RESEARCH   Open Access

Variation in which key motivational and academic resources relate to academic performance disparities across introductory college courses

Lorraine Blatt[*], Christian D. Schunn, Elizabeth Votruba-Drzal and Benjamin Margolin Rottman

Abstract

Background: Differences in post-secondary academic outcomes along dimensions of gender, race/ethnicity, and socioeconomic status are a major concern. Few studies have considered differences in patterns of academic outcomes and underlying mechanisms driving disparities across different STEM disciplines. Using data from about 4000 undergraduates in introductory STEM courses at a large, urban university in the eastern United States, this study examines how differences in course grades by gender, race/ethnicity, and parent education vary in introductory chemistry, physics, and psychology courses. In addition, structural equation modeling techniques examine whether academic resources and discipline-specific motivational attitudes are important mediators of demographic differences in course grades.

Results: This study finds that women have higher course grades than men on average in psychology, and men have marginally higher grades than women in physics. In addition, students whose race/ethnicity is represented or overrepresented in these courses (students who are White and or Asian) have higher course grades in chemistry and physics and marginally higher grades in psychology on average compared with underrepresented students (who are Black, Latinx, Native American, Pacific Islander, and or other racial/ethnic backgrounds). Further, first-generation college students have lower course grades in physics and psychology on average than students with a college-educated parent. The largest average differences in course performance are about half a full letter grade (e.g., the difference between a B and an A−). This study also finds that some demographic differences in physics and chemistry performance are linked to math resources whereas some disparities in psychology are more related to verbal resources. In addition, the results suggest discipline-specific self-efficacy is a motivational attitude associated with course performance in chemistry, physics, and psychology, while discipline-specific interest is only relevant in chemistry.

(Continued on next page)

* Correspondence: lrb66@pitt.edu
Department of Psychology and the Learning Research and Development
Center, University of Pittsburgh, 3939 O'Hara St., Pittsburgh, PA 15213, USA

(Continued from previous page)

Conclusions: Overall, the findings emphasize that there are demographic differences in post-secondary course performance on average, and academic resources and motivational attitudes help explain these differences. Importantly, the specific findings differ across chemistry, physics, and psychology. Understanding these pathways and how they are similar and different across disciplines within STEM is crucial for developing interventions aimed at attenuating disparities in post-secondary academic outcomes.

Keywords: Post-secondary education, Achievement disparities, Academic resources, Academic motivational attitudes, STEM

Introduction

Policymakers, researchers, and practitioners at all levels are concerned about issues of equity in education. When it comes to post-secondary education, there is great attention focused on striving for equity in students' academic outcomes by gender, race/ethnicity, and parents' educational attainment. In particular, women, underrepresented racial/ethnic groups, and first-generation college students often receive lower course grades in large introductory science classes (Boyer Commission, 1998; Johnson, 2007; Salehi et al., 2019). However, there is considerable variation across universities and disciplines in the size of differences in academic outcomes related to demographic factors (Matz et al., 2017). For example, women sometimes outperform men in certain introductory courses (Matz et al., 2017) and some universities implement strategies that better support first-generation college-goers to mitigate disparities (Page et al., 2019). Understanding the causes of specific demographic effects in academic outcomes in a particular context is critical to designing interventions that will improve outcomes.

Demographic disparities in course outcomes are often attributed to differences in factors such as discipline-specific motivational attitudes and general academic resources, including mathematical or verbal ability, that are present before enrollment in higher education (Betancur et al., 2018; Nissen & Shemwell, 2016; Sadler & Tai, 2001; Vincent-Ruz et al., 2018; Wang & Degol, 2017). For example, on average, women score lower on standardized math assessments and also tend to have lower self-efficacy beliefs about their abilities in physical sciences which disrupt studying and test-taking behaviors, and thereby course performance in chemistry and physics classes (Marshman et al., 2018a; Vincent-Ruz et al., 2018).

It is not currently clear why demographic disparities in post-secondary academic outcomes vary by academic discipline. Prior studies tend to focus on specific disciplines in isolation (e.g., only physics or only chemistry; Adams et al., 2008; González & Paoloni, 2015; Hazari et al., 2007). Furthermore, many studies focus on specific academic resources (e.g., only high school grade point average (GPA); Allensworth, & Clark, 2020) or motivational factors (e.g., only interest or only self-efficacy; Adams et al., 2008;

González & Paoloni, 2015), making it difficult to compare effects across studies.

Theoretically, there are a number of reasons to expect variation by course in demographic disparities. For example, academic courses across disciplines have distinct populations of learners (National Science Board, 2018) who differ in prior experiences and in motivational attitudes and academic resources. For example, calculus-based courses such as some physics and economics classes will only enroll students who have previously taken calculus, which is not a universally available high school course. Second, because of discipline-specific stereotypes, there may be different patterns in gender disparities in self-efficacy, for example, between physics and chemistry (Whitcomb et al., 2020). Learning challenges also vary across disciplines (e.g., the amount of mathematical problem-solving in physics versus the volume of verbal information to learn in biology), which likely influences the academic resources that enhance course performance. Disciplines also differ in the sources of support they provide to enhance learning and counteract differences in prior experience (e.g., online home practice systems are most commonly available for mathematics and natural sciences; Magalhães et al., 2020). Thus, there is likely variability in how motivational attitudes and academic resources shape academic performance across disciplines that gives rise to heterogeneous disparities in academic performance.

However, the population of students taking introductory courses does partially overlap across disciplines (e.g., introductory STEM courses tend to draw students who have the interest and prior experiences in STEM; Le et al., 2014). Further, when introductory courses share common instructional formats (e.g., large lectures, grades predominantly based on in-class exams, and a lack of demographic representation in instructors), we can expect similar challenges for students. As a result, systematic work using parallel methodologies across disciplines is necessary to begin to understand similarities and differences in the etiological underpinnings of disparities in academic outcomes. In this study, we present a systematic investigation across introductory chemistry, physics, and psychology courses at the University of

Pittsburgh. In doing so, we aim to identify both common and unique patterns across disciplines in disparities and in linkages between academic resources, motivational attitudes, and academic performance.

Academic resources for introductory college courses

One main goal of the K-12 educational system is to prepare students for college (Conley, 2007), which typically involves developing students' independent academic skills, verbal ability, and mathematical ability (often measured by GPAs and Scholastic Aptitude Test (SAT) scores). As many studies show, because of inequities in opportunities and experiences during K-12, these academic resources (reflected in GPAs and SAT scores) are not equitably obtained by students across gender, race/ethnicity, and socioeconomic status (e.g., Luschei, & Jeong, 2018). This is due to large disparities in access to high-quality core educational experiences and sizeable differences in access to optional educational experiences such as elective coursework, afterschool/extracurricular programming, and at-home learning resources (Betancur et al., 2018; Sen & Wasow, 2016).

Despite concerns about variation in grading standards across high schools, high school GPA remains one of the strongest predictors of college performance (Allensworth & Clark, 2020). High school GPA is often conceptualized as a proxy for general academic work skills, such as keeping track of required tasks, completing classwork and homework, and studying for exams (Geiser & Santelices, 2007; Sawyer, 2013). In other words, high school GPA reflects whether students are good at the general task of "doing school." Large introductory college courses can be especially challenging for students in terms of these general academic work skills because large course sizes translate to little direct oversight of student task completion (Pascarella & Terenzini, 2005). In addition, such courses include a broad range of topics with large weight placed on exam performance, which often requires intensive studying behaviors for successful performance (Putnam et al., 2016).

For many introductory science courses, mathematical skills are an important resource because students must independently interpret graphs and tables, quickly perform mathematical operations when solving problems (e.g., solving for unknowns in systems of equations), and conceptually understand the function of mathematical operations (e.g., the purpose of taking an average or computing a standard deviation; National Research Council, 2002). Indeed, some studies suggest that mathematical ability is as important as prior exposure to disciplinary content in predicting performance in introductory science courses (Sadler & Tai, 2007a).

Verbal ability is also crucial for academic success in introductory college courses. In some disciplines, students must read large textbooks on their own outside of class as a primary source of learning (Conley, 2003; National Survey of Student Engagement, 2006). In addition, when there is a heavy reliance on exams in introductory courses, students' inability to quickly and carefully read long and complex exam questions can be detrimental to exam performance (Conley, 2003).

Discipline-specific motivational attitudes

Research often examines relationships among motivational attitudes and academic performance. There are two broad underlying frameworks commonly used: Expectancy Value Theory (EVT; Wigfield & Eccles, 2000) and Self-Determination Theory (Deci & Ryan, 2012). The two frameworks overlap in several constructs such as intrinsic value (or interest) and self-efficacy (or competence beliefs). Further, motivational attitudes are sometimes framed in very general terms, such as general academic self-efficacy (e.g., Schunk & Pajares, 2002), and sometimes in discipline-specific terms, such as physics self-efficacy (e.g., Marshman et al., 2018a). We focus on discipline-specific motivational attitudes since these are most relevant to predicting performance in particular courses. We also focus on interest and self-efficacy because they predict course performance (e.g., Kalender et al., 2020) and can be addressed through changing course structures and early course interventions (e.g., Nissen & Shemwell, 2016); in other words, they are important and amenable to change.

Interest (also called intrinsic value) can be important for learning outcomes because it leads students to try to master the content (Eccles, 2005) and promotes deeper engagement in learning activities and better conceptual learning (Vansteenkiste et al., 2006). At the university level, student discipline-specific interest predicts academic performance above and beyond academic resource differences and other attitudinal differences (Hsieh, 2014). More enduring interest in a discipline builds from positive experiences in one situation (Hidi & Renninger, 2006). Instructors can shape undergraduates' discipline-specific interest by highlighting the value of the course for larger goals (Hulleman et al., 2010), supporting student autonomy (González & Paoloni, 2015), and extending project-based or rich lab experiences (Hazari et al., 2007).

Self-efficacy (e.g., beliefs about one's abilities to successfully complete tasks) consistently relates to academic outcomes, even when controlling for actual abilities (Honicke & Broadbent, 2016; Lawson et al., 2007). Evidence and theory suggest that students with low self-efficacy avoid studying because they do not think it will result in success (Talsma et al., 2018). Low self-efficacy can also leave students especially vulnerable to stereotype threat (Steele & Aronson, 1995) and worry while taking

exams (Marchand & Taasoobshirazi, 2013; Taasoobshirazi et al., 2019). Processes related to self-efficacy can consume students' working memory and infringe on the mental resources they need for taking exams (Beilock et al., 2004). Discipline-specific self-efficacy predicts undergraduate academic performance to different degrees in physics (Kalender et al., 2020), chemistry (Vincent-Ruz et al., 2018), biology (Lawson et al., 2007), and psychology (Komarraju & Nadler, 2013). Discipline-specific self-efficacy at the university level can be targeted through instructors' beliefs (the beliefs that instructors have about their students influence the beliefs that students have about themselves; e.g., Canning et al., 2019) and collaborative learning structures (Fencl & Scheel, 2005).

Demographic differences in academic resources and discipline-specific motivational attitudes

Colleges in the United States typically use a combination of high school GPA, SAT Verbal, and SAT Math (or similar assessments) as primary determinants of college admissions because they predict academic success (Sawyer, 2013). Unfortunately, such procedures tend to produce inequities in admissions because of differential prior learning opportunities in and out of school. For example, women, underrepresented racial/ethnic groups, and students whose parents did not attend college are less likely to experience opportunities to engage in advanced STEM coursework in high school (Robinson, 2003; Tyson et al., 2007), which negatively impacts mathematical ability compared with their peers who did have such opportunities (Kurban & Cabrera, 2019). In addition, there are large socioeconomic differences in many aspects of K-12 school quality, including the availability of summer and after school learning opportunities (Putnam, 2016), which can impact mathematical, verbal, and general study skills (Bernal et al., 2016; Hanushek & Woessmann, 2017). However, there are also relative strengths, for example, women are generally more successful than men in high school academics (Fortin et al., 2015) and are more likely to enroll in advanced writing courses in high school (College Board, 2018a). Additionally, very selective universities employ resource-based selection procedures that reduce demographic disparities in academic performance in college courses (e.g., only students with high SAT Math scores are admitted to the university or the STEM major). Thus, demographic differences in academic resources in the general pool of high school graduates do not always replicate among college attendees, particularly in programs with a strong focus on academic preparation.

There can also be demographic differences among undergraduates within introductory STEM courses in discipline-specific motivational attitudes, particularly in discipline-specific interest (Marshman et al., 2018b) and discipline-specific self-efficacy (Marshman et al., 2018a; Vincent-Ruz et al., 2018). These may arise from messages students receive from media, teachers, family, and peers that shape beliefs about whether they can be successful in academics. For example, there are negative stereotypes about women and minoritized students in STEM (Moss-Racusin et al., 2012; Seymour & Hewitt, 1997) which can influence the messages students to receive and thereby change course performance (Hazari et al., 2007). Further, differential exposure to certain topics (e.g., in advanced coursework) in high school can affect students' interest in the topics (Osborne et al., 2003). At the same time, students self-select into majors in college, which influences the introductory courses they take. This could result in a reduction or absence of demographic differences in motivational attitudes within those courses. For example, women enrolled in chemistry courses may have a similarly high interest in chemistry as men enrolled in those courses.

It is important to note that academic resources and motivational attitudes are not independent. Throughout high school, grades within particular courses provide feedback to students, which then shapes their discipline-specific self-efficacy (Lopez & Lent, 1992). Additionally, students are keenly aware of their performance on tests like the SAT, which influences their discipline-specific interests and self-efficacy (Vincent-Ruz et al., 2018). Thus, demographic differences in academic resources can also produce demographic differences in motivational attitudes.

