevPhysEducRes.19.020135

I. INTRODUCTION AND THEORETICAL FRAMEWORK

For decades, physics departments have struggled to recruit and retain women [1–3] and generally many women in the broader science, technology, engineering, and mathematics (STEM) workforce have a negative view of physics [4]. In response, researchers have dedicated effort to improving gender equity and diversity (for example, see [5–11]) of physics departments and classrooms. Some of their research has focused on gender differences in motivational beliefs that arise from negative messages in prior and current classrooms as well as broader society. For example, researchers have found that gender differences in physics-specific motivational beliefs (such as physics self-efficacy, perceived recognition from instructors, and intelligence mindset) may account for some of the differences in physics performance and persistence between women and men [12–19]. Other studies also posit that societal stereotypes and biases about who belongs in and can excel in physics also may explain some of these gender differences [20–23], either via messages from media, family, and friends or as the cause of negative messages voiced by instructors, teaching assistants (TAs), and classmates [17–19].

Much of the research about motivational beliefs, performance, and equity in introductory physics courses has focused on courses for students pursuing engineering and physical science majors, rather than for bioscience and health-related majors. These courses in the U.S. often differ in gender or sex makeup: most students in courses for engineering and physical science students are men, but most students in courses for bioscience students are women, similar to the higher participation rate of male students in calculus-based versus algebra-based AP physics

*Corresponding author: alm417@pitt.edu

courses [24]. Some research suggests being a numerical minority in a classroom has negative motivational consequences [25]. On the other hand, negative prior experiences with physics may continue to produce negative attitudes towards physics even when one is not a numerical minority. Other prior research has found that even in physics courses in which women are not underrepresented, men tend to have higher grades and physics-specific motivational beliefs in physics courses than women [5–9,12–19]. For example, women tend to have lower physics self-efficacy (one's belief in their capability to succeed at an activity or subject [26]) than men with the same grades in courses for engineering and physical science students as well as courses for bioscience students [15,27]. However, one motivational belief that has not been studied in the context of bioscience majors taking introductory physics is physics intelligence mindset, which may be particularly skewed towards fixed mindsets among students choosing majors and career paths that involve relatively little physics.

More broadly, intelligence mindset describes a person's views about the nature of intelligence, and was originally conceptualized on a spectrum [28]. On one end of this spectrum is a growth mindset, in which intelligence is believed to be cultivated with effort and can be developed over time [28]. On the other end is a fixed mindset, in which intelligence is believed to be innate and unchangeable [28]. The study of domain-specific intelligence mindset has gained popularity in recent years [12–14]. This is because the mindset for a discipline can be different from a general intelligence mindset and because domain-specific mindsets tend to be more predictive of student performance in that discipline [12,13,27].

In prior work, we developed a new, physics-specific tool to measure intelligence mindsets [12,13], which has been previously used to investigate mindset beliefs of students in physics courses aimed at engineering and physical science majors, but has not yet been used for students in physics courses for bioscience and health-related majors. In this study, we aim to investigate the nature of physics-specific mindsets for this latter group, as well as whether physics intelligence mindsets change from the beginning to the end of the course, differ by gender, or can predict learning outcomes.

A. Intelligence mindset theory

Intelligence mindset theory posits that there are two broad beliefs about intelligence and how it is formed: growth mindsets and fixed mindsets. A growth mindset is one in which intelligence is viewed as something that can be cultivated with effort, like a muscle, whereas a fixed mindset is one in which intelligence is thought to be innate and unchangeable [28]. Mindset beliefs have implications for how people engage with challenges faced while learning. Students with fixed mindsets tend to disengage from or avoid difficult tasks, and tend to view struggle as a sign that they are not smart enough to succeed, rather than a normal part of learning [28–30]. On the other hand, students with growth mindsets tend to welcome challenges and view them as an opportunity to learn and improve their abilities [30,31].

Intelligence mindsets are a useful focus for educational research because of their important role in influencing student learning behaviors but also because relatively brief interventions have been found to successfully change student mindsets for months and even years later. Focusing on their role in student learning behaviors, growth mindsets have been linked to positive learning outcomes even after controlling for prior academic achievement because they can increase students' engagement, propensity to attempt challenging problems, and persistence [28,32–35]. Further, intelligence mindsets often vary by gender and race or ethnicity, and these relationships have been argued to be an important pathway by which inequity of learning outcomes and participation in STEM occur [29,36]. Strong growth mindset beliefs can lead to a greater sense of belonging for both women and students from other underrepresented groups [37].

Turning to interventions focused on student intelligence mindsets, a number of brief interventions have been tested in middle school, high school, and university contexts. Several of these interventions have successfully changed students' reported mindsets [5,35,38,39] and improved students' learning outcomes [34,35,40]. These interventions have tended to be especially effective for students at high risk of failing a class [38,41].

Despite some well-publicized successes with some interventions, a recent meta-analysis by Sisk *et al.* [42] revealed that the effectiveness of mindset interventions varies significantly, with only 12% of included interventions significantly improving academic achievement. One possible factor that could determine the effectiveness of a mindset intervention is the demographic groups a student belongs to. An intervention may be more effective for women or low-income students than for men or high-income students [43]. Indeed it is important to examine which groups experience low growth mindsets or high ability mindsets to understand which groups are likely to be helped by a mindset intervention. However, Sisk *et al.* also raised concerns about mindset's ability to predict learning outcomes in particular contexts. We argue (see below) that general intelligence mindsets may not be as important as discipline-specific mindsets for participation and learning outcomes within disciplines, especially in disciplines like physics for which there are especially strong stereotypes about brilliance [44].

B. Dimensions of intelligence mindset

Researchers initially viewed intelligence mindset as a single continuum in which a growth and fixed mindset sit on either end [28]. However, interviews show that students may simultaneously have some growth mindset beliefs and fixed mindset beliefs, pointing to a need for more nuanced dimensional measures of mindset [14,45]. Technically speaking, the former approach is a "one-factor" model, while the latter is a "multifactor" model. Though the one-factor approach is still popular [43,46,47], there is a growing body of work that uses separable growth and fixed mindset dimensions [12,13,48–50]. For example, in a two-factor model, a student might report both some growth mindset and some fixed mindset beliefs. Such a student may understand that practice and hard work are necessary to excel in physics. However, that student may also believe a base level of ability is also needed and feel disempowered if they think that they do not personally posses that "necessary" talent or ability to excel.

Another conceptual divide in mindset research involves beliefs about self versus others. One study [51] found that high-school students conceptualized intelligence mindsets differently for themselves than for others. They also found that intelligence "self-theory" was a stronger predictor of academic performance than general intelligence mindsets. As noted in the next section, similar patterns were recently found with self versus other physics mindsets.

We aim to investigate if students had separable beliefs about growth and fixed mindsets, as well as if they held different mindset beliefs about themselves versus others. If student mindsets are separable along these divides, then there is an opportunity to learn which more specific mindsets are particularly important for learning outcomes or especially associated with gender differences. Those findings in turn would better enable targeted interventions.

C. Physics intelligence mindsets

Students may have different mindset beliefs in different domains and contexts. For example, they may believe that intelligence in general can change through hard work or that they in general have enough intelligence for most situations, but still have fixed mindsets about particular domains with especially strong stereotypes of innate brilliance such as physics. Physics-specific mindset research is relatively new [12–14,47]. One of these first studies found that physics-specific mindsets are both different from (via a factor analysis) and a better predictor of physics learning outcomes than general intelligence mindsets [47].

Further, many stereotypes about women and intelligence are domain specific. For example, women are perceived to have strengths in the arts and humanities and weaknesses in math and the sciences [21,52]. Physics in particular is a field with particularly strong stereotypes and biases about who belongs in and who can excel in the domain [20,21,53]. Both the general public [20] and working physicists [44] believe that success in physics requires innate talent or brilliance and societal narratives about talent and brilliance tend to ascribe these traits to boys and men [22,53,54]. Parents of girls are less likely to believe their child could succeed in a career that requires mathematical ability [55,56]. Boys are more likely than girls to receive positive recognition from their science instructors, including in physics courses [19,57,58]. Finally, there is evidence that physics intelligence mindsets become more fixed after taking a physics course, especially for women [47].

