Malespina and Singh
International Journal of STEM Education (2023) 10:13
https://doi.org/10.1186/s40594-023-00399-7

International Journal of
STEM Education

RESEARCH
Open Access

Gender gaps in grades versus grade penalties: why grade anomalies may be more detrimental for women aspiring for careers in biological sciences

Alysa Malespina* and Chandralekha Singh

Abstract

Background In the US, bioscience programs now often have more women than men who earn degrees at all levels. Despite this, evidence still exists for gender inequity in bioscience and medical fields. For example, women with careers in these fields tend to get paid less and leave these fields more. Here, we present research investigating grade penalties. Grade penalties describe the difference between a students' grade point average and grade in a particular course when a students' grade in a class is lower than their grade point average (GPA). For example, a student has a grade penalty of 1 in a course if their GPA is 3.0 (B) and they receive a 2.0 (C) in the course. We hypothesize that grade penalties in these STEM courses can affect students' STEM-related self-concept negatively with potentially long-term consequences, e.g., in students' career choices. We investigated grade penalties in science courses for undergraduate students studying biosciences. We propose a framework that posits that students who receive grade penalties in a course may view themselves as less capable in that domain and that gender differences in grade penalties may be a useful measure of gender inequity in the classroom.

Results In this study, we examined grade penalties of 2445 students majoring in biological science or closely related fields across many commonly taken courses, many of which were mandatory. We find that on average, students received grade penalties in the 12 most commonly taken science courses for biosciences students at our institution. We also find that women had more extreme grade penalties than men in seven of the 12 science classes we investigated.

Conclusions The pattern of grade penalties found in this study across a range of courses may result in long-term consequences for students' STEM self-concept. Furthermore, women's decisions to pursue STEM careers may be affected more by the grade penalty received in required science courses than men's at least partly, because their grade penalties are often larger. Finally, the grade penalty measure can be easily computed by institutions concerned with equity.

Keywords Gender, Bioscience, Course performance, Grade anomaly, Grade penalty, Grades

*Correspondence:
Alysa Malespina
alm417@pitt.edu
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA

Introduction

Research on gendered performance and persistence differences in science, technology, engineering, and mathematics (STEM) fields is important and has been conducted in the fields in which women are the least likely to earn degrees, such as physics or engineering (Cavallo et al., 2004; Gonzalez & Kuenzi, 2012; Henderson et al., 2020; Maries et al., 2022; Marshman et al., 2018; Nissen, 2019; Sawtelle et al., 2012; Van Dusen & Nissen, 2020; Whitcomb & Singh, 2020; Whitcomb et al., 2020). However, less research has been conducted on gender differences in fields in which women are not currently underrepresented, such as biology (Aguillon et al., 2020; Ballen, et al., 2017a, b; Haak et al., 2011). Concerns about gender equity in STEM education are not limited to considerations of the percentage of women in a STEM field, but includes the experiences of female students when they do participate (Cwik & Singh, 2022; Eaton et al., 2020; Good et al., 2012; Raelin et al., 2014). For example, past research has raised concerns about equity in the fields of biological science and medicine due to gender differences in classroom experiences (Ballen, et al., 2017a; Elliott, 2016; Grunspan et al., 2016), authorship over all career stages (Holman et al., 2018; Huang et al., 2020), and likelihood to stay in the field (Elliott, 2016; Maries et al., 2022; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). One particularly concerning trend for educators relates to this last point: women studying STEM, including bioscience majors, often leave their disciplines with higher grades than the men who remain in the program (Maries et al., 2022; Seymour & Hewitt, 1997; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). In this regard, men and women in the same programs have different outcomes, and these outcomes are not explained by grade differences.

In this research, we introduce a potential measure of gender inequities in the classroom: grade anomalies (Koester et al., 2016; Malespina & Singh, 2022a, b; Matz et al., 2017; Witteveen & Attewell, 2020). A grade anomaly (GA) is the difference between a student's grade point average (GPA) and their grade in a specific course. We investigated grade penalties, which describe grade anomalies in which a students' grade in a specific course is lower than their GPA. In this study, with equity and inclusion at the center, we investigate the following research questions regarding grade anomalies in the backdrop of gender inequities in biological sciences despite women outnumbering men.