Disciplinary variation in important motivational attitudes and academic resources

From a theoretical perspective, a similar pool of general academic resources and motivational attitudes might matter for all large introductory courses. However, it is likely that there is variability in which academic resources and motivational attitudes matter for which disciplines. First, for self-selection reasons, demographic differences in ability might vary by discipline. For example, if only high math ability students select physical sciences majors, then demographic differences in math ability will be truncated and math skills will have little predictive validity in explaining variability in performance in the physical sciences, but could still explain differences in other disciplines. Discipline-specific stereotypes could give rise to a similar result. For example, women are commonly described as weak in STEM but strong in social sciences (Wang et al., 2013). Disciplines also vary in the extent to which innate talent is thought to be critical, and this is correlated with participation by gender (Leslie et al., 2015). Stereotypes may give rise to limited variability in self-efficacy within disciplines, with higher levels of self-efficacy among women in psychology and lower levels among women in physics and chemistry, for example.

Second, introductory courses may vary substantially in the extent to which they depend upon different academic resources. Most obvious is the relative use of verbal versus mathematical skills, with some disciplines depending more on verbal skills and others depending more on mathematical skills. For example, past studies find a strong role of SAT Verbal in psychology courses (Betancur et al., 2019) and a stronger role of SAT Math in chemistry (Vincent-Ruz et al., 2018) and physics courses (Kalender et al., 2020). With regard to motivation, as previously mentioned, many studies across disciplines find that self-efficacy is an important performance predictor; less obvious is the relative reliance on interest. When courses require students to do a lot of independent learning (e.g., reading in psychology versus required/graded homework completion in physics and chemistry), the relative role of personal interest may be larger.

Theoretical framework

Figure 1 presents a theoretical model of the processes described in the introduction; that demographics can affect academic resources and motivational attitudes (because of a variety of structural factors in society and K-12 experiences), which can then affect academic performance. Furthermore, each step may materialize differently across academic disciplines depending on differences in learning environments.

Research questions

1. What is the magnitude of differences in student academic performance related to gender, race/ethnicity, and parent education, in large, lecture-based introductory courses in chemistry, physics, and psychology?
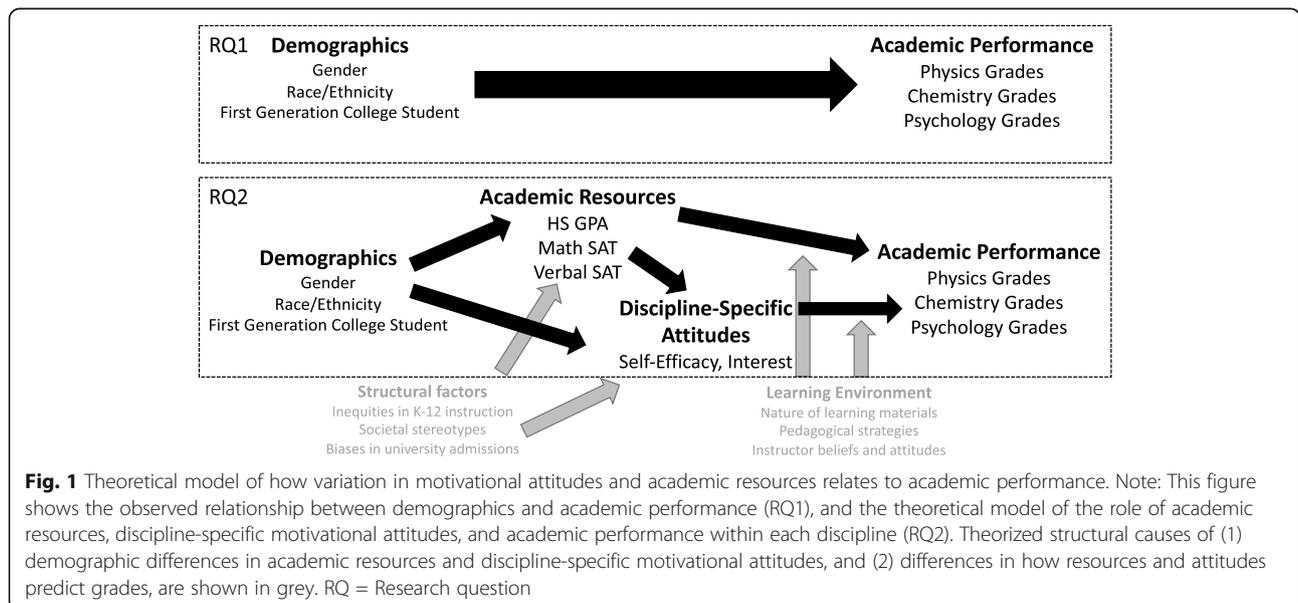2. To what extent do demographic differences in academic resources and discipline-specific motivational attitudes explain differences in academic performance within each of these courses?

We explore these general research questions in a particular set of courses: Chemistry, Physics, and Psychology. The pool of students who choose to enroll at the University of Pittsburgh and register in these specific courses may differ from students who choose to enroll in other universities or register in similar courses at other universities. Thus, the underlying goal of these research questions is not to provide fixed answers by the course that will definitively replicate across contexts. Rather the goal is to show that, even within one university, (1) the size and even direction of demographic differences in course grades can vary by discipline, (2) demographic differences in academic resources and discipline-specific motivational attitudes can also vary by course, rather than being universal statements about student demographic groups, and (3) differences in academic resources and motivational attitudes can provide useful explanations for variation in demographic-based course performance differences.

Method
Overview

This study takes place at the University of Pittsburgh, a large, urban university in the eastern US. In 2018–2019, undergraduate enrollment was 19,330 students and the undergraduate student body was 70 percent White (University of Pittsburgh, 2019). In the fall of 2018, the acceptance rate was 59 percent, the 25th percentile SAT score was 630 for Verbal

Fig. 1 Theoretical model of how variation in motivational attitudes and academic resources relates to academic performance. Note: This figure shows the observed relationship between demographics and academic performance (RQ1), and the theoretical model of the role of academic resources, discipline-specific motivational attitudes, and academic performance within each discipline (RQ2). Theorized structural causes of (1) demographic differences in academic resources and discipline-specific motivational attitudes, and (2) differences in how resources and attitudes predict grades, are shown in grey. RQ = Research question

and 640 for Math, and the 75th percentile SAT score was 700 for Verbal and 730 for Math (Integrated Postsecondary Education Data System, 2019).

This study merges administrative records from the University with course-specific survey data from the University's "Interventions that Matter" project, which aims to understand academic performance disparities and evaluate interventions in undergraduate STEM courses. The motivational attitude measures were collected from students enrolled in General Chemistry I from the fall of 2015 through the fall of 2016, Calculus-based Physics I from the fall of 2015 through the spring of 2017, and Introduction to Psychology from the fall of 2018 through the spring of 2019[1]. Each of these large enrollments, lecture-based courses is the first college-level (non-remedial) introductory course within each of the disciplines, and each is primarily taken by students in their first year at the University. Final grades in all three courses were generally based on multiple midterms and a final exam. However, as is commonly the case across universities, grades in chemistry also involved a separate lab component. grades in both chemistry and physics were also based on weekly homework of a worksheet type, often completed online.

Introductory psychology, chemistry, and (calculus-based) physics courses were selected as cases for study because these large enrollment courses are likely to have substantially different pools of students enrolled. Despite all being predominantly first-year students, enrollees in these introductory courses reflect differences in the types of students who take these courses. The chemistry course includes large numbers of students pursuing medicine or other health careers/life science majors. The calculus-based physics course includes many students from engineering, as well as students intending physics and chemistry majors; students pursuing medicine typically enroll instead in the algebra-based physics course. The psychology course draws students from a wide range of intended majors. In other words, different populations enroll in each of these courses, which may give rise to demographic differences in academic resources and discipline-specific motivational attitudes (e.g., there may be demographic differences in SAT Math in one sample but not in another), and thus grades.

Sample
The samples include 1295, 1102, and 1829 students in chemistry, physics, and psychology respectively. The

same student may be included in more than one sample if they enrolled in more than one of the courses included in this analysis. No student is included in a specific course's sample more than once, and if a student repeated a class, we included data from the earliest time a student was enrolled in the class and completed the survey measures. Overall, the study analyzed 4226 final course grades across the three courses with data from 3625 unique students. Table 1 includes descriptive information on the sample for each course.

Measures
Demographic characteristics
Students' demographic information came from administrative data at the University of Pittsburgh. The main demographic variables of interest are student gender, race/ethnicity, and parent education. Student gender is represented with a dummy variable that indicates whether the student is female or male (reference group)—although gender is a complex, multi-dimensional construct (Hyde et al., 2019), the existing institutional data is only represented in binary terms.

Race/ethnicity is reflected in a dummy variable indicating if the student is a member of an underrepresented racial/ethnic group. If the student is of only White and or Asian descent, they are considered represented (these students serve as the reference group). If the student is of Black, Latinx, Native American, Pacific Islander, and or other descent, they are considered underrepresented. The University of Pittsburgh's undergraduate enrollment is over 70 percent White and over 10 percent Asian (University of Pittsburgh, 2019); Table 1 demonstrates that the introductory courses examined in this study are even more racially/ethnically homogenous. Thus, students of Black, Latinx, Native American, Pacific Islander, and or other descent are disproportionately underrepresented, not only at the University but even more so in these courses. While there is a great deal of racial/ethnic heterogeneity within these dichotomized groups, there were not adequate sample sizes to consider more refined categories of race/ethnicity. Further, we are using students' race/ethnicity as a proxy for measuring the ramifications of being part of a minoritized racial/ethnic group that is underrepresented in a classroom environment. We recognize that using a race/ethnicity dummy variable is an imperfect proxy for these processes (Brown et al., 2019). Unfortunately, we do not have more refined student-level data on specific processes related to the experience of being underrepresented (e.g., students' reports of belonging and experiences of racial aggressions).

First-generation college student status is represented with a dummy variable as well, in this case indicating whether students do not have at least one parent with at least a bachelor's degree. This variable was created from

---

[1]This effort was situated in a larger project that involved different year ranges of participation across disciplines and gradual revisions to surveys that made them more comparable. We used every semester with available data using the most comparable surveys to maximize sample sizes. There were not large changes in population characteristics at the University of Pittsburgh from 2015–2019, so this variation is unlikely to influence the results.

Table 1 Sample characterization in terms of demographics, academic resources, discipline-specific attitudes, and course performance

Variable	Chemistry Mean or %	SD	Physics Mean or %	SD	Psychology Mean or %	SD
Demographics						
Age (years)	18.68	1.16	19.02	1.82	19.17	1.36
Gender						
Female	61%		32%		57%	
Male	39%		68%		43%	
Race/Ethnicity						
Asian	18%		11%		18%	
Black	5%		4%		6%	
Latinx	3%		3%		6%	
Multi-racial	5%		5%		4%	
Other	0.3%		0.1%		0.1%	
White	69%		77%		67%	
Parent Education						
High school or below	10%		8%		10%	
College or higher	90%		92%		90%	
Academic resources						
High School GPA	4.08	0.40	4.07	0.44	4.00	0.48
SAT Verbal/100	6.72	0.77	6.80	0.72	6.59	0.74
SAT Math/100	6.70	0.73	6.97	0.64	6.59	0.75
Discipline attitudes						
Interest	2.88	0.63	3.19	0.52	3.01	0.51
Self-Efficacy	2.74	0.56	2.85	0.45	3.08	0.39
Course performance						
Course grade	2.87	0.91	2.43	1.03	3.53	0.71
N	1,295		1,102		1,829	
Participation rate	95%		60%		79%	
Number of class sections	6		9		9	
Number of instructors	4		5		5	
Average class size	191		122		203	

Note: Sample statistics reflect imputed data. Native American and Pacific Islander students are included with "Other" due to small sample sizes

University administrative data, which originally received the data from the Free Application for Federal Student Aid (FAFSA).

Academic resources

Key student academic resources were measured with three constructs commonly used to predict academic performance, obtained by the University of Pittsburgh as part of the admissions process: high school grade-point averages (GPA), and the Mathematics and Verbal Scholastic Aptitude Test (SAT) scores. About one percent of high school GPAs were either missing or over 5.0. GPAs over 5.0 were removed from analyses since these students' GPAs were not calculated on the usual five-point weighted-by-Advanced-Placement-courses scale,[2] making their high school GPAs not comparable to the majority of the high school GPAs in the sample. Some students only had American College Testing (ACT) scores and no SAT scores. In those cases, the ACT English and math scores were standardized to the SAT scale using the College Board (2018b) ACT/SAT concordance tables to create one SAT Verbal score and one SAT Math score (each with a maximum score of 800 and US-wide means of slightly above 500) for each student. As previously mentioned, these indicators serve as proxies for academic resources and access to academic opportunities more broadly (Bernal et al., 2016; Hanushek & Woessmann, 2017; Putnam, 2016).

Discipline-specific motivational attitudes

Students' interest and self-efficacy in the discipline of each course were measured using self-reported survey responses collected at the start of the course (just after the end of the add-drop period and before any exam feedback was provided). The scales were developed by adapting existing scales in the literature to each discipline and for undergraduate populations as needed. Cognitive interviews were conducted to ensure items were interpreted as intended. These interviews involved 5–10 students per discipline, drawn from the courses being studied at varying times during the semester, and selected to vary in gender and course performance. Exploratory factor analyses with a broader set of constructs (e.g., intelligence mindset, extrinsic value) were conducted to establish a single-factor structure within each scale and discriminant validity between scales. Item response theory (IRT) analyses were conducted to ensure no survey items had differential scale discriminability by gender, race/ethnicity, or first-generation status.