Recent research supports a four-way division of physics intelligence mindsets, and finds that one of the four physics-specific mindsets was especially predictive of introductory physics course grades in the male-dominated courses for physical science and engineering majors [12,13]. In particular, Kalander *et al.* were the first to find that physics intelligence mindsets can be divided into four dimensions along the combinations of me versus others and growth versus ability and the best fitting model to the survey data separately measures the four factors: my ability (students' beliefs about their own abilities), my growth (students' beliefs about their own potential to grow), others' ability (students' beliefs about others'abilities), and others' growth (students' beliefs about others' potential to grow) [12].

However, the Kalender *et al.* study uncovered these four mindset factors using a survey that was not specifically designed to measure four dimensions of physics intelligence mindset (i.e., had too few items per dimension) because this was not the original conception that drove the design of that survey instrument [12]. Malespina *et al.* then built upon this work in the same context by expanding the number of survey items and designing their structure to directly map onto the four hypotheses components and was able to replicate the original findings [13]. Further, both studies (each conducted in the calculus-based context for engineering and physical science majors) found that my ability was the best predictor of physics course grade, had the largest gender differences, and appeared to largely mediate the effects of gender on grades.

D. Research questions

Here the same survey items from the Malespina *et al.* study are used in a new context: introductory physics for bioscience majors [13]. The survey aims to pinpoint specific mindset beliefs (such as if a student holds different mindset beliefs for themselves versus their peers). Additionally, the measure is context-specific to physics. We will also examine whether mindset beliefs predict learning outcomes differently for men and women, as suggested by Yeager and Dweck's [43]. In addition, this research will investigate whether student grades are predicted by mindset across the full range of possible mindset levels or whether there are threshold effects such that mindset differences only matter at the high or low end. Here, we use "low," "medium," and "high" threshold values to measure student mindset. Though these thresholds are specific to the instrument used in this study, such threshold effects could help investigate which courses and students are most in need of intervention. For example, if outcomes for low and medium mindset values are similar, then it would be important to prioritize high scores for students through interventions and other means. We aim to answer the following research questions for students in introductory physics courses for bioscience majors at a large research university:

RQ1. Do physics intelligence mindsets organize into four factors (my ability, my growth, others' ability, and others' growth) as they did for students enrolled in physics for engineering and physical science majors?

RQ2.
a. Are there overall gender or sex differences in the means or distributions (in low, medium, and high categories) of students' physics intelligence mindset beliefs?
b. Are gender or sex differences in the means or distributions of students' physics intelligence mindset beliefs especially localized to particular dimensions?
c. Do gender or sex differences grow or decline during students' first university-level physics course?

RQ3. Do any of the mindset dimensions predict course grade and is the predictive relationship linear?

If the findings replicate what was found in the male-dominated introductory physics courses for physical science and engineering majors, then we expect: (i) four dimensions (my growth, my ability, others' growth, others' ability); (ii) men will have higher mindset scores than women, especially for my ability beliefs, and gender differences in all of the mindset factors will grow over time; and (iii) my ability is the best predictor of grade.

II. METHODOLOGY

A. Participants and procedures

We collected survey data at the beginning and end of the semester. Participants were students enrolled in a physics 1 course for bioscience and health-related majors. At this institution, introductory physics courses for bioscience majors are algebra-based, while courses for physical science and engineering majors are calculus-based. The physics 1 course primarily covered mechanics, though both thermodynamics and waves were also included. Faculty taught the course in a traditional lecture-based format alongside smaller-sized recitations taught by teaching assistants in which students work collaboratively on physics problems. Some active learning approaches were implemented, such as clicker questions, but the primary instructional technique was lecture. The student sample involved three different sections taught by three different instructors in one semester.

Surveys were handed out and collected by the teaching assistants in the first and last recitation class of a semester. Students were given course credit or extra credit for completing the survey, depending on the instructor's preference. We also implemented an attention check (an item that asked all students to answer "C"). In total, 32 ($< 1\%$) survey responses were excluded from the study because the student did not pass the attention check. The completion rate was 83% ($N = 547$) for the pre test and 78% ($N = 500$) for the post test. We focused upon the 428 students who took both surveys so we could observe students' change in motivational beliefs over time; however, similar findings were obtained when using the full set of respondents. An additional 9 students were excluded from the study due to missing demographic information or receiving an "incomplete" grade in the course. The final number of students in the presented analyses was 419.

Based upon institutional data, the longitudinally matched sample was 66% women (compared to 62% for all enrolled students), which is typical for introductory physics for bioscience majors courses at this institution. We note that response rates were slightly different by gender [pre (post) response rates were for 85%(79%) for women and 79% (70%) for men]. Students at this predominantly white institution (PWI) identified with the following races/ethnicities: 68% White, 19% Asian, 3% Hispanic/Latinx, 5% multiracial, and 5% African American/Black. This course is taken almost exclusively by students intending to pursue postgraduate work in the health fields (especially medicine). Most students were in their second (13%) or third (65%) year of university.

This research was carried out in accordance with the principles outlined in this institution's Institutional Review Board ethical policy, and de-identified demographic data were provided through university records. For some variables, such as high school GPA, this approach allows us to rely on records that may be more accurate than students' memories. However, it limits other measures such as student sex or gender, which students could only report as "male" or "female." We acknowledge the harm that collecting data this way can cause [59,60]. This institution recently began to implement more inclusive sex and gender reporting methods for students, which we plan to use once student samples are large enough to be meaningful in quantitative analysis.

B. Measures

1. Physics intelligence mindset

We adapted this mindset survey from previously validated surveys [12,13,47]. The survey was designed to measure mindsets across a self versus others dimension, as across a growth versus ability dimension. Initially, there were 19 items in the mindset survey, which can be found in Table VI in Appendix A.

After the questions were drafted, we conducted 20 1 h semistructured cognitive interviews to ensure that students interpreted questions as intended. Participants were students who had previously taken physics courses ranging from introductory to graduate level. One of the 19 survey items ("I will always be as good at physics as I was in high school.") was removed because the cognitive interviews indicated that students did not interpret it as intended [13].

This survey was designed to have four separable mindset beliefs based upon the combinations of those two dimensions [12,13]: my ability, my growth, others' ability, and others' growth. The items were distributed across mindsets as follows: six my ability items, and four items each for the three other constructs. See Appendix A for the full set of items. Each item used a common set of four response options (strongly disagree, disagree, agree, strongly agree). Responses were correspondingly coded as 1 to 4, with reverse coding for all my ability and others' ability questions such that higher values corresponded to mindsets hypothesized to support learning.

2. Prior academic preparation

High school grade point average (HS GPA) was reported using the weighted 0–5 scale, which is based on the standard 0 (failing)–4 (A) scale with adjustments for honors, Advanced Placement, and International Baccalaureate courses (all of these programs may offer a "weighted" GPA that adds up to one or two grade points as a reward to taking advanced courses, which can allow a GPA higher than 4.0). High school GPA is taken as a measure of general academic skills and generally is a strong predictor of early undergraduate course performance [61].

Students' Scholastic Achievement Test math (SAT math) scores are on a scale of 200–800 and were used as a predictor of performance on high-stakes assessments involving mathematical problem solving (e.g., physics exams) [61–63]. If a student took the American College Testing (ACT) examination, we converted ACT to SAT scores [64]. If a student took a test more than once the school provided the highest section-level score for the SAT and the highest composite score for the ACT. If a student took both ACT and SAT tests, we used their SAT score.

3. Course grade

Course grades were based on the 0–4 scale used at our university, with $A = 4$, $B = 3$, $C = 2$, $D = 1$, $F = 0$, or W (late withdrawal), where the suffixes "+" and "−," respectively, add or subtract 0.25 grade points (e.g., $B- = 2.75$ and $B+ = 3.25$), except for the A+, which is reported as 4.

C. Analysis

1. Survey validation

Confirmatory factor analysis (CFA) using the R package "lavaan" was used to provide quantitative validation for whether the survey items fit the proposed four mindset constructs. To evaluate whether the model was acceptable, we chose the following standards: standardized factor loadings of each item were greater than 0.5 [65] (p. 301), a Comparative Fit Index (CFI) and Tucker-Lewis index (TLI) greater than 0.90, a Root Mean Square Error of Approximation (RMSEA) less than or equal to 0.08, and a Standardized Root Mean Square Residual (SRMR) less than or equal to 0.08 [66].

We first investigated whether the conceptual division into four components in terms of growth or ability and myself or others was replicated in this course context. In particular, in addition to testing the fit of the model based upon the four categories, other models were also evaluated based upon other approaches to intelligence mindset. A one-factor model in which all items were included in a single construct was tested and rejected due to poor model fit. Two-factor models were also evaluated: one model divided items that asked about the self and others, and the other divided questions that asked about growth and ability. Both models were rejected due to poor model fits. The four-factor model resulted in the best overall model fits.