- RQ1. For which of their courses do students majoring in biological sciences receive a "grade penalty" and for which courses do they receive a "grade bonus"?
- RQ2. To what extent do men and women have different grade anomalies in their STEM courses?
- RQ3. To what extent do gender differences in grade anomalies follow the same trends as gender differences in grades?

Theoretical background

Gender and biological sciences

Undergraduate students pursuing biological science and related majors often go on to pursue either graduate programs in biological sciences or medical school in the future. We discuss an investigation of the grade penalties for both groups of students here. Women earn undergraduate and graduate biosciences degrees at higher rates than men (Bachelor's Degrees in Physics & STEM Earned by Women, 2020). Moreover, because biosciences degree recipients make up over half of medical school applicants and matriculants in the United States (56% for the 2021–2022 academic year), there are more women from 2018 to 2021 who entered medical school than men (2021 *FACTS: Applicants and Matriculants Data*, 2021). However, gender equity is not the same as having an equal number of women present in a field. Researchers have investigated many ways in which women who pursue and earn degrees in biological science and related fields do not receive equitable treatment (for example, during careers in biological sciences or medicine or in biology classrooms).

After graduation, women who earn biological sciences degrees are less likely than men to work as scientists after receiving graduate degrees (Elliott, 2016). In addition, one recent study (Huang et al., 2020) found that in biological sciences, women tend to have shorter publishing careers (due to leaving the field) than men, and during their careers, men had higher annual publishing rates than women. The annual publishing rate for biological sciences had a larger gender difference favoring men than any other domain studied, including those in which women are underrepresented at all levels, such as physics and computer science (Huang et al., 2020). Researchers have estimated that biological science authorship in journal articles will not reach gender parity for another 25 years (Holman et al., 2018). In addition, despite their lower publishing rates, women in biological sciences also report feeling more stress arising from the pressure to publish than men do (Husemann et al., 2017). Thus, despite the representation of women in biological sciences, these trends hint at the inequitable culture in sciences that affect the professional advancement of these women. Gender inequities are not just experienced by women who pursue advanced degrees in biology, but those who pursue medical careers as well.

There are gender disparities in compensation and time to promotion for all academic medical specialties (Dandar & Lautenberger, 2021) as well as for physicians (Ly et al., 2016).

While they are pursuing their degrees, undergraduate women are less likely to participate in class discussions in biology classrooms (Eddy et al., 2014), are less likely to be viewed as knowledgeable by their peers (Grunspan et al., 2016), and tend to have lower exam grades than men in their introductory biology classes (Ballen, et al., 2017a). In addition, when women leave STEM disciplines, they often do so with higher grades than the men who remain in the program (Maries et al., 2022; Seymour & Hewitt, 1997; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019).

Gender differences in outcomes cannot be explained by a lack of representation in the classroom for students pursuing biological science and health-related majors, because women make up more than half of these populations. Some researchers offer potential explanations such as societal stereotypes and biases about who can excel in these disciplines (Bian et al., 2017; Bleeker & Jacobs, 2004; Eaton et al., 2020; Ganley et al., 2018; Leslie et al., 2015) and gender discrimination in hiring (Moss-Racusin et al., 2012). Others posit that differences in STEM motivational beliefs may explain these differences. One such motivational belief is self-efficacy. Self-efficacy is a student's belief in their ability to succeed in a domain, and often predicts learning outcomes (Bandura, 1997; Cavallo et al., 2004; Marshman et al., 2018; Nissen & Shemwell, 2016; Pintrich & De Groot, 1990; Raelin et al., 2014; Sawtelle et al., 2012; Van Dusen & Nissen, 2020). Another such belief is academic self-concept, which describes a long-term expectation of success that students hold regarding their academic abilities that depends on outside feedback, such as grades (Eccles & Wigfield, 2020; Gniewosz et al., 2015; Spence, 1983). Low academic self-concept from negative grade feedback may lead to lower future achievement and persistence, because it discourages student engagement in a domain (Spence, 1983). Gender differences in either self-efficacy or academic self-concept may account for some differences in outcomes, because women may opt out of majors or careers if they do not believe they can excel.