Discipline-specific interest scales capture how fascinated a student is by each course's subject matter. The discipline-specific self-efficacy scale captures how confident and capable a student feels about their ability to do well on a variety of tasks involving content from

---

[2]US high schools use a 0–4 GPA scale with 0 representing a failure and 4 representing the top grade, an A. However, when high school students take Advance Placement (AP) courses that provide them early access to university-level coursework, grades in these courses are inflated to a 0 to 5 scale. The combination of grades across courses types is called a weighted GPA. We did not have access to students' unweighted GPAs; there is evidence that weighted GPAs can be a better predictor than unweighted GPAs (Sadler & Tai, 2007b), although there is some counter-evidence as well (Warne et al., 2014).

that particular discipline. Table 2 includes the full list of survey items used to measure discipline-specific interest and self-efficacy in chemistry, physics, and psychology as well as the corresponding Cronbach's alphas. The survey items were collected from existing scales and surveys and were not originally created for the purpose of this study; hence, there are minor differences in the items across disciplines. The source datasets also contain many

Table 2 Survey questions used for discipline-specific interest and self-efficacy variables

Chemistry interest (α = 0.77)	In general I find chemistry topics	Very boring (1) Boring (2) Interesting (3) Very Interesting (4)
	I want to know everything I can about chemistry.	NO! (1) no (2) yes (3) YES! (4)
Chemistry self-efficacy (α = 0.85)	Chemistry intimidates me.	Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4) IDU (5)
	I have trouble understanding anything based on chemistry.	Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4) IDU (5)
	I have always had difficulty understanding arguments that require chemical knowledge.	Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4) IDU (5)
	I can usually figure out a way to solve chemistry problems.	Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4) IDU (5)
	After I study a topic in chemistry and feel that I understand it, I have difficulty solving problems on the same topic.	Strongly Disagree (1) Disagree (2) Agree (3) Strongly Agree (4) IDU (5)
Source: (Adams et al., 2008; Bauer, 2005; Vincent-Ruz et al., 2018)		
Physics interest (α = 0.63)	I wonder about how nature works.	Never(1) Once a month (2) Once a week (3) Every day (4)
	In general, I find physics	Very boring (1) Boring (2) Interesting (3) Very Interesting (4)
	I want to know everything I can about physics.	NO! (1) no (2) yes (3) YES! (4)
Physics self-efficacy (α = 0.75)	I can complete the physics activities I get in a lab class.	Rarely (1) Half the time (2) Most of the time (3) All the time (4)
	If I went to a museum, I could figure out what is being shown about physics in	None of it (1) A few areas (2) Most areas (3) All areas (4)
	I am often able to help my classmates with physics in the laboratory or in recitation.	NO! (1) no (2) yes (3) YES! (4)
	I get a sinking feeling when I think of trying to tackle difficult physics problems.	NO! (1) no (2) yes (3) YES! (4)
	If I wanted to, I could be good at doing physics research.	NO! (1) no (2) yes (3) YES! (4)
	If I study, I will do well on a physics test.	NO! (1) no (2) yes (3) YES! (4)
Source: (Marshman et al., 2018b)		
Psychology interest (α = 0.82)	I wonder about how psychology works.	Never(1) Once a month (2) Once a week (3) Every day (4)
	In general, I find psychology	Very boring (1) Boring (2) Interesting (3) Very Interesting (4)
	I want to know everything I can about psychology.	NO! (1) no (2) yes (3) YES! (4)
	I am curious about recent discoveries in psychology.	NO! (1) no (2) yes (3) YES! (4)
	I want to know about the current research that psychologists are doing.	NO! (1) no (2) yes (3) YES! (4)
Psychology self-efficacy (α = 0.77)	If I study, I will do well on a psychology test.	NO! (1) no (2) yes (3) YES! (4)
	If I wanted to, I could be good at psychology research.	NO! (1) no (2) yes (3) YES! (4)
	I understand concepts I have studied in psychology.	NO! (1) no (2) yes (3) YES! (4)
	I am able to help my classmates with psychology in the laboratory or in recitation.	NO! (1) no (2) yes (3) YES! (4)
	If I encounter a setback in a psychology exam, I can overcome it.	NO! (1) no (2) yes (3) YES! (4)
	When I think about getting ready for a psychology exam, I get a sinking feeling.	NO! (1) no (2) yes (3) YES! (4)
Source: (Marshman et al., 2018b; Vincent-Ruz & Schunn, 2017)		

Note: Only responses 1–4 were included in analyses of composites

other survey constructs (different constructs within each discipline), so the number of survey items per construct were minimized to avoid poor participation due to survey fatigue.

Course performance

The main outcomes of interest in this study are students' course grades in introductory chemistry, physics, and psychology. Course grades are measured on a continuous four-point scale with a 4.0 being an A or A+ and a 0 equaling an F: A = 4, B = 3, C = 2, D = 1, and +/− add or subtract 0.25 (e.g., a B− is a 2.75 and a B+ is 3.25).

Analytic approach

The first research question (Fig. 1) involved identifying the magnitude of differences in students' course grades based on gender, race/ethnicity, and parent education, in large, lecture-based introductory courses in psychology, chemistry, and physics. This analysis used multiple regression in Stata 15.0 to predict course grades in chemistry, physics, and psychology with just demographic characteristics.

For the second research question (Fig. 1), this study used structural equation modeling (SEM) in Stata 15.0. We chose an SEM framework in order to test several direct and indirect effects simultaneously, which is a more precise method than multiple regression (Li, 2011). The SEMs tested academic resources and discipline-specific motivational attitudes as mediators of the relationships of gender, race/ethnicity, and parent education on course grades. The models also tested pathways from academic resources to motivational attitudes. The statistical significance of indirect effects was tested simultaneously in a SEM framework (Keith, 2006) using a Sobel (1987) test with 200 bootstrapped standard errors. After initially testing all possible direct and indirect effects, insignificant ($p > .05$) pathways and insignificant correlated errors of mediators were trimmed to improve the goodness of fit of the model. Model fit was assessed using the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) to compare the model with a baseline model and the Root Mean Squared Error of Approximation (RMSEA) to examine parsimony (Hu & Bentler, 1999). A good fit is achieved if the CFI and TLI values are above .95 and the RMSEA values are below .06 (Hu & Bentler, 1999).

Missing data

Only 72 percent of the chemistry sample, 74 percent of the physics sample, and 69 percent of the psychology sample have complete data for all variables included in the analyses. Most missing data is due to the missing parent education variables for students who did not complete a FAFSA. Given the biases introduced with listwise deletion, regression analyses used imputation methods and SEMs used maximum likelihood with missing values (MLMV; Enders & Bandalos, 2001). For the regression models, missing data were imputed using chained equations (ICE) to create 20 complete datasets in Stata 15.0 (Royston, 2004, 2005). The SEMs with MLMV were also conducted in Stata 15.0. The models were analyzed both with and without imputed data and were consistent; for simplicity, we present only the results with imputed data and include the unimputed results in Online Supplementary Tables 1–6.

Results

Table 3 provides group mean differences in academic resources, discipline-specific motivational attitudes, and course grades, separated by gender, underrepresented racial/ethnic backgrounds, and parent education. Although the University of Pittsburgh is somewhat selective, there are no ceiling effects on the academic resource variables. Additionally, while students self-selected into these courses (by larger academic pathway, by major, or by general interest), there were no ceiling effects on the interest or self-efficacy measures. Thus, differential predictiveness of academic outcomes could not be attributed to a lack of variation across students enrolled in each course. The variation, highlighted by large within-group standard deviations (Table 3), also illustrates that despite the results' focus on average group differences, there are students of all backgrounds with high levels of interest and self-efficacy and high course grades.

Table 4 provides bivariate correlations for all variables included in the SEMs. The three demographic variables were generally independent of one another, but with small variation across courses. The three academic resource variables were moderately correlated, with the largest correlation being between SAT Verbal and SAT Math. Interest and self-efficacy moderately correlated as well, but not so strongly as to prevent the analysis of separate mediation pathways. Academic resources and motivational variables were essentially uncorrelated in psychology, rarely correlated in physics, and consistently but weakly correlated in chemistry. Overall, the correlations among demographic variables, academic resources, and motivational variables varied substantially across disciplines and are probed further in the SEMs.

Demographic performance disparities in introductory courses

In accordance with the first research question, Table 5 shows the results of three regression models with introductory course grades for chemistry, physics, and psychology each regressed on gender, race/ethnicity, and parent education. Associations between gender and introductory grades varied across courses, from the marginally lower

Table 3 Mean and SD for academic resources and discipline-specific motivational attitudes by course and demographic characteristics

	%	High School GPA (range: 0–5)		SAT Verbal (/100) (range: 3.5–8)		SAT Math (/100) (range: 3.6–8)		Discipline-based interest (range: 1–4.5)		Discipline-based self-efficacy (range: 1–4.1)		Course grade (range: 0–4)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Chemistry													
Female	61%	4.13	0.37	6.76	0.75	6.57	0.71	2.82	0.63	2.62	0.57	2.84	0.87
Male	39%	4.02	0.42	6.66	0.79	6.89	0.72	2.98	0.60	2.91	0.49	2.92	0.97
Underrepresented	11%	4.02	0.41	6.59	0.70	6.41	0.73	2.72	0.66	2.63	0.55	2.49	1.03
White and or Asian	89%	4.09	0.39	6.73	0.78	6.73	0.72	2.90	0.62	2.75	0.56	2.91	0.89
First generation	10%	4.06	0.38	6.38	0.82	6.32	0.73	2.89	0.65	2.64	0.54	2.75	0.91
Non-first gen	90%	4.09	0.40	6.76	0.76	6.74	0.72	2.88	0.62	2.75	0.56	2.88	0.93
Physics													
Female	32%	4.18	0.36	6.91	0.74	6.87	0.67	3.05	0.58	2.67	0.46	2.34	0.96
Male	68%	4.02	0.47	6.74	0.71	7.02	0.63	3.25	0.48	2.93	0.42	2.47	1.06
Underrepresented	9%	4.03	0.42	6.72	0.69	6.83	0.61	3.18	0.55	2.80	0.51	2.22	1.07
White and or Asian	91%	4.07	0.44	6.80	0.73	6.99	0.65	3.19	0.52	2.85	0.44	2.45	1.03
First generation	8%	3.93	0.56	6.41	0.71	6.63	0.75	3.28	0.54	2.86	0.46	1.94	1.16
Non-first gen	92%	4.08	0.43	6.83	0.71	7.00	0.63	3.18	0.52	2.85	0.45	2.47	1.01
Psychology													
Female	57%	4.05	0.42	6.61	0.75	6.46	0.73	3.08	0.50	3.09	0.37	3.59	0.63
Male	43%	3.94	0.53	6.57	0.73	6.76	0.73	2.91	0.50	3.07	0.41	3.46	0.81
Underrepresented	14%	3.86	0.46	6.45	0.74	6.30	0.69	3.03	0.51	3.09	0.40	3.45	0.69
White and or Asian	86%	4.03	0.47	6.62	0.74	6.63	0.74	3.01	0.51	3.08	0.39	3.54	0.72
First generation	10%	3.88	0.53	6.17	0.76	6.28	0.76	3.07	0.57	3.10	0.42	3.35	0.89
Non-first gen	90%	4.02	0.47	6.64	0.72	6.62	0.74	3.00	0.50	3.08	0.39	3.55	0.69

Note: Group ranges and mean statistics reflect imputed data

performance of female students on average (.06 standard deviation units (SD), $p = .065$) in physics, to no significant difference in chemistry, to significantly higher performance of women in psychology (.09 SD, $p < .001$). Students from underrepresented racial/ethnic groups performed marginally or significantly worse than students from only White and or Asian backgrounds across all of the introductory courses, but the size of the relationship varied by discipline, with differences of .14 SD ($p < .001$) in chemistry, .06 SD ($p = .039$) in physics, and only .04 SD ($p = .064$) in psychology. Finally, students with at least one college-educated parent performed better than first-generation college students in physics (.14 SD, $p < .001$) and psychology (.09 SD, $p < .001$), whereas the difference was small and non-significant in chemistry. Thus, all demographic disparities varied in significance, effect size, and or direction across disciplines.

Mediators of demographic performance disparities in introductory college courses

In accordance with the second research question, Figs. 2–4 display the direct effects of pathways with significant indirect effects relevant to course grade outcomes in the SEMs across the chemistry, physics, and psychology samples. Figure 5 shows the direct effects of full SEMs. The complete coefficients, standard errors, and p values of every direct and indirect effect estimated in the SEMs are displayed in Appendix Tables 6 and 7. Overall, the final models for each course provided good fits to the data[3]: chemistry RMSEA <.01 (90% CI .00, .03), CFI > .99, and TLI > .99; physics RMSEA = 0.01 (90% CI 0.00, 0.03), CFI > .99, and TLI > .99, and psychology RMSEA = 0.02 (90% CI 0.00, 0.03), CFI > .99, and TLI = .99. As a robustness check to examine whether there was strong support for using different final models for each discipline, each model specification was run with each course sample (e.g., the psychology model specification was applied to the chemistry sample). In all of these instances, the models demonstrated worse fit with the SEM specifications for other disciplines. This supports that each model specification is distinctly predictive of its course sample. Details of

---

[3]We report p values rounded to three digits and fit statistics rounded to two digits.