After deciding on a four-factor model, the item with the lowest factor loading was dropped, and the model was iteratively reevaluated with the remaining items. Items were dropped as long as fit indices improved or remained consistent and each factor had at least three items. This process produced a robust model while eliminating excess variables. After determining the items to include, we calculated Cronbach's α, a measure of internal consistency between items within a construct. A generally accepted value for Cronbach's α is between 0.70 and 0.90 [67].

To create latent variables, we calculated the mean score of the questions in each validated category using the reduced set of twelve survey items. As a reminder, all the mindset dimensions are scored from 1 to 4, and are coded such that a high score corresponds to agreeing strongly with growth or malleable physics mindset beliefs or disagreeing with fixed or ability mindset beliefs. We used mean scores for constructs because prior Rasch modeling [67] with this four-point scale for mindset items had found roughly equal psychological distance between levels [12] and because the correlation between simple mean scores and Rasch-adjusted person estimates are very high (e.g., usually above 0.99).

We also tested different levels of measurement invariance in the final CFA model to make sure the survey items functioned equally across gender groups given the focus of the current study. In each step, we fixed different elements of the model to equality across gender and compared the results to the previous step using the likelihood ratio test [65]. We did not find any statistically significant moderation by gender, supporting the use of mean scales scores in analyses of gender differences.

After completing the CFAs to determine the mindset scales, we addressed outliers in all mindset scale values (as well as in SAT/ACT math, course grade, and high school GPA) by winsorizing [67]. To winsorize the scores, we replaced outliers with values two standard deviations above or below the mean, so that we maintained the direction of the outlier without introducing extreme values that produce poor performance in the regression models.

Pearson correlations were calculated between the generated latent variables within a time point to provide information on potential problems of collinearity among predictors in the multiple regressions (e.g., Pearson $r > 0.70$). Further, pre-post Pearson correlations for each attitude were used to examine attitude stability over time: pre-post correlations below 0.3 would indicate low stability, correlations above 0.8 would indicated high stability, and intermediate values would indicated moderate stability. We also calculated Pearson correlations between each mindset dimension and course grades as a baseline prediction model.

2. Descriptive statistics

To analyze gender differences on all measures, we calculated means and standard deviations by gender and then we compared men and women's scores using unpaired t tests to measure statistical significance of the differences [67] and Cohen's d to measure the size of the difference [68]. Cohen's d is calculated using

$$d = \frac{\mu_1 - \mu_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}}, \tag{1}$$

where μ_1 and μ_2 are the mean values of each group and σ_1 and σ_2 are the standard deviations of each group [68]. Group one was women and group two was men. Cohen's d is considered small if $d \sim 0.2$, medium if $d \sim 0.5$, and large if $d \sim 0.8$ [68]. Levene's test was implemented to ensure that the homogeneity of variance assumption was met for the unpaired t tests [67].

Similarly, to compare students' mindset scores from pre to post, paired t tests [67] and Cohen's d effect size measures were also used. The change-over-time analyses were also conducted for all instructors separately to check for consistency of the patterns across instructors. Trends were generally similar between instructors. One instructor's class did not have statistically significant mindset decreases in all constructs for men, though the decrease was similar in magnitude to other instructors. This may be due to small class size, as there were only 22 men in that group of participants.

Finally, we divided students into groups that reported low (< 2.5), medium (2.5–3.5), and high (≥ 3.5) on the 1-to-4 scales (after recoding) for each mindset dimension. The specific thresholds were selected given the distribution of the data, as it was rare for students to select the lowest values for each survey item. We divided students into categories for two reasons. First, for instructors with large class sizes, strategically dividing students into groups with low, medium, or high mindset scores may be easier to manage than placing students into groups based upon a continuum. Second, analyzing the data this way provides a test of the linearity assumption in the regression analyses. Third, if effects were nonlinear, this could shape the scale of interventions that would be needed (e.g., for moving students from low all the way to high).

3. Predicting learning outcomes

First, multiple linear regression analysis was used to find partial correlations between mindset components and grades, controlling for gender or sex and prior academic preparation. For the quantitative analyses, gender or sex was coded as an indicator variable: women = 1, men = 0.

Regression analysis was chosen over hierarchical linear modeling because the interclass correlation coefficients of the motivational measure data in these larger lecture courses are adequately small (always < 0.10 and regularly < 0.04). Multiple models were evaluated in order to find which was the best predictor of learning outcomes and show robustness of relationships across model specification. All models used standardized regression coefficients as a measure of effect size. The models were implemented using the regress command in Stata [69]. To test the normality of errors, we compared a kernel density estimate of each model's residuals with a normal distribution. Each model had a normal distribution of residuals. We also ensured that predictor effects were not miss-estimated due to multicollinearity by implementing a variance inflation factor cutoff of 2.0 for each model.

The regression models were built incrementally to assess the robustness of the findings across different models. A baseline model predicted course grade using only gender or sex, high school GPA, and SAT math scores. Next, we added the mindset variables one by one in order of correlation strength with course grade until all mindset variables were included. All models with significant mindset variables were kept, along with the final model with all variables included as a robustness test. The regression analyses used an average across prescores and postscores. Average scores were used instead of presurvey or postsurvey scores for two reasons. First, using postsurvey scores raises a question of causality (did course performance affect mindset or did mindset affect course performance?). Second, the average score is a proxy for students' mindset during the semester, while they were taking the course, rather than after the class. Using average rather than only presurvey data is particularly important given the sizable changes from pre to post alongside only moderate pre-post stability that were observed in several of the attitudinal variables. The results of linear regressions

TABLE I. Survey items included in the study and standardized factor loadings for pre and postsurveys.

	Construct name or item	λ Pre	λ Post
	My growth ($\alpha_{\text{pre}} = 0.83$, $\alpha_{\text{post}} = 0.89$)		
1.	I can become even better at solving physics problems through hard work	0.77	0.85
2.	I am capable of really understanding physics if I work hard	0.85	0.91
3.	I can change my intelligence in physics quite a lot by working hard	0.80	0.84
	My ability ($\alpha_{\text{pre}} = 0.77$, $\alpha_{\text{post}} = 0.88$)		
4.	I won't get better at physics if I try harder	0.61	0.72
5.	I could never excel in physics because I do not have what it takes to be a physics person	0.80	0.88
6.	I could never become really good at physics even if I were to work hard because I don't have natural ability	0.84	0.90
	Others' growth ($\alpha_{\text{pre}} = 0.85$, $\alpha_{\text{post}} = 0.81$)		
7.	People can change their intelligence in physics quite a lot by working hard	0.84	0.77
8.	If people were to spend a lot of time working on difficult physics problems, they could develop their intelligence in physics quite a bit	0.83	0.80
9.	People can become good at solving physics problems through hard work	0.74	0.80
	Others' ability ($\alpha_{\text{pre}} = 0.71$, $\alpha_{\text{post}} = 0.75$)		
10.	Only a few specially qualified people are capable of really understanding physics	0.68	0.70
11.	To really excel in physics, people need to have a natural ability in physics	0.73	0.80
12.	If a student were to often make mistakes on physics assignments and exams, I would think that maybe they are just not smart enough to excel in physics	0.62	0.65

using either presurvey or postsurvey scores can be found in the Supplemental Material [70].

After using a linear model, we also implemented regression analysis using a threshold method based upon the low, medium, and high categories described in the preceding section. Instead of using continuous variables for each mindset score, these models used dummy variables for the two higher thresholds, treating low as the contrast group. For each mindset component, we performed a regression controlling for SAT math, high school GPA, and gender. Finally, if a mindset belief dummy variable predicted grade, we performed each regression again, but included an interaction between gender and that mindset belief dummy variable. Such an interaction test could reveal whether a dimension predicts course grade for one gender but not another or to a much larger extent for one gender.

III. RESULTS

A. RQ1. Do physics intelligence mindsets organize into four factors (my ability, my growth, others' ability, and others' growth) as they did for students taking physics for engineering and physical science majors?