Broadly, researchers have proposed many partial explanations of gender differences in classroom experiences and professional trajectories of students pursuing biological sciences. Some of these, such as cultural biases, are not limited to our classrooms, but may also exist within them (Bian et al., 2017; Bleeker & Jacobs, 2004; Eaton et al., 2020; Ganley et al., 2018; Leslie et al., 2015). Other partial explanations, such as motivational beliefs about learning, may be primarily relevant in the classroom (Cavallo et al., 2004; Marshman et al., 2018; Nissen & Shemwell, 2016; Pintrich & De Groot, 1990; Raelin et al., 2014; Sawtelle et al., 2012; Van Dusen & Nissen, 2020). Regardless, to improve outcomes for women, instructors should aim to eliminate these inequities by taking steps to create an inclusive classroom culture and equitable learning environment.

Grade anomalies

One aspect of creating such a learning environment may be to reassure students that they can excel in any domain in which they choose to major. For example, students often have a fairly fixed view of what "kind" of student they are, e.g., students may endorse the idea that ``If I get As, I must be an A kind of person. If I get a C, I am a C kind of person" (Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). A GA, which is the difference between a student's GPA and their grade in a specific course, may challenge or reinforce students' ideas about what kind of student they are, and if they are capable of succeeding in their chosen domain. Many students who leave STEM explicitly cite lower grades than they are used to as a reason for doing so (Seymour & Hewitt, 1997; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). While students' raw grades are a useful measure, because they allow for direct comparison between students and because they are used by institutions to award scholarships and track student academic standing, we propose that using grade penalty in addition to raw grades gives researchers and instructors more insight into students' view of what "kind" of student they are.

Matz et al. (2017) conducted the first study we know of that focuses on average grade anomaly (AGA). A GA is the difference between a students' grade point average and grade in a particular course, and AGA is the mean grade anomaly for students who took a course. They used AGA, because it was perceived to be a better measure of how students view their comparative performance than their raw grades across different courses. They found that, at their institution, grade penalties were greater for STEM than non-STEM courses. A student who receives lower grades in their science courses than their humanities courses may take this as a sign that they are not capable of excelling in the sciences, even if the grades they earn are high enough for them to continue in their major (Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). Furthermore, within STEM courses, grade penalties were smaller for men than women. In particular, they found that physics courses had the largest grade penalty and largest gender difference in AGA. The Matz et al. (2017) study had similar findings but with a larger student sample across multiple institutions. Across five universities, STEM courses had larger grade penalties and larger gender differences in AGA that usually favored men.

Past work provides evidence that grade penalties are more common and extreme in STEM disciplines than in humanities or social science departments (Koester et al., 2016; Matz et al., 2017; Rask, 2010; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019; Tomkin & West, 2022), and women tend to have larger grade penalties than men in STEM subjects, especially in courses students take early in their time at university (Malespina & Singh, 2022a, b; Matz et al., 2017). Here, we present an investigation that focuses on GA in various courses for biological sciences majors in which we analyze data to study if these trends hold in a more homogeneous population of largely pre-health and pre-medical students at a single large university in the US, rather than combining students across institutions and many majors as in prior studies.

Materials and methods

Participants

Participants in this study were enrolled in biosciences majors or students with interest in health professions at a large research university in the US, which is a large, public, and urban institution. The student major breakdown was as follows: 37% Biological Sciences, 1% Bioinformatics/Computational Biology, 3% Ecology and Evolution, 6% Microbiology, 6% Molecular Biology, 30% Neuroscience, 4% Pharmacy, and 13% Rehabilitation Science. All major except for Neuroscience, Pharmacy, and Rehabilitation Science are offered through the Department of Biological Sciences. These students were chosen because of their similar course requirements, especially for large introductory science courses.