Table 4 Bivariate correlations for each variable analyzed by course

	Course grade	Female	Under-represented	First generation	High school GPA	SAT Verbal	SAT Math	Interest	Self-efficacy
Chemistry									
Course grade	1.00								
Female	−0.04	1.00							
Underrepresented racial/ethnic group	−0.14	0.06	1.00						
First generation	−0.04	0.06	0.04	1.00					
High school GPA	0.36	0.13	−0.06	−0.03	1.00				
SAT Verbal	0.31	0.06	−0.06	−0.15	0.29	1.00			
SAT Math	0.41	−0.22	−0.14	−0.17	0.26	0.53	1.00		
Interest	0.28	−0.13	−0.09	0.01	0.05	0.04	0.12	1.00	
Self-efficacy	0.38	−0.25	−0.07	−0.06	0.16	0.14	0.34	0.55	1.00
Physics									
Course grade	1.00								
Female	−0.06	1.00							
Underrepresented racial/ethnic group	−0.07	0.05	1.00						
First generation	−0.14	0.01	0.01	1.00					
High school GPA	0.32	0.17	−0.02	−0.09	1.00				
SAT Verbal	0.28	0.11	−0.03	−0.16	0.30	1.00			
SAT Math	0.47	−0.11	−0.07	−0.15	0.21	0.42	1.00		
Interest	0.08	−0.18	−0.001	0.05	−0.05	−0.02	0.02	1.00	
Self-efficacy	0.16	−0.27	−0.04	0.01	−0.04	−0.01	0.16	0.50	1.00
Psychology									
Course grade	1.00								
Female	0.09	1.00							
Underrepresented racial/ethnic group	−0.05	0.01	1.00						
First generation	−0.09	0.03	0.04	1.00					
High school GPA	0.38	0.12	−0.12	−0.08	1.00				
SAT Verbal	0.23	0.03	−0.08	−0.19	0.29	1.00			
SAT Math	0.18	−0.20	−0.15	−0.14	0.28	0.42	1.00		
Interest	0.02	0.17	0.01	0.04	0.003	0.07	−0.05	1.00	
Self-efficacy	0.08	0.02	0.01	0.02	0.06	0.08	0.03	0.39	1.00

the fit indices for all model specifications are in Appendix Table 8.

In the following sections, we discuss the mediators between demographic factors and course grades to determine whether the demographic-based grade disparities revealed in the regression analyses in research question one (Table 5) can be explained in terms of the mediation models we ran for research question two. We organize sections based on gender (Fig. 2), race/ethnicity (Fig. 3), and parent education (Fig. 4). Each figure also shows the initial demographic effect on course grades from the first research question (Table 5). Figures 2, 3, and 4 show only the significant mediational pathways (i.e., that

contributed significantly in positive or negative ways towards demographic grade effects). Figure 5 includes all significant pathways in the mediational models. We conclude with the findings of moderation analyses to verify that the tested mediational pathways are appropriate pathways across demographic groups (e.g., whether SAT math predicts physics self-efficacy to an equal extent for women and men).

Gender

Chemistry The SEM results reveal that the non-significant relationship between gender and course

Table 5 Multiple regression results for demographic characteristics predicting course grade in introductory chemistry, physics, and psychology

Predictors	Chemistry			Physics			Psychology		
	Coeff.	SE	Standardized Coeff.	Coeff.	SE	Standardized Coeff.	Coeff.	SE	Standardized Coeff.
Gender									
Female	−0.06	0.05	−0.03	−0.12+	0.07	−0.06	0.14***	0.03	0.09
Race/ethnicity									
Underrepresented	−0.41***	0.08	−0.14	−0.22*	0.11	−0.06	−0.09+	0.05	−0.04
Parent education									
First generation	−0.10	0.10	-0.03	−0.53***	0.14	−0.14	−0.20***	0.06	−0.09
Intercept	2.96	0.04		2.53	0.04		3.49	0.03	
N	1,295			1,102			1,829		

Note. All of the variables in the models are dummy variables, so the coefficients reflect the difference in mean performance between the variable and the reference group

***$p < .001$. **$p < .01$. *$p < .05$. +$p < .1$

performance in the regression analysis for chemistry may be the result of significant and countervailing indirect effects operating through academic resources and motivational attitudes (Fig. 2a; Appendix Table 7). For academic resources, there was a positive indirect effect of high school GPA on the association between gender (women had higher high school GPAs than men on average) and course grade (mediation effect $\beta = .03$, $p < .001$). By contrast, there was a negative indirect effect of SAT Math scores (women had lower SAT Math scores on average) on the association between gender and course grade (mediation effect $\beta = −.05$, $p < .001$). For motivational attitudes, women had lower chemistry interest and self-efficacy on average and there were negative indirect effects of chemistry interest and self-efficacy on the association between gender and chemistry course grades (mediation effects $\beta = −.01$, $p = .004$ and $\beta = −.04$, $p < .001$, respectively).

Similar to the countervailing indirect effects between gender and course performance, there were also countervailing indirect effects of academic resources on the associations between gender and motivational attitudes. There was a positive indirect effect of high school GPA on the association between gender and self-efficacy (mediation effect $\beta = .01$, $p = .001$). By contrast, there were negative indirect effects of SAT Math on the associations between gender and chemistry self-efficacy (mediation effect $\beta = −.06$, $p < .001$) and gender and chemistry interest (mediation effect $\beta = −.02$, $p = .002$).

Physics The patterns in physics were similar to those in chemistry, but with fewer pathways. In physics, the regression found a marginal relationship between gender and course performance, with women receiving lower grades than men. This predicted association is also mediated by countervailing indirect effects through high school GPA, SAT Math, and physics self-efficacy (Fig. 2b, Appendix Table 7). There was a positive indirect effect of high school GPA (women had higher high school GPAs on average) on the association between gender and course grade (mediation effect $\beta = .04$, $p < .001$). By contrast, there was an equally large negative indirect effect of SAT Math scores (women had lower SAT Math scores on average) on the association between gender and course grade (mediation effect $\beta = −.04$, $p = .001$). Additionally, women had lower physics self-efficacy on average, and there was a negative indirect effect of self-efficacy on the association between gender and physics course grades (mediation effect $\beta = −.03$, $p = .001$).
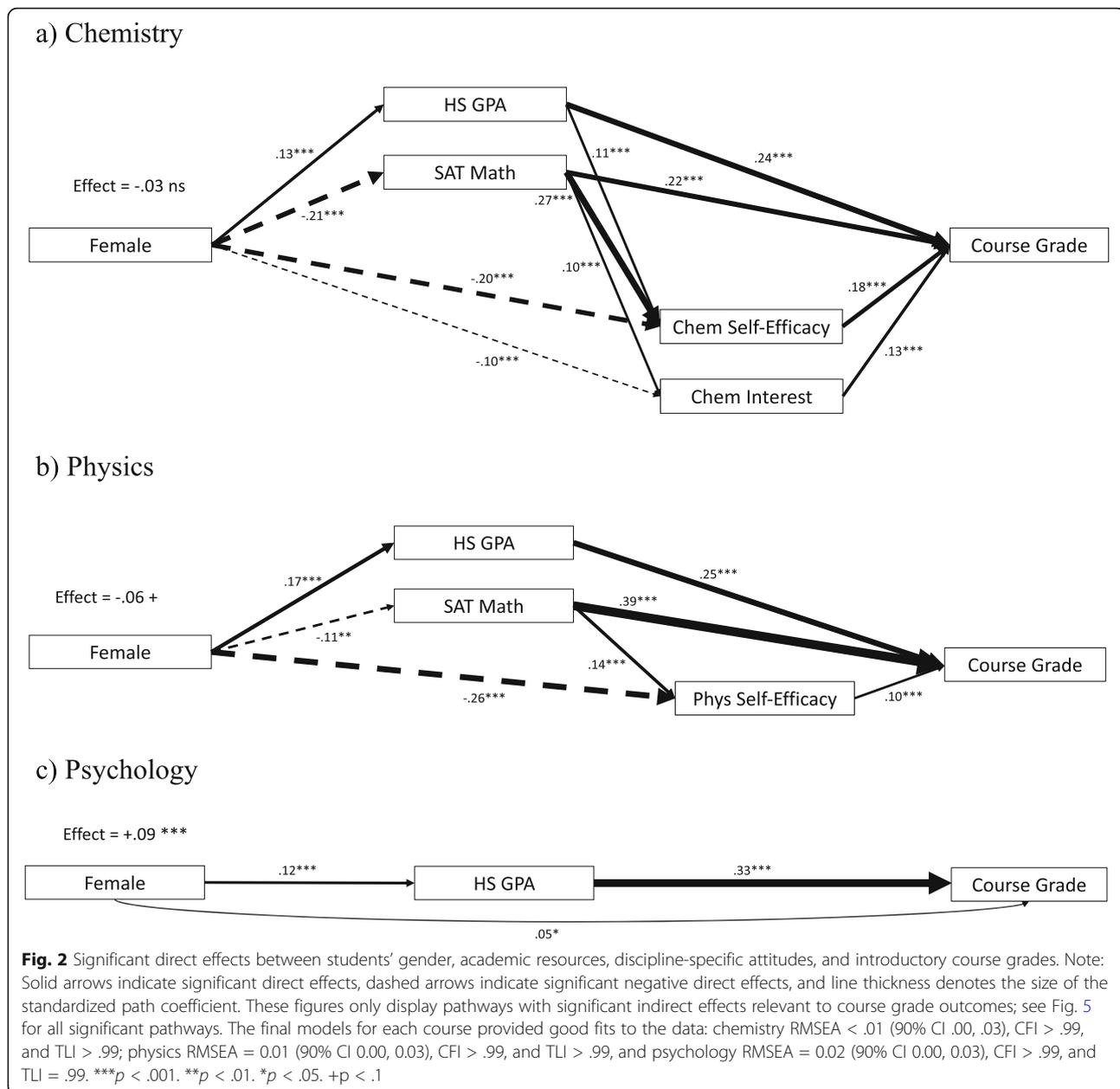
There was only one indirect effect of academic resources on the associations between gender and motivational attitudes: there was a negative indirect effect of SAT Math scores on the association between gender and physics self-efficacy (mediation effect $\beta = −.02$, $p = .004$).

Psychology In psychology, unlike chemistry and physics, women's course grades were significantly higher than men's on average. The SEM results (Fig. 2c; Appendix Table 7) partially explain this disparity with a positive indirect effect of high school GPA (women had higher high school GPAs on average) on the association between gender and course grade (mediation effect $\beta = .04$, $p < .001$).

However, the SEM does not fully explain the estimated association between gender and psychology course grade. There is an enduring direct effect of gender on course grade (direct effect $\beta = .05$, $p = .034$) after accounting for academic resources and motivational attitudes (Fig. 2c; Appendix Table 6).
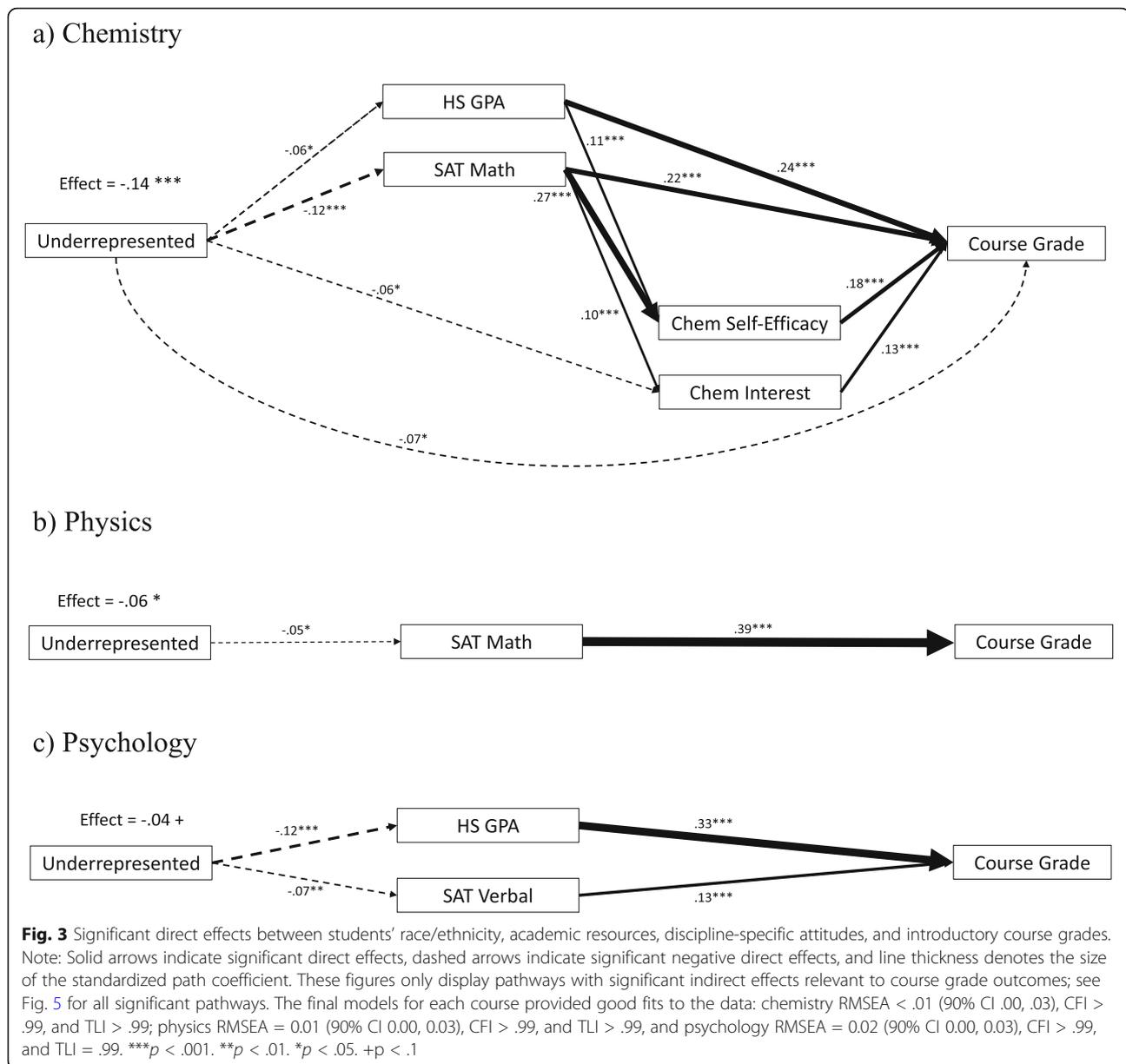
Race/ethnicity