Initially, there were 19 items in the mindset survey, which can be found in Appendix A. A one-factor (in which all items were contained in the same construct) and both two-factor models (in which one model construct used "growth" and "ability" items, and another model used "me" and "others" factors) were rejected due to poor overall model fit. After deciding on a four-factor model, six additional survey items were removed (that is, they were completely excluded from the analysis) during the CFA model testing process due to low factor loadings or cross loading that led to a poor model fit. The factor loadings for the 12 remaining items in the final model can be seen in Table I. This model meets all chosen fit index cutoffs. Standardized factor loadings of each item were all greater than 0.5 [65] (p. 301), as seen in Table I. All other fit indices (CFI, TLI, SRMR, and RMSEA) along with their cutoff values [66], can be seen in Table II. Thus, answering RQ1, the final model had the four predicted mindset constructs: My growth, my ability, others' growth, and others' ability. Further, each construct was based upon three items, similar to prior work on mindsets [12,13,71]. In Table I, Cronbach's α values were between 0.71 and 0.89 for all constructs for both the presurveys and postsurveys. There are some changes in α for each construct from pre to post. However, small changes in α are generally not treated as a topic of concern if both fall within the acceptable values range of 0.70 to 0.90 [67].

TABLE II. Fit indices for the CFAs testing survey validity at pre and post, along with the applied fit index cutoffs. There were 419 students included in the factor analysis.

	CFI	TLI	RMSEA	SRMR
Cutoff	$\geq$0.90	$\geq$ 0.90	$\leq$ 0.08	$\leq$ 0.08
Pre	0.97	0.95	0.06	0.05
Post	0.97	0.95	0.07	0.04

TABLE III. Pearson correlations between each mindset construct as well as physics 1 course grade. The following abbreviations are used: my ability (MA), my growth (MG), others' ability (OA), and others' growth (OG). $p < 0.001$ unless otherwise noted by $^{*} = p < 0.05$, $^{**} = p < 0.01$, and $^{\mathrm{ns}}$ = not statistically significant.

	Pre				Post				
	MG	MA	OG	OA	MG	MA	OG	OA	Grade
MG pre					0.28				0.08^{ns}
MA pre	0.53					0.47			0.10^{*}
OG pre	0.58	0.47					0.36		-0.03^{ns}
OA pre	0.42	0.51	0.43					0.44	0.06^{ns}
MG post									0.25
MA post					0.67				0.34
OG post					0.59	0.48			0.16
OA post					0.53	0.67	0.46		0.17

Finally, the upper left and lower right of Table III reveal that intercorrelations among the scales at pre and at post are all moderate and positive (after reverse coding of ability mindsets), but none are so high ($r > 0.7$) as to represent redundant measures. Correlations within each construct from pre to post (upper right of Table III) showed low to moderate stability of the mindsets during this course experience, with especially low stability of the my growth mindset.

B. RQ2. Gender differences

1. Are there overall gender or sex differences in the means or distributions (in low, medium, and high categories) of students' physics intelligence mindset beliefs?

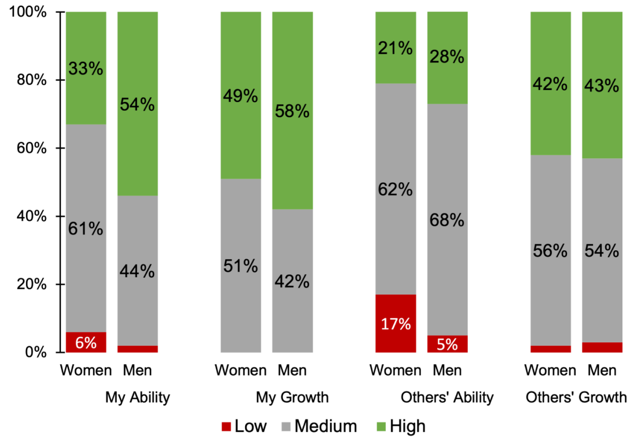
The winsorized means of both men's and women's mindset dimensions can be found in Table IV. As a reminder, all the mindset constructs are scored from 1 to 4, and are coded such that high scores correspond to a strong agreement with growth mindset beliefs or rejection of a fixed mindset beliefs. Table IV also shows the unpaired t tests and Cohen's d effect sizes comparing men and women's mindsets at the beginning of the semester.

Answering the means component of RQ2a, as expected based upon prior work, men generally had higher mindset scores than women, as shown in Table IV. The smallest gender differences were nonsignificant, while the largest differences had moderate effect sizes. Men had higher mindset scores than women in every mindset category both pre and post.

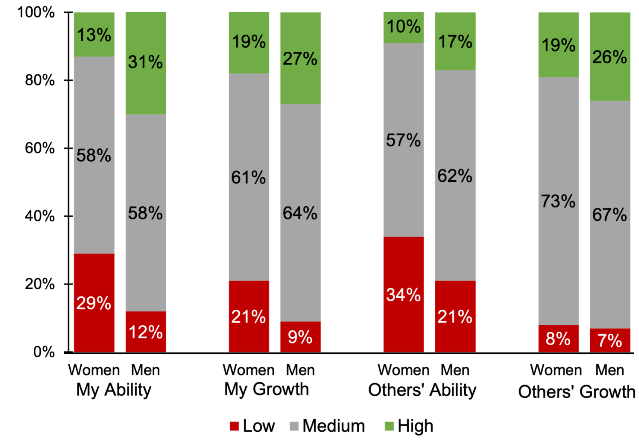
Next, we divided students into groups that reported low (< 2.5), medium (2.5–3.5), and high (≥ 3.5) on a 1-to-4 scale for each mindset dimension. These student

TABLE IV. Mean and standard deviation (SD) by gender or sex of each mindset factor at pre and post, along with SAT Math, HS GPA, and Physics 1 grade, Cohen's d and t test of gender or sex differences. Positive values of d indicate that women had a higher score or that scores increase over time. $^{*} = p < 0.05$, $^{**} = p < 0.01$, and $^{***} = p < 0.001$.

		Women ($n = 276$)		Men ($n = 143$)		
Variable		Mean	SD	Mean	SD	d between genders
My growth	Pre	3.45	0.47	3.54	0.45	−0.19
	Post	2.95	0.60	3.20	0.57	-0.42^{***}
	d over time	-0.92^{***}		-0.69^{***}		
My ability	Pre	3.28	0.48	3.52	0.48	-0.51^{***}
	Post	2.82	0.64	3.20	0.57	-0.61^{***}
	d over time	-0.81^{***}		-0.63^{***}		
Others' growth	Pre	3.37	0.47	3.40	0.48	−0.08
	Post	3.09	0.44	3.20	0.51	-0.24^{*}
	d over time	-0.62^{***}		-0.41^{***}		
Others' ability	Pre	3.03	0.56	3.24	0.47	-0.40^{***}
	Post	2.73	0.60	3.00	0.61	-0.44^{***}
	d over time	-0.52^{***}		-0.47^{***}		
HS GPA		4.19	0.38	4.07	0.42	0.30^{**}
SAT/ACT math		669	65	688	65	-0.29^{**}
Course grade		2.91	0.76	3.22	0.71	-0.41^{***}



(a) Pre survey distributions



(b) Post survey distributions

FIG. 1. Percentages of students who reported low (< 2.5), medium (2.5–3.5), or high (≥ 3.5) on a 4-point Likert scale, by gender. (a) Presurvey distributions; (b) postsurvey distributions. If any category contains ≤ 5% of students, the percent is not labeled.

distributions can be seen in Fig. 1(a) for the presurvey and Fig. 1(a) for the postsurvey. As an important context note, more students in this course had growth rather than fixed mindsets: average scores for both genders are closer to 4 than to 1 for the presurvey. Distributions of mindset scores by gender showed similar trends to the means. For all constructs, men were more likely than women to fall into the high category, which can be seen in Fig. 1. Similarly, for all constructs women were more likely than men to fall into the low category. There was only one exception to this trend, but this category did not have any student in the low category. Figure 1 also reveals that at the end of the semester men continue to be more likely to fall into the high category and are less likely to fall into the low category.

2. Are gender or sex differences in the means or distributions of students' physics intelligence mindset beliefs especially localized to particular dimensions?

Though men generally had higher mindset scores than women, the size of the gender differences varied by mindset construct, which we will now discuss individually. My growth beliefs had no statistically significant mean gender difference between men and women at the start of the semester (see Table IV). For this construct, approximately half of women report high scores for prebeliefs (compared to 58% of men). By contrast, pre my ability beliefs had the largest mean gender difference of any of the four precon-structs. These gender differences are also apparent in the score distributions that are found in Fig. 1(a). At pre, over half of men report a high my ability score, while only one-third of women do.