Grade data collected over 13 years. and we excluded courses that were taken during the summer semester. We excluded summer courses, because they are not a typical representation of courses at our institution. For example, many summer students do not primarily attend this institution, but are local students visiting home for the summer. In addition, summer courses are typically taught by graduate students, the class sizes are an order of magnitude smaller than those in the Fall and Spring semesters, many students work full time while taking an introductory course which is not as common during the Fall and Spring semesters. In addition, the class sizes are an order of magnitude smaller than those in the Fall and Spring semesters. In total, we had 2445 students who took a total of 89,560 courses. Of these courses, 33,543 were instances of courses of interest (for example, Biochemistry or Physics 1) and 56,017 were instances of other courses (for example, Medical Sociology). The student sample was 58.1% women and 41.9% men. Less than 0.1% did not list their gender, so they were excluded from the study due to small sample size. Students identified with the following races/ethnicities: 68% White, 17% Asian, 3% Hispanic/Latinx, 3% multiracial, 7% African American/Black, and 1% unknown or unspecified. This research was carried out in accordance with the principles outlined in the institution's Institutional Review Board (IRB) ethical policy, and de-identified demographic data were provided through university records.

Course selection

We chose to study courses that were taken by the largest number of students, excluding non-major electives (for example, "Introduction to Piano" or "Public Speaking") and courses that make up general education requirements. Thus, many courses were mandatory for students in the majors we focus on. The courses we chose are listed in Table 1, along with information about the year in which the students typically take the course. Not all courses were required for students in all the majors in our sample. Information about if a course was required, optional (i.e., an elective that count toward the major), or not required is included in Table 2. We would like to note that, though it is not required for most majors studied, Human Physiology met our criteria, because it is a commonly chosen elective for both Biology and Rehabilitation Science Students. In addition, students in this sample may take either algebra-based or calculus-based physics to fulfil their two-semester physics course requirement, but so few ($N=61$) students chose calculus-based physics that they were excluded from this study.

Measures

Course grade

Course grades were based on the 0–4 scale used at our university, with A = 4, B = 3, C = 2, D = 1, F = 0 or W (late withdrawal); the suffixes '+' and '−', respectively, add or subtract 0.25 grade points (e.g., B− = 2.75 and B+ = 3.25), except for the A+, which is reported as 4. We are unable to report grading schemes of each instructor, type of course (i.e., traditional lectures or active learning), or any other detailed course-level information due to the large number of courses sampled.

Demographic information

Students provide demographic information as part of university enrollment. Students were given the binary options "male" and "female" to identify their gender upon entering the university, although this conflates gender and sex (Schudson, 2021). This approach also marginalizes gender minority students (Traxler et al., 2016; Van Dusen & Nissen, 2020). We acknowledge the harm that such data collection practices cause (Traxler et al., 2016;

Table 1 Courses studied by year

Course	Department	1st	2ed	3ed	4th	≥ 5th
Calculus 1	Mathematics	**61**	22	9	6	2
Biology 1	Biology	**81**	14	3	1	1
Biology 2	Biology	**55**	34	7	2	2
Chemistry 1	Chemistry	**86**	10	2	1	1
Chemistry 2	Chemistry	**64**	28	4	3	1
Genetics	Biology	6	**44**	31	13	6
Organic Chemistry 1	Chemistry	7	**72**	15	4	2
Organic Chemistry 2	Chemistry	3	**50**	32	9	6
Physics 1	Physics	21	**37**	30	7	5
Human Physiology	Biology	2	18	**56**	19	5
Biochemistry	Biology	1	4	**48**	35	12
Physics 2	Physics	6	26	**60**	5	3

List of courses studied, the department that offers them, and the percentage of students in our sample who took each course in a given year. For example, 61% of students take Calculus 1 during their first year of university, and 22% of students take Calculus 1 during their second year. The year in which students take the course most often has its percentage of students in bold

Table 2 Course requirements by major

	Calc. 1	Biology		Chem		Genetics	Orgo Chem		H.P	B.C	Physics	
Major		**1**	**2**	**1**	**2**		**1**	**2**			**1**	**2**
Biology	R	R	R	R	R	R	R	R	O	R	R	R
Computational Biology	R	R	R	R	R	R	R			R	O	
E&E	R	R	R	R	R	R	R	R		R	R	R
Microbiology	R	R	R	R	R	R	R	R	O	R	R	R
Molecular Biology	R	R	R	R	R	R	R	R			R	R
Neuroscience	R	R	R	R	R		R	R	R	R	R	R
Pharmacy	R	R	R	R	R		R	R		R	O	O
Rehab. Science		R		R					O		R	

R designates a required course, O designates a course that can be taken for elective credit in the major, and no letter designates a course that does not fulfill any credits for the major. H.P. stands for Human Physiology, B.C. stands for Biochemistry, E&E stands for Ecology and Evolution, and Rehab stands for Rehabilitation

Van Dusen & Nissen, 2020), and we are pleased to report that our university has recently switched to collecting gender information using more than binary options.