Chemistry The regression results highlight that Black, Latinx, Native American, Pacific Islander, and or

Fig. 2 Significant direct effects between students' gender, academic resources, discipline-specific attitudes, and introductory course grades. Note: Solid arrows indicate significant direct effects, dashed arrows indicate significant negative direct effects, and line thickness denotes the size of the standardized path coefficient. These figures only display pathways with significant indirect effects relevant to course grade outcomes; see Fig. 5 for all significant pathways. The final models for each course provided good fits to the data: chemistry RMSEA < .01 (90% CI .00, .03), CFI > .99, and TLI > .99; physics RMSEA = 0.01 (90% CI 0.00, 0.03), CFI > .99, and TLI > .99, and psychology RMSEA = 0.02 (90% CI 0.00, 0.03), CFI > .99, and TLI = .99. ***p < .001. **p < .01. *p < .05. +p < .1

students who classify themselves as other (all of whom are underrepresented in these courses) demonstrated lower chemistry course grades on average than did White and or Asian students (who make up the racial majority of these courses). The SEM results (Fig. 3a; Appendix Table 7) suggest there are negative indirect effects of academic resources and motivational attitudes on the association between race/ethnicity and chemistry course grade. For academic resources, there are indirect effects of high school GPA and SAT Math (underrepresented students had lower high school GPAs and SAT Math scores on average) on the association between race/ethnicity and chemistry course grade (mediation

effects β = −.02, p = .030 and β = −.03, p < .001, respectively). For motivational attitudes, there was a negative indirect effect of chemistry interest (underrepresented students demonstrated less interest) on the relationship between race/ethnicity and chemistry course grades (mediation effect β = −.01, p = .020).

The model also suggests negative indirect effects of academic resources on the associations between race/ethnicity and motivational attitudes. There was a negative indirect effect of SAT Math on the association between race/ethnicity and chemistry interest (mediation effect β = −.01, p = .010). There were also negative indirect effects of high school GPA and SAT Math on the
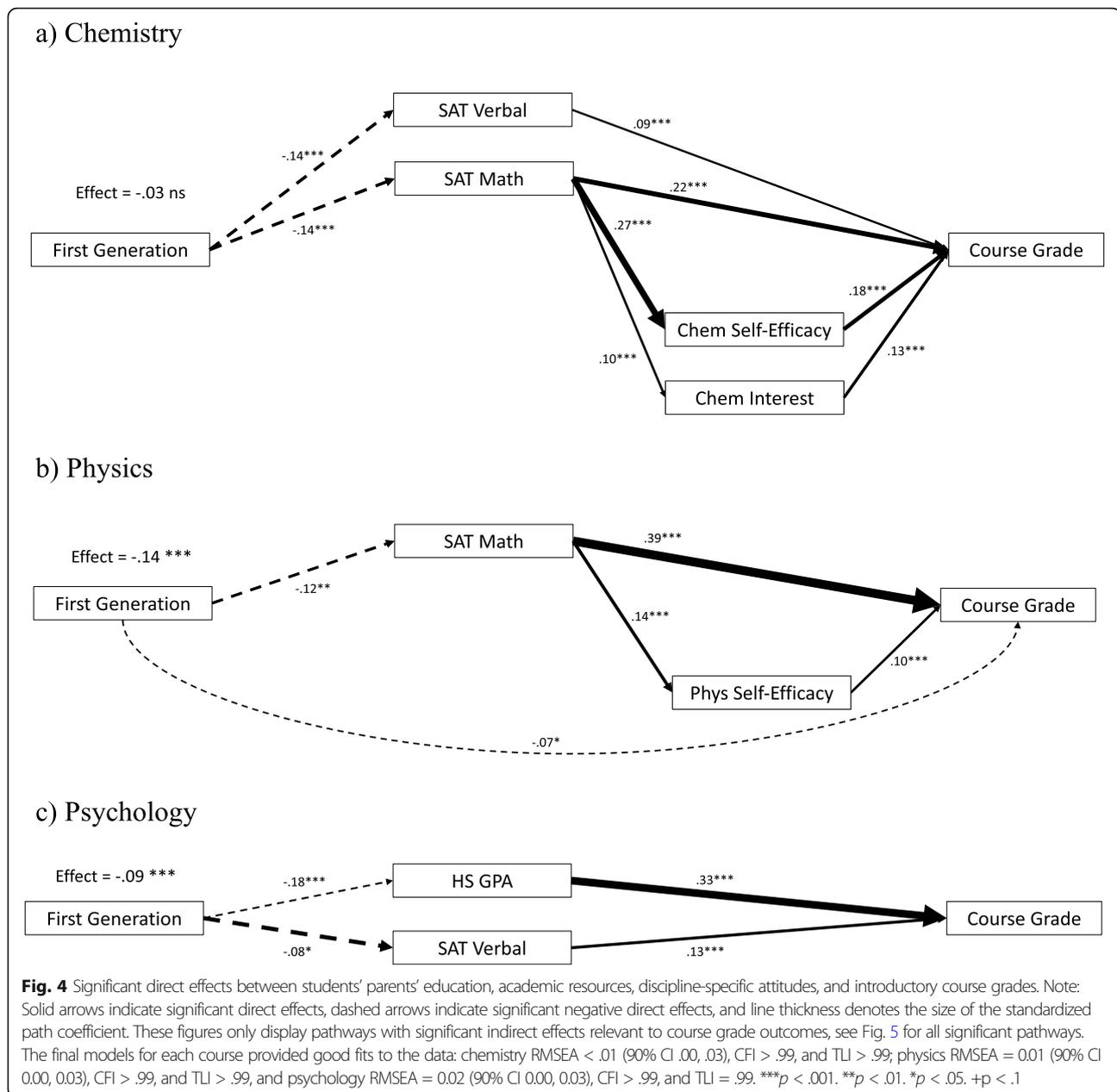
Fig. 3 Significant direct effects between students' race/ethnicity, academic resources, discipline-specific attitudes, and introductory course grades. Note: Solid arrows indicate significant direct effects, dashed arrows indicate significant negative direct effects, and line thickness denotes the size of the standardized path coefficient. These figures only display pathways with significant indirect effects relevant to course grade outcomes; see Fig. 5 for all significant pathways. The final models for each course provided good fits to the data: chemistry RMSEA < .01 (90% CI .00, .03), CFI > .99, and TLI > .99; physics RMSEA = 0.01 (90% CI 0.00, 0.03), CFI > .99, and TLI > .99, and psychology RMSEA = 0.02 (90% CI 0.00, 0.03), CFI > .99, and TLI = .99. ***p < .001. **p < .01. *p < .05. +p < .1

association between race/ethnicity and chemistry self-efficacy (mediation effects β = −.01, p = .036; β = −.03, p < .001, respectively).

However, these significant mediators do not explain the full course grade disparity. The SEM suggests a negative direct effect of underrepresented racial/ethnic backgrounds on chemistry course grades (direct effect β = −.07, p = .011) even when accounting for academic resources and motivational attitudes.

Physics Similar to chemistry, physics has a racial/ethnic disparity in course grades with White and or Asian students receiving higher course grades on average than students from underrepresented racial/ethnic

backgrounds. The SEM (Fig. 3b, Appendix Table 7) highlights that the significant negative relationship may be explained by a negative indirect effect of SAT Math scores (underrepresented students had lower SAT Math scores on average) on the association between race/ethnicity and physics course grades (mediation effect β = −.02, p = .049).

Psychology Unlike chemistry and physics, students from underrepresented racial/ethnic backgrounds had only marginally lower course grades in psychology compared to White and or Asian students. The SEM (Fig. 3c; Appendix Table 7) illustrates that this small effect may be due to negative indirect effects of high school GPA

Fig. 4 Significant direct effects between students' parents' education, academic resources, discipline-specific attitudes, and introductory course grades. Note: Solid arrows indicate significant direct effects, dashed arrows indicate significant negative direct effects, and line thickness denotes the size of the standardized path coefficient. These figures only display pathways with significant indirect effects relevant to course grade outcomes, see Fig. 5 for all significant pathways. The final models for each course provided good fits to the data: chemistry RMSEA < .01 (90% CI .00, .03), CFI > .99, and TLI > .99; physics RMSEA = 0.01 (90% CI 0.00, 0.03), CFI > .99, and TLI > .99, and psychology RMSEA = 0.02 (90% CI 0.00, 0.03), CFI > .99, and TLI = .99. ***$p < .001$. **$p < .01$. *$p < .05$. +$p < .1$

and SAT Verbal scores (underrepresented students had lower high school GPAs and SAT Verbal scores on average) on the association between race/ethnicity and psychology course grades (mediation effects $\beta = -.04$, $p < .001$; $\beta = -.01$, $p = .019$, respectively).

Parent education

Chemistry The regression analysis found that slightly lower chemistry course performance for first-generation college students was not statistically significant. However, the SEM (Fig. 4a; Appendix Table 7) found negative indirect effects of SAT Math and Verbal scores

(first-generation students had lower SAT scores on average than students who have a parent with a college degree) on the association between parent education and chemistry course grades (mediation effects $\beta = -.03$, $p < .001$ and $\beta = -.01$, $p = .015$, respectively).
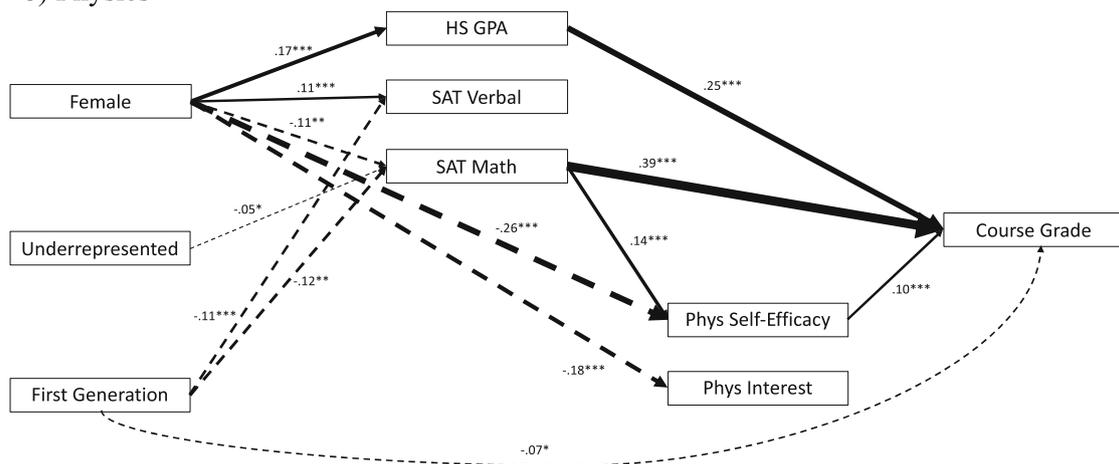
There was also a negative indirect effect of SAT Math scores on the associations between parent education and interest in chemistry (mediation effect $\beta = -.01$, $p = .006$) and parent education and chemistry self-efficacy (mediation effect $\beta = -.04$, $p < .001$).