The dimension that had the smallest pre gender or sex mean difference was others' growth. In this category, men and women had indistinguishable scores, shown in Table IV. Others' growth also had the smallest gender differences in prescore distributions. Figure 1(a) shows that less than 5% of both men and women reported low prescores, and a similar portion of men and women reported high scores (42% versus 43%).

Others' ability had the lowest prescores of any construct. Additionally, others' ability had statistically significant gender differences in prescores. The others' ability gender differences are moderate at the start of the semester. As others' ability had the lowest scores of any construct, it also had the largest portion of students reporting low beliefs. In particular, 17% of women reported low pre others' ability beliefs, compared to < 5% of men. Men were also more likely than women to report high pre others' ability scores.

Broadly, we found that men tended to report higher mean prescores in all four mindset constructs than women. The categorical distributions had a similar trend: women were more likely to report low scores and men were more likely to report high scores for all four pre mindset factors. However, to answer RQ2b, the gender differences were smallest in the my growth and others' growth factors, and largest in the my ability and others' ability factors.

3. Do gender or sex differences grow or decline during students' first university-level physics course?

All students had statistically significant drops in mindset beliefs over time. The decreases in mean scores for each factor can be seen in Table IV and the changes in student score distributions can be found in Fig. 1. Men had similar but usually smaller decreases than women, with quantitative variations by construct.

The my growth construct did not have statistically significant mean gender differences at the start of the semester. However, by the end of the semester, there is a moderate and significant mean gender difference in this construct. This growth in gender difference is the largest of any construct, and appears to be due to a dramatic drop ($d = -0.92$) in women's my growth beliefs. This decrease is also shown in Figs. 1(a) and 1(b), which shows the distributions of students low, medium, and high mindset beliefs at post. At the end of the semester, only 19% of women have high my growth beliefs, compared to 27% of men. Both men and women have lower my growth beliefs at the end of the semester, but the shift was more drastic for women.

Turning to the mindset with the largest gender difference, Table IV reveals that my ability gender differences grew from $d = -0.51$ at pre to $d = -0.61$ at post. These gender differences are also apparent in the score distributions that are found in Fig. 1. Both men and women show decreases in the number of students reporting high scores from pre to post, but at the end of the semester men were over twice as likely as women to report high my ability scores. Women were also more than twice as likely as men to report low post my ability scores.

By contrast, focusing on the dimension that had the smallest gender or sex differences, others' growth, men and women were indistinguishable at pre, and then there was a small gender difference at post. For pre others' growth, Fig. 1 shows that less than 5% of both men and women reported low scores, and a similar portion of men and women reported high scores (42% versus 43%). Post others' growth distributions show minimal gender differences for low scores, but 7% more men than women report high my growth scores at the end of the semester.

The construct with the lowest scores, others' ability, also showed the lowest (but still substantial) declines, and these declines were similar for men and women. Thus, the gender difference was similar at pre and post, but in the context of overall relatively low scores. At the end of semester, one-third of women and one-fifth of men had low scores in this construct. Women were also less likely than men to report high post scores.

In sum, to answer RQ3c, mindset scores generally declined over time, but they tended to decrease more for women than men, leading to larger gender differences at the end of the semester than at the start. The largest increases in gender differences from the start to the end of the semester were for my growth and others' growth beliefs. The smallest increases in gender differences from the start to the end of the semester were for my ability and others' ability beliefs. However, my ability had the largest differences both pre and post.

C. RQ3. Do any of the mindset dimensions predict course grade and is the predictive relationship linear?

First, we conducted multiple regression analysis to find which of the four mindset beliefs best predicted physics course grade. We conducted this analysis two ways. First, we used linear regression using students' mean scores for each construct (see Table V). Each model was conducted using the average of pre and postsurvey mindset scores due to the large changes in mindsets across the semester.

TABLE V. Linear regression models predicting final course grade using average mindset beliefs. The regression (β) coefficiants are standardized, and the gender or sex variable was coded such that women $= 1$ and men $= 0$. $^{*} = p < 0.05$, $^{**} = p < 0.01$, and $^{***} = p < 0.001$. $N = 418$.

Variable	Model 1	Model 2	Model 3
Gender	-0.18^{***}	-0.12^{**}	-0.11^{**}
HS GPA	0.34^{***}	0.32^{***}	0.32^{***}
SAT/ACT math	0.39^{***}	0.39^{***}	0.38^{***}
My ability average		0.21^{***}	0.19^{**}
My growth average			0.14^{**}
Others' ability average			-0.04
Others' growth average			-0.09
Adjusted R^2	0.37	0.41	0.41

Similar models using pre and postsurvey results can be found in the Supplemental Material [70].

These models needed to include controls because there were also gender or sex differences in prior academic performance, although in opposing directions: for our sample women tended to have higher high school GPAs than men, but lower SAT math scores, as seen in Table IV. However, men in the sample tended to have higher physics 1 course grades than women. Here we investigate whether mindset differences could account for this gender difference in grade outcome.

Model 1, which can be seen in Table V uses only gender, SAT/ACT math scores, and HS GPA to predict students' physics 1 course grades. All three predictors are statistically significant. This model establishes that women had lower physics 1 course grades than men, even when controlling for High School GPA and standardized test scores, formally establishing that other factors are needed to account for gender or sex differences in course performance.

Model 2 includes my ability as a fourth predictor. My ability was chosen the first predictor to add because it has the strongest correlation with course grade for both the pre and post mindset components (see Table III). Model 2 in Table V reveals that adding average my ability to the model weakens the relationship between gender or sex and physics 1 grade, though it remains a statistically significant predictor. Model 2 also has an increase in adjusted R-squared compared to model 1. This means that model 2 explains more of the variance in course grades than model 1 even with a penalty for having an additional predictor [67].

Model 3 includes all four mindset components. This model reveals that both my growth and my ability are positively correlated with physics 1 course grades. Adding these other constructs marginally decreases the regression (β) coefficients of gender, SAT/ACT Math, and my ability. Additional model testing revealed that my growth average is a not statistically significant predictor of course grade unless either or both others' ability or others' growth are

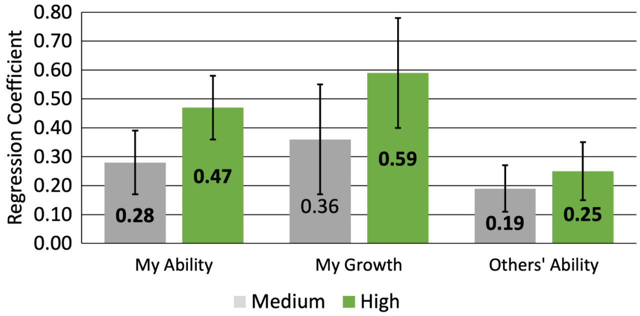


FIG. 2. Unstandardized regression coefficients predicting course grade, controlling for gender, high school GPA, and SAT or ACT math scores. Error bars represent standard error. On a 4-point Likert scale, a medium score is 2.5–3.5, and a high score is ≥ 3.5. Statistically significant regression coefficients are in bold. Others' growth is excluded because it did not predict course grade.

added to the model. Most importantly, Model 3 shows that the "my" dimensions of mindset (especially my ability) are stronger predictors of physics 1 grades than "others" dimensions.

Next, we conducted the regression analyses predicting grades using two dummy variables for each mindset construct: one for students categorized as reporting having medium mindsets and one for students categorized as having high mindsets; the regressions then treat the low group as the reference category [[67], p. 551]. As a reminder, on the 4-point Likert scale, low was < 2.5, medium was 2.5–3.5, and high was ≥ 3.5.

We focus on the models of this type that added dummy codes for each mindset factor individually, rather than using all four factors simultaneously. Models that used all mindset constructs simultaneously can be found in the supplementary materials. For each construct, we first introduce a model that predicts physics 1 course grade using each of the four mindset constructs (my ability, my growth, others' ability, and others' growth), gender or sex, SAT/ACT math scores, and high school GPA. These models can all be seen in Table VII in Appendix B. When both gender and mindset predictors were statistically significant, we proceeded to include interaction terms to test whether men and women's mindset scores predicted course grade to a different extent.