Grade anomaly

GA was found by first finding each student's grade point average excluding the course of interest (GPA_{exc}). This was accomplished using the equation:

$$GPA_{exc} = \frac{GPA_c \times Units_c - Grade \times Units}{Units_c - Units},$$

where GPA_c is the student's cumulative GPA, $Units_c$ is the cumulative number of units the student has taken (including the semester the course of interest is taken), Grade is the grade the student received in an individual course, and Units is the number of units associated with an individual course. After finding GPA_{exc} we can calculate (GA by finding the difference between a student's GPA_{exc} and the grade received in that class:

$$GA = Grade - GPA_{exc}.$$

A negative GA corresponds to a course grade lower than a students' GPA in other classes and we call this a "grade penalty". A positive GA corresponds to a course grade higher than a students' GPA in other classes and we call this a "grade bonus". AGA is the mean of GAs for each course, and is the metric by which we compare courses.

Analyses

To characterize both AGA and grades, we found the sample size, mean, standard deviation, and standard error of each measurement for each course of interest. We calculated these statistics for women and men separately, and then for all students combined. We also compared the effect size of gender on both grade and GA, using Cohen's d to describe the size of the mean differences and unpaired t tests to evaluate the statistical robustness of the differences. Cohen's d is calculated as follows:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 - \sigma_2^2}},$$

where μ_1 and μ_2 are the means of the two groups, σ_1 and σ_2 are the standard deviations (Frey, 2018) and Cohen's d is considered small if $d \sim 0.2$, medium if $d \sim 0.5$, and large if $d \sim 0.8$ (Cohen, 1988). All analysis was conducted using R (R Core Team, 2020), using the package plotrix (Lemon, 2006) for descriptive statistics, lsr (Navarro, 2015) for effect sizes, and ggplot2 (Wickham, 2016) to create plots.

Results

RQ1: For which of their courses do biosciences students receive a ``grade penalty" and for which courses do they receive a ``grade bonus"?

To answer RQ1, we calculated AGA for the most popular courses taken by biological sciences students and students in related majors focusing on future careers in health professions. We show the descriptive statistics for both grades and AGA in Table 3 and Fig. 1. We find that students generally received grade penalties in all STEM courses we studied. One course, Human Physiology, has a smaller grade penalty than other courses, while Organic Chemistry 1 and 2 have much larger grade penalties than all other courses. For Organic Chemistry courses, students have an AGA of approximately 1, meaning that on average, students receive one full letter grade lower in these courses than in their other courses.

Table 3 Average grade anomalies (AGAs) and grades

Course	Department		AGA		Grade	
		N	Mean	SD	Mean	SD
Calculus 1	Mathematics	1018	−0.87	1.53	2.43	1.22
Biology 1	Biology	1740	−0.70	1.10	2.65	1.03
Biology 2	Biology	1630	−0.65	0.78	2.69	0.92
Chemistry 1	Chemistry	1746	−0.47	1.08	2.89	0.87
Chemistry 2	Chemistry	1673	−0.58	0.83	2.79	0.96
Genetics	Biology	775	−0.63	0.89	2.72	1.05
Organic Chemistry 1	Chemistry	1614	−1.00	1.08	2.40	1.15
Organic Chemistry 2	Chemistry	1135	−1.04	1.01	2.35	1.21
Physics 1	Physics	2685	−0.64	0.95	2.60	1.08
Human Physiology	Biology	1010	−0.30	0.78	3.11	0.93
Biochemistry	Biology	886	−0.68	0.89	2.73	1.06
Physics 2	Physics	1350	−0.58	0.77	2.79	0.97

Mean and standard deviation (SD) of average grade anomalies (AGA) and grades, as well as number of students (*N*) of each course of interest