Physics The largest course grade disparity in physics is between first-generation college students and college
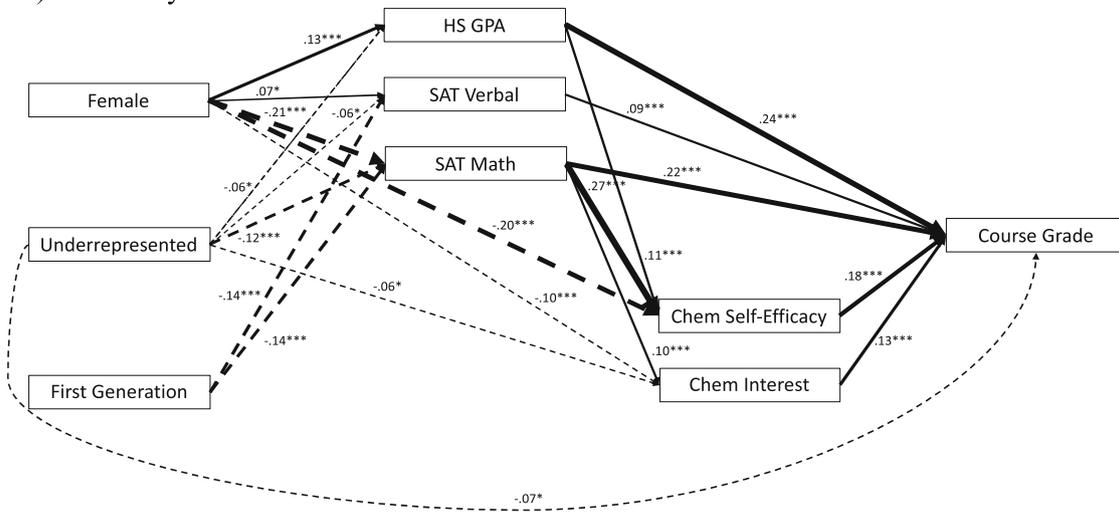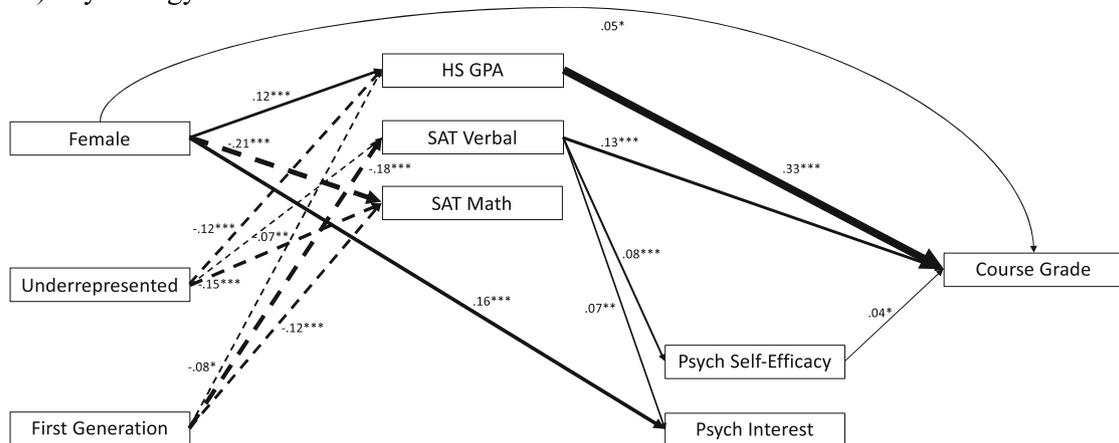
Fig. 5 (See legend on next page.)

(See figure on previous page.)
Fig. 5 All significant direct associations between students' demographics, academic resources, discipline-specific attitudes, and introductory course grades. Note: Solid arrows indicate significant direct effects, dashed arrows indicate significant negative direct effects, and line thickness denotes the size of the standardized path coefficient. The final models for each course provided good fits to the data: chemistry RMSEA < .01 (90% CI .00, .03), CFI > .99, and TLI > .99; physics RMSEA = .01 (90% CI 0.00, 0.03), CFI > .99, and TLI > .99, and psychology RMSEA = 0.02 (90% CI 0.00, 0.03), CFI > .99, and TLI = .99. ***p < .001. **p < .01. *p < .05. +p < .1

students with at least one college-educated parent. Similar to chemistry, but with greater magnitude, the SEM results (Fig. 4b; Appendix Table 7) suggest a negative indirect effect of SAT Math scores (first-generation students have lower SAT Math scores on average) on the relationship between parent education and physics course grades (mediation effect β = −.05, p = .001). Additionally, there is a negative indirect effect of SAT Math scores on the relationship between first-generation students and physics self-efficacy (mediation effect β = −.02, p = .005). However, there is also a remaining direct effect of first-generation status on physics course grades (direct effect β = −.07, p = .029); so the physics course grade disparity is not entirely explained through academic resources and motivational attitudes about physics (Fig. 4b; Appendix Table 6).

Psychology Similar to physics, first-generation students in psychology demonstrated significantly worse course performance on average, although with a smaller effect size. The SEM (Fig. 4c; Appendix Table 7) suggests negative indirect effects of high school GPA and SAT Verbal scores (first-generation students have lower high school GPAs and SAT Verbal scores on average) on the association between parent education and psychology course grades (mediation effects β = −.03, p = .011 and β = −.02, p < .001, respectively).

Moderation

Finally, we were concerned that the pathways highlighted above in the mediation analyses may have been operating differently for different demographic subgroups. For example, discrimination and biases in the learning environment may lead one subgroup to (1) more heavily revise their self-efficacy based upon prior academic performance or (2) rely more on motivational attitudes for course performance. For the latter group, motivational attitudes would be a stronger predictor of course performance. While we lacked sufficient power to simultaneously test every potential moderating pathway using a multi-group SEM framework, we felt it was important to test the moderating effects of the demographic variables (gender, racial/ethnic background, and parent education) on the pathways from academic resources to self-efficacy and self-efficacy to the course grade. There were only a small number of statistically significant moderation effects. In physics, the positive effect of self-efficacy on course grades was significantly stronger for men (unstandardized simple slope (SS) = .36) than women (SS = −.04, p < .001). In

psychology, the positive effect of self-efficacy on course grades was significantly stronger for first-generation college students (SS = .47) than non-first generation students (SS = .04, p = .043). Also in psychology, the positive effect of SAT Verbal on self-efficacy was significantly stronger for men (SS = .09) than women (SS = .01, p = .003). There were no significant moderating effects in chemistry. Given the small number of moderating effects, we only include tables and figures with the mediation analyses in the manuscript and the moderation results are available in online supplementary Tables 7, 8, and 9.

Discussion

The current study extends a core understanding of inequities in educational outcomes within large introductory college courses in four fundamental ways. First, the study highlights robust consistent trends across disciplines and demographic predictors. Some combination of academic resources and discipline-specific motivational attitudes always had direct effects on course grades in chemistry, physics, and psychology (see Fig. 5 and Appendix Table 6). High school GPA and discipline-specific self-efficacy in particular were significant predictors of course grades in each SEM, replicating prior findings in the literature (Robbins et al., 2004). SAT was also important for each discipline, but whether it was SAT Math, SAT Verbal, or both varied, this is also consistent with prior findings that SAT Math is especially important for chemistry and physics and SAT Verbal predicts psychology grades (Betancur et al., 2019; Kalender et al., 2020; Vincent-Ruz et al., 2018).

Furthermore, academic resources always had direct effects on motivational attitudes (see Fig. 5 and Appendix Table 6): that is, the general predictiveness of academic resources on course grades is partially mediated through motivational attitudes. Replicating prior work on self-efficacy (Vincent-Ruz et al., 2018), SAT scores generally predict self-efficacy, and the mediational effects (see Appendix Table 7) of SAT Math scores to course grades via self-efficacy were statistically significant in chemistry and physics. The findings regarding discipline-specific interest were more novel. The SEMs provide evidence that SAT scores were significant predictors of discipline-specific interest in chemistry and psychology, and the mediational effect of SAT Math on course grades through discipline-specific interest was statistically significant in chemistry.

There were also consistent direct effects between particular demographic variables and both academic resources and

motivational attitudes, even though most of these patterns varied substantially by demographic variable (see Fig. 5 and Appendix Table 6). Gender was consistently related to academic resources and motivational attitudes, whereas parent education was only directly connected to academic resources, and race/ethnicity was associated with academic resources with only one significant direct connection to discipline-specific interest in Chemistry. Moreover, the associations of students from underrepresented racial/ethnic backgrounds and first-generation students to academic resources were consistently negative (although of varying size), whereas the associations for female students to both academic resources and motivational attitudes were positive in some cases and negative in others. For example, as generally found in the literature (Fortin et al., 2015) and in recent reports on the SAT national sample (e.g., College Board, 2016), women in all three samples had higher high school GPAs but lower SAT Math scores than men on average.

The second fundamental takeaway is the finding that disparities in course performance are variable by course, even within the same University and for courses that all predominantly enroll first-year students. In one case, there are even effects in opposite directions: women outperform men on average in psychology while they underperform on average compared with men in physics. The other academic performance disparities may have been directionally equivalent, but they all yield different effect sizes across courses. For example, in chemistry, the largest academic disparities relate to race/ethnicity, while in physics they relate to parent education (Table 5). Each discipline's sample included over one thousand students, and only three disciplines were examined, thus reducing the likelihood of by-chance variation. Further, the variation was not likely a matter of individual instructor effects since each course involved a range of instructors. Therefore, these findings support that research on education inequity in large introductory lecture courses needs to attend to discipline-specific disparities, rather than treating all courses as the same, even within STEM.

Third, the current study uncovered significant indirect effects of the relationships between gender, race/ethnicity, and parent education on course performance operating through academic resources and motivational attitudes. Importantly, the analyses of indirect effects uncovered course-variations in demographic differences in both academic resources and motivational attitudes. For example, first-generation students have lower high school GPAs on average in psychology but show no significant difference from non-first generation students within chemistry or physics. Additionally, women have significantly higher SAT Verbal scores on average in the two natural science courses, but not in psychology. Demographic differences in motivational attitudes include women demonstrating significantly less self-efficacy on average in chemistry and physics while there

are no significant gender differences in psychology self-efficacy on average. These variations in demographic differences by course may be attributable to self-selection into courses and may also reflect broader self-selection trends within majors. For example, gender differences in SAT Verbal scores may only occur in the natural science courses because men with higher SAT Verbal scores are less likely to take natural science courses (Wang et al., 2013). Additionally, women may have significantly less interest in and self-efficacy regarding chemistry and physics due to beliefs and stereotypes about women in STEM that may not apply to psychology.

Furthermore, the SEMs revealed differences in links between academic resources, motivational attitudes, and course performance across subpopulations of students. In all three courses, combinations of academic resources and motivational attitudes were significant predictors of introductory course grades. Thus, the general conceptual model is robust and generalizable; however, the specific significant academic resources and motivational attitudes varied by course. As expected, physics and chemistry were more dependent upon math resources whereas psychology was more dependent upon verbal resources. Self-efficacy was the key motivational attitude in psychology and physics, while both self-efficacy and interest were significant predictors for chemistry.

Fourth and finally, this study highlights important null findings and small effect sizes that may dispel assumptions about disparities in academic performance and assumptions about what explains these differences. As previously mentioned, the difference between a B and a B+, for example, is 0.25 grade points. Many of the significant demographic differences were less than 0.25 grade points (Table 5), and thus rarely reflected different letter grades. The largest effects are the difference in average chemistry course grades (–0.41 grade points) between White and or Asian students (who are overrepresented in these courses) and Black, Latinx, Native American, Pacific Islander, and students of other descent (who are underrepresented in these courses), and the difference in average physics course grades (–0.53 grade points) between first-generation college students and students who have a parent with a college degree. These larger disparities operationally represent the difference between a B and an A–.

This study also highlights important small and null findings with regard to the mediators explaining associations between demographic factors and course grades. For example, across all three courses, first-generation students only exhibit significantly different motivational attitudes related to self-efficacy and interest through indirect effects of SAT scores. The fact that there are no direct effects of parent education on motivational attitudes highlights that the reason for first-generation students' lower grades on average cannot likely be

attributed to motivational attitudes independently of academic resources. In other words, the results largely support that academic resources in high school have more to do with academic performance disparities for underrepresented and first-generation students than motivational attitudes in college. However, given the estimated significant pathways from academic resources to course grades through motivational attitudes (see Appendix Table 7), the results support that addressing motivational attitudes through interventions may have the capacity to attenuate disparities in academic performance resulting from differential access to academic resources in high school. This is discussed further in the conclusion.

Caveats and future research

It should be noted that these observational findings cannot be used to make strong causal claims about the mechanisms underlying associations between demographic characteristics and academic outcomes. However, the SEM framework certainly provides compelling evidence for indirect effects that mediate these direct associations. In regression or SEM approaches, there should also be consideration of possible confounding factors that might result in alternative causal pathways. The current analyses build upon prior studies conducted on these courses at the University of Pittsburgh to identify which demographic, academic resource, and motivational factors from a much broader set were significant predictors (Vincent-Ruz et al., 2018; Betancur et al., 2019; Kalender et al., 2020; Witherspoon et al., 2019). The current analyses focused on the factors previously found to be significant. In applying our analytic approach at other universities, we encourage broader initial exploration to identify which factors are most important to control for.

It is also important to emphasize that the undergraduates attending the University and especially those enrolled in introductory STEM courses are predominantly White and non-first generation (University of Pittsburgh, 2019; Table 1). Thus, the findings may not generalize to institutions with different populations of students. However, the methods and analytic framework presented in this study could be productively applied at other institutions, and with other courses, to highlight the specific factors related to demographic-based performance differences. Further, dashboards could be created to help administrators and faculty attend to gaps and underlying factors, without needing to have access to large administrative datasets or master SEM techniques. Hence, future research should examine associations between demographics, prior academic resources, discipline-specific motivational attitudes, and academic performance at institutions with more socioeconomic and racial/ethnic variability. Future research should also address the previously mentioned limitations that arise from using binary gender categories and measuring race/ethnicity with dummy variables (Brown et al., 2019; Hyde et al., 2019).

One limitation of comparing discipline-specific interest and self-efficacy across chemistry, physics, and psychology courses is that slightly different survey items were used to measure these constructs, so they are not operationally equivalent across disciplines. As previously mentioned, the items used to create these constructs were collected from existing scales and surveys that are discipline-specific and not necessarily intended to be compared across disciplines. Additional details on the items and measures can be found in Table 2. Future research should explore the mediating pathways of motivational attitudes on scales that are standardized and designed to be comparable across disciplines. However, it is also important to reiterate that the self-efficacy differences observed in this study replicate past findings in those specific disciplines using a variety of other self-efficacy instruments. Thus, it is unlikely that there is high sensitivity to a small variation in survey structure.