Looking across the factors, the dummy code approach replicated the high-level findings of the linear modeling approach in that strong agreement with my growth beliefs was the best predictor of course grade. On the other hand, both medium and high agreement with my ability beliefs predicted course grade (Fig. 2). Others' ability also was a statistically significant predictor, in contrast to being not significant as a linear predictor. On a related point, the support for linearity of effects was weak. Saliently, medium and high effects for others' ability were almost identical, potentially explaining why the linear model was not significant. However, even in the cases of my ability and my growth, the effect of high levels of agreement was not statistically different from medium levels. Others' growth was not included in Fig. 2 because neither medium nor high levels predicted course grade. Additionally, no statistically significant gender interaction term was found for any of the average mindset components. In other words, the relationship between mindset and course grade was similar for men and women, validating the use of simpler models that did not include interaction terms.

IV. DISCUSSION

Regarding RQ1, we found four components to students' mindsets (my ability, my growth, others' ability, and others' growth) using a survey instrument designed to specifically test for these components. This result replicated the findings of past work using previous iterations of the survey [12,13]. These findings build on past work that separated mindset into multiple constructs, either between my versus others' mindset dimensions [51] or between growth and ability dimensions [48–50].

The four components were only moderately correlated with one another (18%–34% shared variance at pre) and were separable in CFA models. Further, though each component showed similar patterns of gender difference, and change over time, the magnitudes of these effects were different and each component predicted course grades with different strengths. Thus, our components were not only separated by psychometric analyses, but by empirical patterns as well. Therefore, future research should avoid collapse measurement of mindsets into overall intelligence mindset scores.

Regarding RQ2, there were gender differences in the pre survey means of my ability and others' ability, and men tended to report higher scores than women. There were gender or sex differences in the distributions of all mindset beliefs. More men fell into the high score range for each mindset component, and more women fell into the low score ranges. These differences were more pronounced for my ability and others' ability pre. Women in this context were more likely than men to believe that physics requires innate ability. They were also more likely than men to believe that they did not personally have this innate ability. This is particularly concerning for this student sample, which consists of relatively high-achieving students who had decided to pursue bio- and health-science majors.

Even though women were the numerical majority in this context, the gender or sex differences in the means of each mindset component increased substantially from pre to post for my growth and others' growth. As a result, men reported much higher mean scores than women for all four mindset constructs at the end of the semester. By the end of the semester, women were more likely to report low mindset scores than men in three out of four constructs (men and women reported low others' growth scores at similar rates). Men were more likely than women to report high mindset scores for all four mindset constructs. These inequities are not present because men had steady or increasing mindset beliefs. Instead, both men and women had moderate-to-large drops in all four mindset components during the semester. Women tended to have larger drops than men, creating or widening gender differences. It is important to note that, while there are gender inequities in student mindsets, the decreases over time are concerning themselves. Research in physics mindsets has found decreases in mindset beliefs for both men and women [12,13]. Motivational factors, such as self-efficacy or interest, commonly tend to decrease over time in introductory physics courses, both for physical science and engineering majors [72,73], as well as for bioscience majors [74].

A large body of research shows that women tend to have lower motivational characteristics in physics courses than men [47,75,76], including mindsets [12,13]. Most of this research has been conducted in physics courses in which men outnumber women, such as courses for engineering and physical science majors. This study shows that gender differences also exist in courses in which women outnumber men, such as physics courses aimed at bioscience majors.

One similarity between mindset trends in these two types of courses is that students enter the course with gender differences in mindset. One important difference is that, while women in both categories of courses show moderate-to-large decreases in mindset scores, men show much larger decreases in mindset scores in courses aimed at bioscience, rather than engineering or physical science students [12,13]. In this context, efforts to increase mindset scores may benefit all students while simultaneously decreasing gender differences in physics intelligence mindset.

Our data do not allow us to know why women show larger declines in mindset scores over time, but we hypothesize that classroom culture and instructors' efforts to encourage equity and inclusion in the classroom may affect gendered student outcomes. Ideally, instructors will adopt and adapt equitable and inclusive classroom policies and pedagogies in their classroom. In physics, this may include countering common societal stereotypes and beliefs that physics requires innate "brilliance" [44]. These beliefs, coupled with a lack of female physicists as role models throughout their life may lead women to feel that they are not personally capable of excelling in physics. One example of how to counter these beliefs is explicitly stating to students that failure is a normal and necessary part of learning, and that physics can only be learned through effort and practice.

Successful student-level mindset interventions tend to provide opportunities for self-reflection and show students that they can change their own intelligence. For example, students may be asked to remember instances during which they were able to improve their abilities [5,77]. Interventions may also share stories from a diverse group of peers or experts about overcoming academic challenges, so that all students may find someone they can relate to. If a relatable role model shares that they were able to work hard to achieve success instead of relying on innate talent, students may realize they can do the same and develop growth mindsets for themselves [5,26]. It is important to note that successful interventions generally require training and mentoring for the person who implements them, as poorly implemented interventions may be ineffective or harmful.

Instructor-focused change can be useful as well. Instructor mindsets can predict student achievement in their courses [78]. In addition, students in courses taught by instructors with growth mindsets report increased interest in their courses as well as fewer concerns about fair treatment and low grades [79]. Instructors with growth mindsets encourage students to accept mistakes and failures as a part of a normal learning process, congratulate persistence, praise effort rather than intelligence when students succeed [80–84], and are more likely to implement active learning in their courses [85]. On the other hand, instructors with fixed mindsets tend to have low expectations of students they believe lack natural talent, which can lead instructors to give easier assignments or encourage students to drop difficult classes because of presumed low ability [81].

V. CONCLUSIONS

This study shows that intelligence mindset can be divided into four constructs: my ability, my growth, others' ability, and others' growth. Previous work in studying mindset has divided along either growth or ability or me or others categories, but rarely simultaneously. Next, this work reveals that gender or sex differences are more pronounced in the ability categories than the growth categories. Gender or sex differences developed or widened over the semester for all mindset constructs. These differences are the result of substantial drops in all four mindset factor scores from all students, which tended to be even larger for women than for men.

Next, students' mindset scores decrease over the semester for all four constructs. They also show women's mindset scores decrease more than men's. We also find that my ability and my growth consistently predict the course grade. my ability positively predicts grades if students report a medium or high score, but my growth only predicts grades if students report a high score. This information may be useful to target mindset interventions to student beliefs. A student who believes nobody can become more intelligent through hard work may have different needs than one who believes that most people can become more intelligent but that they personally lack the ability to do so.

VI. LIMITATIONS AND FUTURE DIRECTIONS

We now note several limitations of this study. First, the analyses were correlational in nature: the causal nature of physics intelligence mindsets would need to be further supported through intervention studies and interview data. The established benefits of other mindset interventions [80]) suggest such a causal link is plausible, and we note that future interventions that focus on my ability and my growth may show even larger effects.

Another limitation is the generalizability of the findings. The studied institution is predominantly white, so we were unable to study if mindset beliefs differ or predict grades differently for students of different racial or ethnic backgrounds due to low sample size. Because of the focus on gender or sex in this study, future work should also explicitly include students who fall outside of the binary gender or sex categories included here, as well as transgender students who may not have their gender accurately recorded by the university. Though this university recently began to include more sex or gender options for students, qualitative studies may be more appropriate to understand mindsets in these marginalized populations until student samples are large enough to be meaningful in quantitative analysis.

ACKNOWLEDGMENTS

This work was funded by NSF Award No. DUE-1524575.

APPENDIX A: ORIGINAL SURVEY ITEMS

TABLE VI. Original survey items. Italicized items were removed from analysis during validation. Item 10 was removed during interviews because students did not interpret the question as intended. All other items were removed to during statistical survey validation.