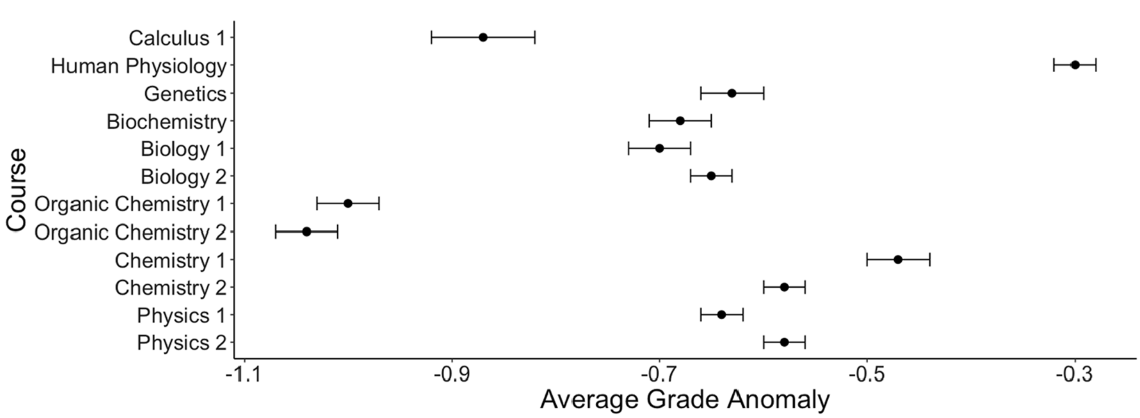

Fig. 1 Average grade anomalies (AGAs) for each course of interest. The error bars represent standard errors

There are no grade bonuses in the table, because not all courses that students take are included. For example, students may receive grade bonuses in general education courses, e.g., in social sciences and humanities as well as in laboratory courses.

RQ2: Do men and women have different "grade anomalies" in their STEM courses?

To investigate if there are differences in grade anomalies between men and women, we grouped students by their self-reported gender and calculated the AGA for both groups for each course of interest. We then calculated Cohen's *d* as a measure of effect size between the two groups (Frey, 2018), which can be seen in Table 4. Group differences can also be seen in Fig. 2. Women had statistically indistinguishable AGA outcomes to men in Calculus, Human Physiology, Genetics, Biochemistry, and Chemistry 2. By indistinguishable, we mean there is no statistically significant difference between men and women's AGAs in Table 4. Women did not have favorable AGA outcomes (e.g., smaller grade penalties) compared to men in any course. Men had favorable AGA outcomes compared to women in several courses, meaning there was a statistically significant difference in AGA between man and women, and men tended to have smaller grade penalties than women. These courses included Biology 1, Biology 2, Organic Chemistry 1, Organic Chemistry 2, Chemistry 1, Physics 1, and Physics 2.

Table 4 Average grade anomalies (AGAs), grades, and between-gender effect sizes for each course of interest

Course	*N*	Women					Men				Cohen's *d*	
		AGA		Grade			AGA		Grade			
		Mean	SD	Mean	SD	N	Mean	SD	Mean	SD	AGA	Grades
Calculus 1	527	− 0.85	1.55	2.53	1.17	491	− 0.89	1.52	2.32	1.26	− 0.03	− 0.18[b]
Biology 1	1078	− 0.76	1.16	2.63	1.03	662	− 0.60	1.00	2.68	1.05	0.14[b]	0.04
Biology 2	993	− 0.70	0.77	2.67	0.90	637	− 0.57	0.78	2.72	0.94	0.16[b]	0.05
Chem. 1	1046	− 0.51	1.11	2.87	0.87	700	− 0.40	1.04	2.91	0.88	0.10[c]	0.04
Chem. 2	986	− 0.59	0.80	2.80	0.92	687	− 0.56	0.88	2.76	1.02	0.03	− 0.05
Genetics	463	− 0.67	0.92	2.70	1/08	312	− 0.58	0.84	2.75	1.09	0.07	0.02
Organic 1	983	− 1.07	1.12	2.35	1.16	631	− 0.89	1.02	2.48	1.11	0.16[b]	0.11[c]
Organic 2	673	− 1.10	1.00	2.29	1.19	462	− 0.94	1.02	2.43	1.22	0.16[b]	0.11
Physics 1	1572	− 0.75	0.94	2.56	1.04	1113	− 0.49	0.95	2.67	1.14	0.28[a]	0.11[b]
Hum. Phy	626	− 0.27	0.80	3.11	0.94	384	− 0.32	0.76	3.12	0.92	0.07	0.02
Biochem	508	− 0.70	0.88	2.72	1.04	378	− 0.64	0.91	2.75	1.09	0.07	0.02
Physics 2	776	− 0.61	0.68	2.78	0.89	574	− 0.54	0.86	2.80	1.06	0.09	0.02