Additionally, while these findings highlight significant indirect effects through academic resources and motivational attitudes, there are still enduring demographic disparities in course grades (see Fig. 5 and Appendix Table 6). In chemistry, there are estimated remaining direct effects of students from underrepresented racial/ethnic backgrounds on course grades; in physics, there are remaining direct effects of parent education; and in psychology, there are remaining direct effects of gender even when controlling for pathways through academic resources and motivational attitudes. This highlights that there are residual explanations for why some students are systemically underperforming academically that are not explored in this study. Future research should examine additional possible operationalizations of motivational attitudes (e.g., the roles of sense-of-belonging, discipline-specific identity, or discipline-specific theories of intelligence). Just as importantly, research should consider other explanations for disparities in outcomes, such as characteristics of the learning environment (e.g., instructor and peer beliefs and stereotypes that might produce racial aggressions or stereotype-reinforcing messages in class towards particular groups).

Finally, future research should examine whether these results are robust beyond introductory and or first-year courses. Two different cases seem particularly salient. First is the case of organic chemistry or algebra-based physics, which is usually taken by medical-school intending students after the first year of their studies (Witherspoon et al., 2019). Attrition during the first year can systematically change what population remains (Witherspoon et al., 2019). Second is the case of challenging "majors only" courses. How does the student population of an introductory course map on to the population of majors in a given discipline? Self-efficacy in a discipline may not be a robust predictor of course performance in an upper-level course comprised only of majors. Note that in studies of later coursework, high school GPA and

SAT should be replaced by measures reflecting more temporally proximate measures of academic resources because those are malleable factors that can fundamentally change with experience. Additional studies might also examine these pathways in other disciplines beyond chemistry, physics, and psychology that demonstrate performance disparities by gender, race/ethnicity, and or parent education.

Conclusion

This study highlights that demographic differences in post-secondary course performance differ by discipline within STEM, and distinct pathways through academic resources and discipline-specific motivational attitudes help explain these differences. Therefore, this study emphasizes the importance of considering the unique and complex ways students' gender, race/ethnicity, and parents' education interface with academic resources and motivational attitudes, and how these constructs work directly and indirectly to differentially predict students' academic outcomes.

From a theory-building perspective, it now becomes important to understand when and why academic resources and motivational attitudes are predictors of course performance. Conceptualizations of why high school GPA, SAT scores, and discipline-specific interest and self-efficacy predict academic outcomes should apply at least somewhat to all learning contexts. Course contexts that are radically different (e.g., lab courses vs. lecture courses vs. small seminars) might require very different learning behaviors and thus depend upon different resources. However, it is somewhat surprising from a theoretical perspective that differences in significant predictors were found in these relatively homogeneously structured courses.

In addition to highlighting important theoretical pathways to consider when examining disparities in course performance, these findings also highlight opportunities for instructors to play to the relative strengths of underperforming students. For example, women in physics demonstrated superior SAT Verbal skills on average yet have inferior course grades on average compared with men. With this knowledge, instructors interested in improving the average grades of women can diversify their assignments to play on verbal strengths through more open-ended responses or written explanations. Further, different degree specializations (ecology specialization within biology) or special-topic versions of core courses (e.g., health sciences General Chemistry I) will involve student self-selection and thus other potential population differences that should be measured and carefully considered in the design of these programs and courses.

The findings also highlight key areas for supporting students' perceptions of their efficacy in specific courses. Chemistry, physics, and psychology exhibited positive direct effects of discipline-specific self-efficacy on course grades. Thus, instructors can employ teaching methods and interventions aimed at improving underperforming students' self-efficacy in their disciplines. For example, Miyake et al. (2010) introduced a course-specific values affirmation intervention that reduced the gender performance gap in an introductory physics course. Social belonging and mindset interventions are also proven to be effective for changing attitudes and reducing disparities (Walton & Cohen, 2011; Chen et al., 2020). More general field-specific personal values and or framing interventions may also have the potential to improve academic outcomes of women, racially/ethnically minoritized students, and first-generation students in introductory college science courses through motivational attitudes such as interest and self-efficacy (Harackiewicz & Priniski, 2018). However, the findings emphasize that interventions should be targeted to the subpopulations that actually stand to benefit from them depending on the course context. For example, in this study, women significantly outperform men on average in psychology, but the opposite is true in physics. Women also only demonstrate significantly lower self-efficacy on average in chemistry and physics. Thus, interventions to improve women's discipline-specific self-efficacy in order to improve course grades are likely not necessary in psychology courses while they may be instrumental in chemistry and physics courses.

At the same time, we should caution against focusing only on interventions that aim to improve students' attitudes directly. Classroom environments also need to change. Students should not receive stereotype-reinforcing messages from their instructors; instructor mindsets about innate talent versus possibility for growth are shown to be strong predictors of racial/ethnic disparities (Canning et al., 2019). Importantly, students too regularly encounter toxic cultures in certain disciplines within STEM. For example, a recent report revealed continuing issues of sexual harassment of undergraduates in physics (Aycock et al., 2019). Factors such as bias and harassment may explain the unmediated disparities in our own study.

Overall, this study identifies the magnitude of differences in student academic performance related to gender, race/ethnicity, and parent education, in large, lecture-based introductory courses in chemistry, physics, and psychology in one university setting. The study also highlights the extent to which demographic differences in academic resources and discipline-specific motivational attitudes explain differences in academic performance within each of these courses. Finally, the similarities and differences in findings across the different disciplines provide a compelling argument for implementing this methodology to examine questions about academic disparities and mediators of these disparities separately by fields within STEM.

Appendix

Table 6 Direct effect results of structural equation models for demographic characteristics predicting course grades

Direct Effect Results of Structural Equation Models for Demographic Characteristics Predicting Course Grades

Path Estimated	Chemistry				Physics				Psychology			
	Coeff.	Bootstrapped SE	Standardized Coeff.	p-value	Coeff.	Bootstrapped SE	Standardized Coeff.	p-value	Coeff.	Bootstrapped SE	Standardized Coeff.	p-value
High School GPA to Course Interest												
High School GPA to Course Self-Efficacy	0.15	0.04	0.11	0.000								
High school GPA to Course Grade	0.56	0.05	0.24	0.000	0.59	0.08	0.25	0.000	0.50	0.06	0.33	0.000
SAT Verbal to Course Interest									0.05	0.02	0.07	0.005
SAT Verbal to Course Self-Efficacy									0.04	0.01	0.08	0.000
SAT Verbal to Course Grade	0.11	0.03	0.09	0.000					0.12	0.02	0.13	0.000
SAT Math to Course Interest	0.09	0.02	0.10	0.000								
SAT Math to Course Self-Efficacy	0.21	0.02	0.27	0.000	0.10	0.02	0.14	0.000				
SAT Math to Course Grade	0.28	0.04	0.22	0.000	0.63	0.04	0.39	0.000				
Course Interest to Course Grade	0.19	0.05	0.13	0.000								
Course Self-Efficacy to Course Grade	0.29	0.05	0.18	0.000	0.24	0.06	0.10	0.000	0.08	0.04	0.04	0.043
Gender (Female=1) **to High School GPA**	0.11	0.02	0.13	0.000	0.16	0.03	0.17	0.000	0.11	0.02	0.12	0.000
Gender to SAT Verbal	0.11	0.04	0.07	0.017	0.17	0.05	0.11	0.000				
Gender to SAT Math	-0.31	0.04	-0.21	0.000	-0.15	0.05	-0.11	0.002	-0.32	0.03	-0.21	0.000
Gender to Course Interest	-0.13	0.03	-0.10	0.000	-0.21	0.04	-0.18	0.000	0.16	0.02	0.16	0.000
Gender to Course Self-Efficacy	-0.23	0.03	-0.20	0.000	-0.25	0.03	-0.26	0.000				
Gender to Course Grade									0.07	0.03	0.05	0.034
Race/Ethnicity (Underrepresented background=1) **to High School GPA**	-0.08	0.03	-0.06	0.019					-0.17	0.03	-0.12	0.000
Race/Ethnicity to SAT Verbal	-0.14	0.07	-0.06	0.041					-0.15	0.05	-0.07	0.005
Race/Ethnicity to SAT Math	-0.29	0.07	-0.12	0.000	-0.10	0.05	-0.05	0.040	-0.32	0.05	-0.15	0.000
Race/Ethnicity to Course Interest	-0.13	0.05	-0.06	0.018								
Race/Ethnicity to Course Self-Efficacy												
Race/Ethnicity to Course Grade	-0.20	0.08	-0.07	0.011								
Parent Education (First generation=1) **to High School GPA**									-0.12	0.05	-0.08	0.012
Parent Education to SAT Verbal	-0.34	0.10	-0.14	0.000	-0.31	0.09	-0.11	0.000	-0.45	0.06	-0.18	0.000
Parent Education to SAT Math	-0.33	0.08	-0.14	0.000	-0.30	0.10	-0.12	0.002	-0.30	0.06	-0.12	0.000
Parent Education to Course Interest												
Parent Education to Course Self-Efficacy												
Parent Education to Course Grade					-0.27	0.12	-0.07	0.029				

Note. After initially testing all possible mediating pathways, insignificant ($p > .05$) pathways and insignificant correlated errors of mediators were trimmed to improve the goodness of fit of the model. Only the pathways displayed in this table were estimated in the final structural equation models (SEMs). This table includes each individual direct effect, not the total direct effects.

Table 7 Indirect effect results of structural equation models with academic resources and motivational attitudes as mediators

Indirect Effect Results of Structural Equation Models with Academic Resources and Motivational Attitudes as Mediators

Path Estimated	Mediator	Chemistry				Physics				Psychology			
		Coeff.	Bootstrapped SE	Standardized Coeff.	p-value	Coeff.	Bootstrapped SE	Standardized Coeff.	p-value	Coeff.	Bootstrapped SE	Standardized Coeff.	p-value
High school GPA to Course Grade	Course Interest												
	Course Self-Efficacy	0.04	0.01	0.02	0.000								
SAT Verbal to Course Grade	Course Interest												
	Course Self-Efficacy									0.004	0.002	0.004	0.102
SAT Math to Course Grade	Course Interest	0.02	0.01	0.01	0.007								
	Course Self-Efficacy	0.06	0.01	0.05	0.000	0.02	0.01	0.01	0.002				
Gender to Course Interest (Female=1)	High School GPA												
	SAT Verbal												
	SAT Math	-0.03	0.01	-0.02	0.002								
Gender to Course Self-Efficacy	High School GPA	0.02	0.005	0.01	0.001								
	SAT Verbal												
	SAT Math	-0.07	0.01	-0.06	0.000	-0.01	0.01	-0.02	0.004				
Gender to Course Grade	High School GPA	0.06	0.01	0.03	0.000	0.09	0.02	0.04	0.000	0.06	0.01	0.04	0.000
	SAT Verbal	0.01	0.01	0.01	0.061								
	SAT Math	-0.09	0.02	-0.05	0.000	-0.09	0.03	-0.04	0.001				
	Course Interest	-0.02	0.01	-0.01	0.004								
	Course Self-Efficacy	-0.07	0.02	-0.04	0.000	-0.06	0.02	-0.03	0.001				
Race/Ethnicity to Course Interest (Underrepresented background=1)	High School GPA												
	SAT Verbal									-0.01	0.004	-0.004	0.097
	SAT Math	-0.02	0.01	-0.01	0.010								
Race/Ethnicity to Course Self-Efficacy	High School GPA	-0.01	0.01	-0.01	0.036								
	SAT Verbal									-0.01	0.003	-0.01	0.042
	SAT Math	-0.06	0.01	-0.03	0.000	-0.01	0.01	-0.01	0.071				
Race/Ethnicity to Course Grade	High School GPA	-0.05	0.02	-0.02	0.030					-0.08	0.02	-0.04	0.000
	SAT Verbal	-0.02	0.01	-0.01	0.085					-0.02	0.01	-0.01	0.019
	SAT Math	-0.08	0.02	-0.03	0.000	-0.06	0.03	-0.02	0.049				
	Course Interest	-0.02	0.01	-0.01	0.020								
	Course Self-Efficacy												
Parent Education to Course Interest (First generation=1)	High School GPA												
	SAT Verbal									-0.02	0.01	-0.01	0.017
	SAT Math	-0.03	0.01	-0.01	0.006								
Parent Education to Course Self-Efficacy	High School GPA												
	SAT Verbal									-0.02	0.01	-0.01	0.001
	SAT Math	-0.07	0.02	-0.04	0.000	-0.03	0.01	-0.02	0.005				
Parent Education to Course Grade	High School GPA									-0.06	0.02	-0.03	0.011
	SAT Verbal	-0.04	0.02	-0.01	0.015					-0.06	0.02	-0.02	0.000
	SAT Math	-0.09	0.03	-0.03	0.000	-0.19	0.06	-0.05	0.001				
	Course Interest												
	Course Self-Efficacy												

Note. After initially testing all possible mediating pathways, insignificant ($p > .05$) pathways and insignificant correlated errors of mediators were trimmed to improve the goodness of fit of the model. Only the pathways displayed in this table were estimated in the final structural equation models (SEMs). This table includes each individual indirect effect, not the total indirect effects through multiple mediators.