Item No.	Item
	My growth
1.	I can become even better at solving physics problems through hard work
2.	I am capable of really understanding physics if I work hard
3.	I can change my intelligence in physics quite a lot by working hard
4.	*Struggling with difficult physics problems would help me develop mastery in physics*
	My ability
5.	I won't get better at physics if I try harder
6.	I could never excel in physics because I do not have what it takes to be a physics person
7.	I could never become really good at physics even if I were to work hard because I don't have natural ability
8.	*If I were to often make mistakes on physics assignments and exams, I would think that maybe I'm just not smart enough to excel in physics.*
9.	*I won't get better at physics if I try harder*
10.	*I will always be as good at physics as I was in high school*
11.	*I will always get the same physics grade whether I try or not*
	Others' growth
12.	People can change their intelligence in physics quite a lot by working hard
13.	If people were to spend a lot of time working on difficult physics problems, they could develop their intelligence in physics quite a bit
14.	People can become good at solving physics problems through hard work
15.	*If people were to persist in struggling with difficult physics problems, they would develop mastery in physics*
	Others' ability
16.	Only a few specially qualified people are capable of really understanding physics
17.	To really excel in physics, people need to have a natural ability in physics
18.	If a student were to often make mistakes on physics assignments and exams, I would think that maybe they are just not smart enough to excel in physics
19.	*If people really have to struggle to solve physics problems, that means they are just not physics people.*

APPENDIX B: PREDICTING GRADES WITH CATEGORICAL MINDSET COMPONENTS

TABLE VII. Unstandardized regression coefficients for each average mindset categorical component. SAT/ACT math scores have been divided by 10 (such that they are on a 20–80). Medium scores are ≥ 2.5, and high scores are ≥ 3.5. The gender or sex variable was coded such that women = 1 and men = 0. An "×" indicates an interaction term between two variables. $^{*} = p < 0.05$, $^{**} = p < 0.01$, and $^{***} = p < 0.001$.

	My ability		My growth		Others' ability		Others' growth
Variable	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1
Gender	-0.22^{**}	-0.03	-0.25^{***}	-0.02	-0.26^{**}	-0.52^{*}	-0.29^{***}
HS GPA	0.62^{***}	0.62^{***}	0.63^{***}	0.63^{***}	0.66^{***}	0.66^{***}	0.65^{***}
SAT/ACT math	0.05^{***}	0.05^{***}	0.04^{***}	0.04^{***}	0.05^{***}	0.05^{***}	0.05^{***}
Medium	0.28^{**}	0.48	0.36	0.63	0.19^{*}	-0.05	-0.11
High	0.47^{***}	0.61	0.59^{**}	0.82	0.25^{*}	0.06	-0.04
Gender × Medium		-0.23		-0.31		0.30	
Gender × High		-0.13		-0.23		0.22	
Adjusted R^2	0.39	0.39	0.40	0.39	0.38	0.38	0.3

[1] National Science Foundation, *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2019* (National Science Foundation, Washington, DC, 2020), https://ncses.nsf.gov/pubs/nsf19304/digest/field-of-degree-women.

[2] N. Abramzon, P. Benson, E. Bertschinger, S. Blessing, G. L. Cochran, A. Cox, B. Cunningham, J. Galbraith-Frew, J. Johnson, L. Kerby, E. Lalanne, C. O'Donnell, S. Petty, S. Sampath, S. Seestrom, C. Singh, C. Spencer, K. S. Woodle, and S. Yennello, Women in physics in the United States: Recruitment and retention, AIP Conf. Proc. **1697**, 060045 (2015).

[3] D. Doucette and C. Singh, Why are there so few women in physics? Reflections on the experiences of two women, Phys. Teach. **58**, 297 (2020).

[4] E. H. Ecklund, A. E. Lincoln, and C. Tansey, Gender segregation in elite academic science, Gender and Society **26**, 693 (2012).

[5] K. R. Binning, N. Kaufmann, E. M. McGreevy, O. Fotuhi, S. Chen, E. Marshman, Z. Y. Kalender, L. Limeri, L. Betancur, and C. Singh, Changing social contexts to foster equity in college science courses: An ecological-belonging intervention, Psychol Sci. **31**, 1059 (2020).

[6] E. Brewe, V. Sawtelle, L. H. Kramer, G. E. O'Brien, I. Rodriguez, and P. Pamelá, Toward equity through participation in modeling instruction in introductory university physics, Phys. Rev. ST Phys. Educ. Res. **6**, 010106 (2010).

[7] D. Doucette, R. Clark, and C. Singh, Hermione and the secretary: How gendered task division in introductory physics labs can disrupt equitable learning, Eur. J. Phys. **41**, 035702 (2020).

[8] A. Traxler and E. Brewe, Equity investigation of attitudinal shifts in introductory physics, Phys. Rev. ST Phys. Educ. Res. **11**, 020132 (2015).

[9] B. Van Dusen and J. Nissen, Equity in college physics student learning: A critical quantitative intersectionality investigation, J. Res. Sci. Teach. **57**, 33 (2020).

[10] L. Holman, D. Stuart-Fox, and C. E. Hauser, The gender gap in science: How long until women are equally represented?, PLOS Biology **16**, 1 (2018).

[11] A. Johnson, M. Ong, L. T. Ko, J. Smith, and A. Hodari, Common challenges faced by women of color in physics, and actions faculty can take to minimize those challenges, Phys. Teach. **55**, 356 (2017).

[12] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Framework for unpacking students' mindsets in physics by gender, Phys. Rev. Phys. Educ. Res. **18**, 010116 (2022).

[13] A. Malespina, C. D. Schunn, and C. Singh, Whose ability and growth matter? Gender, mindset and performance in physics, Int. J. STEM Educ. **9**, 28 (2022).

[14] A. J. Little, B. Humphrey, A. Green, A. Nair, and V. Sawtelle, Exploring mindset's applicability to students'

experiences with challenge in transformed college physics courses, Phys. Rev. Phys. Educ. Res. **15**, 010127 (2019).

[15] 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).

[16] 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).

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

[18] J. Wang and Z. Hazari, Promoting high school students' physics identity through explicit and implicit recognition, Phys. Rev. Phys. Educ. Res. **14**, 020111 (2018).

[19] Z. Hazari and C. Cass, Towards meaningful physics recognition: What does this recognition actually look like?, Phys. Teach. **56**, 442 (2018).

[20] E. Reuben, P. Sapienza, and L. Zingales, How stereotypes impair women's careers in science, Proc. Natl. Acad. Sci. U.S.A. **111**, 4403 (2014).

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

[22] L. Bian, S.-J. Leslie, and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, Science **355**, 389 (2017).

[23] L. Bian, S.-J. Leslie, and A. Cimpian, Evidence of bias against girls and women in contexts that emphasize intellectual ability, Am. Psychol. **73**, 1139 (2018).

[24] The College Board, Advanced Placement 2021 Program Summary report (2021), https://reports.collegeboard.org/media/pdf/2021-ap-program-summary-report_1.pdf.

[25] S. Cheryan and V. C. Plaut, Explaining underrepresentation: A theory of precluded interest, Sex Roles **63**, 475 (2010).

[26] A. Bandura, W. Freeman, and R. Lightsey, Self-efficacy: The exercise of control, J. Cognit. Psychotherapy **13**, 158 (1999).

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

[28] C. S. Dweck, *Mindset: The New Psychology of Success* (Random House, New York, 2006).

[29] A. Rattan, K. Savani, D. Chugh, and C. S. Dweck, Leveraging mindsets to promote academic achievement: Policy recommendations, Perspect. Psychol. Sci. **10**, 721 (2015).

[30] D. S. Yeager and C. S. Dweck, Mindsets that promote resilience: When students believe that personal characteristics can be developed, Educ. Psychol. **47**, 302 (2012).

[31] K. Muenks and D. B. Miele, Students' thinking about effort and ability: The role of developmental, contextual, and individual difference factors, Rev. Educ. Res. **87**, 707 (2017).

[32] C. S. Dweck, *Is Math a Gift? Beliefs That Put Females at Risk* (American Psychological Association, Washington, DC, 2007).

[33] L. B. Limeri, N. T. Carter, J. Choe, H. G. Harper, H. R. Martin, A. Benton, and E. L. Dolan, Growing a growth mindset: Characterizing how and why undergraduate students' mindsets change, Int. J. STEM Educ. **7**, 35 (2020).

[34] C. Good, J. Aronson, and M. Inzlicht, Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat, J. Appl. Dev. Psychol. **24**, 645 (2003).

[35] L. S. Blackwell, K. H. Trzesniewski, and C. S. Dweck, Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention, J. Child Dev. **78**, 246 (2007).

[36] K. Kricorian, M. Seu, D. Lopez, E. Ureta, and O. Equils, Factors influencing participation of underrepresented students in STEM fields: Matched mentors and mindsets, Int. J. STEM Educ. **7**, 16 (2020).

[37] A. Rattan, K. Savani, M. Komarraju, M. M. Morrison, C. Boggs, and N. Ambady, Meta-lay theories of scientific potential drive underrepresented students' sense of belonging to science, technology, engineering, and mathematics (STEM), J. Pers. Soc. Psychol. **115**, 54 (2018).