Table 8 Fit indices for all model specifications

Course sample	Chemistry	Chemistry	Chemistry	Physics	Physics	Physics	Psychology	Psychology	Psychology	Good Fit Indices (Hu & Bentler, 1999)
SEM specification	Chemistry	Physics	Psychology	Physics	Chemistry	Psychology	Psychology	Chemistry	Physics	
RMSEA	**0.00**	0.03	0.04	**0.01**	0.06	0.05	**0.02**	0.13	0.13	<.06
90% CI lower bound	**0.00**	0.00	0.03	**0.00**	0.05	0.04	**0.00**	0.12	0.12	
90% CI upper bound	**0.03**	0.05	0.05	**0.03**	0.08	0.06	**0.03**	0.14	0.15	
CFI	**1.00**	0.99	0.98	**1.00**	0.95	0.95	**1.00**	0.84	0.76	>0.95
TLI	**1.00**	0.98	0.94	**1.00**	0.91	0.89	**0.99**	0.62	0.44	>0.95

Supplementary information

> **Additional file 1: Table S1.** Chemistry unimputed regression. **Table S2**. Physics Unimputed Regression Results. **Table S3.** Psychology Unimputed Regression Results. **Table S4.** Chemistry SEM Results without MLMV. **Table S5.** Physics SEM Results without MLMV. **Table S6.** Psychology SEM Results without MLMV. **Table S7.** Chemistry SEM Moderation Results. **Table S8.** Physics SEM Moderation Results. **Table S9** Psychology SEM Moderation Results.

Abbreviations
ACT: American College Testing; EVT: Expectancy value theory; ICE: Imputed chained equations; IRT: Item response theory; CFI: Comparative fit index; GPA: Grade point average; MLMV: Maximum likelihood with missing values; RMSEA: Root mean squared error of approximation; SAT: Scholastic Aptitude Test; SEM: Structural Equation Model(ing); SS: Simple slope; STEM: Science, technology, engineering, and mathematics; TLI: Tucker-Lewis Index

Authors' contributions
LB analyzed the data and contributed to writing and revising the manuscript; CS conceived and designed the study and contributed to writing and revising the manuscript; EVD contributed to the design of the study and revising manuscript; BR contributed to the conception of the study and revising the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interest.

References
Adams W.K., Wieman C. E., Perkins K. K., and Barbera, J. (2008). Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry. Journal of Chemistry Education, 85(10), 1435.

Allensworth, E. M., & Clark, K. (2020). High school GPAs and ACT scores as predictors of college completion: Examining assumptions about consistency across high schools. *Educational Researcher, 49*(3), 198–211.

Aycock, L. M., Hazari, Z., Brewe, E., Clancy, K. B., Hodapp, T., & Goertzen, R. M. (2019). Sexual harassment reported by undergraduate female physicists. *Physical Review Physics Education Research, 15*(1), 010121.

Bauer, C. F. (2005). Beyond "student attitudes": Chemistry self-concept inventory for assessment of the affective component of student learning. *Journal of Chemistry Education, 82*(12), 1864.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133*(4), 584.

Bernal, P., Mittag, N., & Qureshi, J. A. (2016). Estimating effects of school quality using multiple proxies. *Labour Economics, 39*, 1–10.

Betancur, L., Rottman, B. M., Votruba-Drzal, E., & Schunn, C. D. (2019). Analytical assessment of course sequencing: The case of methodological courses in psychology. *Journal of Educational Psychology, 111*(1), 91–103.

Betancur, L., Votruba-Drzal, E., & Schunn, C. D. (2018). Socioeconomic gaps in science achievement. *International Journal of STEM Education, 5*, 38.

Boyer Commission (1998). Reinventing undergraduate education: a blueprint for America's research universities.

Brown, K. S., Kijakazi, K., Runes, C., & Turner, M. A. (2019). *Confronting structural racism in research and policy analysis*. Urban Institute https://www.urban.org/research/publication/confronting-structural-racism-research-and-policy-analysis.

Canning, E. A., Muenks, K., Green, D. J., & Murphy, M. C. (2019). STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. *Science Advances, 5*, eaau4734.

Chen, S., Binning, K. R., Manke, K. J., Brady, S. T., McGreevy, E. M., Betancur, L., … Kaufmann, N. (2020). Am I a science person? A strong science identity bolsters minority students' sense of belonging and performance in college. *Personality and Social Psychology Bulletin*, 1–14.

College Board (2016). 2016 college-bound seniors total group profile report. https://secure-media.collegeboard.org/digitalServices/pdf/sat/total-group-2016.pdf

College Board. (2018a). Program summary report. https://secure-media.collegeboard.org/digitalServices/pdf/research/2018/Program-Summary-Report-2018.pdf

College Board. (2018b). Guide to the ACT®/SAT® concordance. https://collegereadiness.collegeboard.org/educators/higher-ed/scoring/concordance

Conley, D. T. (2003). *Understanding university success. A report from standards for success*. Center for Educational Policy Research.

Conley, D. T. (2007). *Redefining college readiness*. Educational Policy Improvement Center.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd.

Eccles, J. S. (2005). *Subjective task value and the Eccles et al. model of achievement-related choices*, (pp. 105–121). Handbook of Competence Motivation.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457.

Fencl, H., & Scheel, K. (2005). Engaging students. *Journal of College Science Teaching, 35*(1), 20–24.

Fortin, N. M., Oreopoulos, P., & Phipps, S. (2015). Leaving boys behind gender disparities in high academic achievement. *Journal of Human Resources, 50*(3), 549–579.

Geiser, S., & Santelices, M. V. (2007). *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*. UC Berkeley Research and Occasional Paper Series https://escholarship.org/uc/item/7306z0zf.

González, A., & Paoloni, P. V. (2015). Perceived autonomy-support, expectancy, value, metacognitive strategies and performance in chemistry: a structural equation model in undergraduates. *Chemistry Education Research and Practice, 16*(3), 640–653.

Hanushek, E. A., & Woessmann, L. (2017). School resources and student achievement: A review of cross-country economic research. In M. Rosén, K. Yang Hansen, & U. Wolff (Eds.), *Cognitive Abilities and Educational Outcomes*, (pp. 149–171). Springer.

Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: the science of targeted intervention. *Annual Review of Psychology, 69*, 409–435.

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

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127.

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: a systematic review. *Educational Research Review, 17*, 63–84.

Hsieh, T. L. (2014). Motivation matters? The relationship among different types of learning motivation, engagement behaviors and learning outcomes of undergraduate students in Taiwan. *Higher Education, 68*(3), 417–433.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*(4), 880–895.

Hyde, J. S., Bigler, R. S., Joel, D., Tate, C. C., & van Anders, S. M. (2019). The future of sex and gender in psychology: five challenges to the gender binary. *American Psychologist, 74*(2), 171.

Integrated Postsecondary Education Data System (IPEDS); (2019). University of Pittsburgh-Pittsburgh Campus. https://nces.ed.gov/ipeds/datacenter/institutionprofile.aspx?unitId=215293

Johnson, A. C. (2007). Unintended consequences: How science professors discourage women of color. *Science Education*, 91(5), 805–821.

Kalender, Z. Y., Marshman, E., Schunn, C., Nokes-Malach, T., & Singh, C. (2020). Damage caused by women's lower self-efficacy on physics learning. *Physical Review Physics Education Research*, 16(1).

Keith, T. (2006). *Multiple regression and beyond*. Pearson Education.

Komarraju, M., & Nadler, D. (2013). Self-efficacy and academic achievement: why do implicit beliefs, goals, and effort regulation matter? *Learning and Individual Differences*, 25, 67–72.

Kurban, E. R., & Cabrera, A. F. (2019). Building readiness and intention towards STEM fields of study: using HSLS: 09 and SEM to examine this complex process among high school students. *The Journal of Higher Education*, 1–31.

Lawson, A. E., Banks, D. L., & Logvin, M. (2007). Self-efficacy, reasoning ability, and achievement in college biology. *Journal of Research in Science Teaching*, 44(5), 706–741.

Le, H., Robbins, S. B., & Westrick, P. (2014). Predicting student enrollment and persistence in college STEM fields using an expanded PE fit framework: A large-scale multilevel study. *Journal of Applied Psychology*, 99(5), 915–947.

Leslie, S., Cimpian, A., Meyer, M., & Freeland, E. (2015). Women are underrepresented in disciplines that emphasize brilliance as the key to success. *Science*, 347, 262–265.

Li, S. D. (2011). Testing mediation using multiple regression and structural equation modeling analyses in secondary data. *Evaluation Review*, 35(3), 240–268.

Lopez, F. G., & Lent, R. W. (1992). Sources of mathematics self-efficacy in high school students. *The Career Development Quarterly*, 41(1), 3–12.

Luschei, T. F., & Jeong, D. W. (2018). Is teacher sorting a global phenomenon? Cross-national evidence on the nature and correlates of teacher quality opportunity gaps. *Educational Researcher*, 47(9), 556–576.

Magalhães, P., Ferreira, D., Cunha, J., & Rosário, P. (2020). Online vs traditional homework: a systematic review on the benefits to students' performance. *Computers & Education*, 103869.

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

Marshman, E., Kalender, Z. Y., Schunn, C. D., Nokes-Malach, T., & Singh, C. (2018b). A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences. *Canadian Journal of Physics*, 96, 391–405.

Marshman, E. M., Kalender, Z. Y., Nokes-Malach, T., Schunn, C., & Singh, C. (2018a). Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: a cause for alarm? *Physical Review Physics Education Research*, 14(2).

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., … McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open*, 3(4).

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science*, 330(6008), 1234–1237.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474–16479.

National Research Council (2002). *Learning and understanding: Improving advanced study of mathematics and science in U.S. high schools*. National Academies Press.

National Science Board. (2018). Science & Engineering Indicators 2018 (NSB-2018-1). https://www.nsf.gov/statistics/2018/nsb20181/report/

National Survey of Student Engagement. (2006). Engaged learning: Fostering success for all students.

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

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049-1079.

Page, L. C., Kehoe, S. S., Castleman, B. L., & Sahadewo, G. A. (2019). More than dollars for scholars: The impact of the Dell Scholars Program on college access, persistence, and degree attainment. *Journal of Human Resources*, 54(3), 683-725.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research (Vol. 2)*. Jossey-Bass.

Putnam, A. L., Sungkhasettee, V. W., & Roediger, H. L. (2016). Optimizing learning in college: Tips from cognitive psychology. *Perspectives on Psychological Science*, 11(5), 652-660.

Putnam, R. D. (2016). *Our kids: The American dream in crisis*. Simon and Schuster.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261-288.

Robinson, M. (2003). Student enrollment in high school AP sciences and calculus: How does it correlate with STEM careers? *Bulletin of Science, Technology, Society*, 23(4), 265-273.

Royston, P. (2004). Multiple imputation of missing values. *The Stata Journal*, 4, 227–241.

Royston, P. (2005). Multiple imputation of missing values: Update of ICE. *The Stata Journal*, 5, 527–536.

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

Sadler, P. M., & Tai, R. H. (2007a). Transitions - The two high-school pillars supporting college science. *Science*, 317(5837), 457-458.

Sadler, P. M., & Tai, R. H. (2007b). Weighting for recognition: Accounting for advanced placement and honors courses when calculating high school grade point average. *NASSP Bulletin*, 91(1), 5-32.

Salehi, S., Burkholder, E., LePage, G. P., Pollock, S., & Wieman, C. (2019). The impact of incoming preparation and demographics on performance in Physics I: A multi-institution comparison. *arXiv preprint arXiv:1905.00389*.

Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, 26(2), 89-112.

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In Wigfield, A., & Eccles, J. S. (Eds.), *Development of Achievement Motivation* (pp. 15-31). Academic Press.

Sen, M., & Wasow, O. (2016). Race as a bundle of sticks: Designs that estimate effects of seemingly immutable characteristics. *Annual Review of Political Science*, 19, 499-522.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences (Vol. 12)*. Westview Press.

Sobel, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociological Methods and Research*, 16, 155–176.

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

Taasoobshirazi, G., Puckett, C., & Marchand, G. (2019). Stereotype threat and gender differences in biology. *International Journal of Science and Mathematics Education*, 17(7), 1267-1282.

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning & Individual Differences*, 61, 136-150.

Tyson, W., Lee, R., Borman, K. M., & Hanson, M. A. (2007). Science, technology, engineering, and mathematics (STEM) pathways: High school science and math coursework and postsecondary degree attainment. *Journal of Education for Students Placed at Risk*, 12(3), 243-270.

University of Pittsburgh (2019). *Fact Book, 2019*. https://catalog.upp.pitt.edu/mime/media/view/170/15158/Fact-Book-2019.pdf.

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: another look at the quality of academic motivation. *Educational Psychologist*, 41(1), 19-31.

Vincent-Ruz, P., Binning, K., Schunn, C. D., & Grabowski, J. (2018). The effect of math SAT on women's chemistry competency beliefs. *Chemistry Education Research and Practice*, 19(1), 342-351.

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

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. *Science*, 331(6023), 1447-1451.

Wang, M.-T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Education Psychology Review*, 29(1), 119-140.

Wang, M.-T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice: Individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science, 24*(5), 770-775.

Warne, R. T., Nagaishi, C., Slade, M. K., Hermesmeyer, P., & Peck, E. K. (2014). Comparing weighted and unweighted grade point averages in predicting college success of diverse and low-income college students. *NASSP Bulletin, 98*(4), 261-279.

Whitcomb, K. M., Kalender, Z. Y., Nokes-Malach, T. J., Schunn, C. D., & Singh, C. (2020). A mismatch between self-efficacy and performance: Undergraduate women in engineering tend to have lower self-efficacy despite earning higher grades than men. *arXiv preprint arXiv:2003.06006*.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology. Special Issue: Motivation and the Educational Process, 25*(1), 68-81.

Witherspoon, E. B., Vincent-Ruz, P., & Schunn, C. D. (2019). When making the grade isn't enough: The gendered nature of premed science course attrition. *Educational Researcher, 48*(4), 193-204.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.