[38] D. S. Yeager, P. Hanselman, G. M. Walton, J. S. Murray, R. Crosnoe, C. Muller, E. Tipton, B. Schneider, C. S. Hulleman, C. P. Hinojosa, D. Paunesku, C. Romero, K. Flint, A. Roberts, J. Trott, R. Iachan, J. Buontempo, S. M. Yang, C. M. Carvalho, P. R. Hahn, M. Gopalan, P. Mhatre, R. Ferguson, A. L. Duckworth, and C. S. Dweck, A national experiment reveals where a growth mindset improves achievement, Nature (London) **573**, 364 (2019).

[39] D. S. Yeager, V. Purdie-Vaughns, J. Garcia, N. Apfel, P. Brzustoski, A. Master, W. T. Hessert, M. E. Williams, and G. L. Cohen, Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide, J. Exper. Psychol. **143**, 804 (2014).

[40] C. Bagès, C. Verniers, and D. Martinot, Virtues of a hardworking role model to improve girls' mathematics performance, Psychol. Women Q. **40**, 55 (2015).

[41] D. S. Yeager, C. Romero, D. Paunesku, C. S. Hulleman, B. Schneider, C. Hinojosa, H. Y. Lee, J. O'Brien, K. Flint, A. Roberts, J. Trott, D. Greene, G. M. Walton, and C. S. Dweck, Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school, J. Educ. Psychol. **108**, 374 (2016).

[42] V. F. Sisk, A. P. Burgoyne, J. Sun, J. L. Butler, and B. N. Macnamara, To what extent and under which circumstances are growth mind-sets important to academic achievement? Two meta-analyses, Psychol. Sci. **29**, 549 (2018).

[43] D. S. Yeager and C. S. Dweck, What can be learned from growth mindset controversies?, Am. Psychol. **75**, 1269 (2020).

[44] S.-J. Leslie, A. Cimpian, M. Meyer, and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, Science **347**, 262 (2015).

[45] R. E. Scherr, M. Plisch, K. E. Gray, G. Potvin, and T. Hodapp, Fixed and growth mindsets in physics graduate admissions, Phys. Rev. Phys. Educ. Res. **13**, 020133 (2017).

[46] C. Kepple, M. Tilahun, N. Matti, and K. Coble, Pedagogy training for the development of GTA mindsets and inclusive teaching practices, presented at PER Conf. 2017, Virtual Conference, 2020, 10.1119/perc.2020.pr.Kepple, pp. 272–277.

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

[48] D. A. Cook, R. M. Castillo, B. Gas, and A. R. Artino Jr, Measuring achievement goal motivation, mindsets and cognitive load: Validation of three instruments' scores, Med. Educ. **51**, 1061 (2017).

[49] S.-S. Shih, Perfectionism, implicit theories of intelligence, and Taiwanese eighth-grade students' academic engagement, J. Educ. Res. **104**, 131 (2011).

[50] S. J. Troche and A. Kunz, The factorial structure and construct validity of a German translation of Dweck's implicit theories of intelligence scale under consideration of the wording effect, Psychol. Test Assess. Model. **62**, 386 (2020), https://www.proquest.com/scholarly-journals/factorial-structure-construct-validity-german/docview/2450655754/se-2.

[51] K. De Castella and D. Byrne, My intelligence may be more malleable than yours: The revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement, Eur. J. Psychol. Educ. **30**, 245 (2015).

[52] C. M. Ganley, C. E. George, J. R. Cimpian, and M. B. Makowski, Gender equity in college majors: Looking beyond the STEM/non-STEM dichotomy for answers regarding female participation, Am. Educ. Res. J. **55**, 453 (2018).

[53] A. Deiglmayr, E. Stern, and R. Schubert, Beliefs in "brilliance" and belonging uncertainty in male and female STEM students, Front. Psychol. **10**, 1114 (2019).

[54] S. Upson and L. F. Friedman, Where are all the female geniuses?, Sci. Am. Mind **23**, 63 (2012).

[55] M. M. Bleeker and J. E. Jacobs, Achievement in math and science: Do mothers' beliefs matter 12 years later?, J. Educ. Psychol. **96**, 97 (2004).

[56] J. E. Jacobs and J. S. Eccles, The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions, J. Pers. Soc. Psychol. **63**, 932 (1992).

[57] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Gendered patterns in the construction of physics identity from motivational factors, Phys. Rev. Phys. Educ. Res. **15**, 020119 (2019).

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

[59] A. L. Traxler, X. C. Cid, J. Blue, and R. Barthelemy, Enriching gender in physics education research: A binary past and a complex future, Phys. Rev. Phys. Educ. Res. **12**, 020114 (2016).

[60] B. Van Dusen and J. Nissen, Equity in college physics student learning: A critical quantitative intersectionality investigation, J. Res. Sci. Teach. **57**, 33 (2020).

[61] 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).

[62] Z. Hazari, R. H. Tai, and P. M. Sadler, Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors, Sci. Educ. **91**, 847 (2007).

[63] P. Vincent-Ruz, K. Binning, C. Schunn, and J. Grabowski, The effect of math SAT on women's chemistry competency beliefs, Chem. Educ. Res. Pract. **19**, 342 (2018).

[64] The College Board and ACT, Guide to the 2018 ACT/SAT Concordance (2018), https://www.act.org/content/dam/act/unsecured/documents/ACT-SAT-Concordance-Information.pdf.

[65] R. B. Kline, *Principles and Practice of Structural Equation Modeling* (Guilford Press, New York, 2016), 4th ed.

[66] R. C. MacCallum, M. W. Browne, and H. M. Sugawara, Power analysis and determination of sample size for covariance structure modeling., Psychol. Methods **1**, 130 (1996).

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

[68] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (L. Erlbaum Associates, Hillsdale, NJ, 1988).

[69] StataCorp., Stata Statistical Software: Release 17, StataCorp LLC., College Station, TX (2021).

[70] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.19.020135 for additional models predicting final course grade using either pre or post mindset material, as well as measurement invariance testing between men and women.

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

[72] Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach, and C. Singh, Damage caused by women's lower self-efficacy on physics learning, Phys. Rev. Phys. Educ. Res. **16**, 010118 (2020).

[73] Y. Li and C. Singh, Do female and male students' physics motivational beliefs change in a two-semester introductory physics course sequence?, Phys. Rev. Phys. Educ. Res. **18**, 010142 (2022).

[74] S. Cwik and C. Singh, Role of inclusiveness of learning environment in predicting students' outcomes in courses in which women are not underrepresented, J. Higher Educ. Theory Pract. **22**, 176 (2022).

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

[76] 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).

[77] G. M. Walton and G. L. Cohen, A question of belonging: Race, social fit, and achievement, J. Pers. Soc. Psychol. **92**, 82 (2007).

[78] E. A. Canning, K. Muenks, D. J. Green, and M. C. Murphy, STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes, Sci. Adv. **5**, 4734 (2019).

[79] J. LaCosse, M. C. Murphy, J. A. Garcia, and S. Zirkel, The role of STEM professors' mindset beliefs on students' anticipated psychological experiences and course interest, J. Educ. Psychol. **113**, 949 (2021).

[80] C. M. Mueller and C. S. Dweck, Praise for intelligence can undermine children's motivation and performance, J. Pers. Soc. Psychol. **75**, 33 (1998).

[81] A. Rattan, C. Good, and C. S. Dweck, It's OK–not everyone can be good at math: Instructors with an entity theory comfort (and demotivate) students, J. Exp. Soc. Psychol. **48**, 731 (2012).

[82] S. Cwik and C. Singh, Framework for and review of research on assessing and improving equity and inclusion in undergraduate physics learning environments, in *International Handbook of Physics Education Research: Special Topics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing, Melville, NY, 2023), https://doi.org/10.1063/9780735425514_002.

[83] A. Malespina and C. Singh, Gender gaps in grades versus grade penalties: Why grade anomalies may be more detrimental for women aspiring for careers in biological sciences, Int. J. STEM Educ. **10**, 13 (2023).

[84] K. Whitcomb, D. Doucette, and C. Singh, Comparing major declaration, attrition, migration, and completion in physics with other STEM disciplines: A sign of inequitable physics culture? J. Higher Educ. Theory Pract. **22**, 63 (2022).

[85] B. J. Yik, J. R. Raker, N. Apkarian, M. Stains, C. Henderson, M. H. Dancy, and E. Johnson, Evaluating the impact of malleable factors on percent time lecturing in gateway chemistry, mathematics, and physics courses, Int. J. STEM Educ. **9**, 15 (2022).