A bold Cohen's *d* signifies that a *t* test showed significant differences between men and women. The following abbreviations are used: Human Physiology (Hum. Phy.), Biochemistry (Biochem.), Organic Chemistry (Organic), and Chemistry (Chem)

[a] indicates that $p < 0.001$

[b] indicated that $p < 0.01$

[c] indicates $p < 0.05$

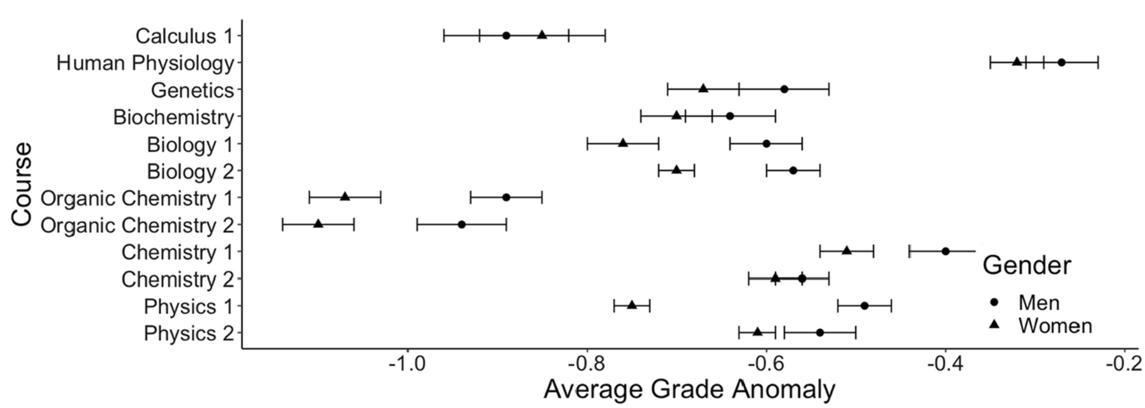

Fig. 2 Average grade anomalies (AGAs) for each course of interest, by gender. The error bars represent standard errors

RQ3: Do gender differences in "grade anomalies" follow the same trends as gender differences in grades?

There are many courses for which there was no statistically significant gender differences in grades or grade anomalies, such as Human Physiology, Genetics, Biochemistry, Chemistry 2, and Physics 2, which can be seen in Table 4. With the exception of Chemistry 2, these are all courses that students tend to take in their second year of university or later. There was one course in which the gender difference was both statistically significant and similar for grades and AGA: Organic Chemistry 1. In this course, Cohen's d between genders was similar for Grade and AGA, as shown in Table 4. These similar but significant gaps favoring men can also be seen in Figs. 2 and 3. For the aforementioned six courses, AGA and grades provide similar information.

There are also courses that show difference trends in AGA versus raw grades. For example, Table 4 reveals that Biology 1, Biology 2, Organic Chemistry 2, and Chemistry 1 have no statistically significant gender difference in grades, but all have a statistically significant difference between men and women when in AGA. Similarly, Table 4 also shows that the gender difference in AGA ($d = 0.26$) is larger than in raw grades ($d = 0.11$) for Physics 1, even if the gender difference is statistically significant for both measures. Comparing Figs. 2 and 3, it is clear that the gender differences are larger for AGA than raw grades for all of these courses.

There is one course that has a larger gender difference in grades than in AGA: Calculus 1, which can be seen in Table 4. In this course, women tend to have indistinguishable AGAs to men, but statistically significantly higher grades than men.

Discussion

Our primary finding is that there are average grade penalties in all courses studied for undergraduate biosciences majors or those with interest in health-related professions. First, we discuss why these grade penalties can be harmful. According to some studies by Seymour (1997; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019), lower than expected grades, even in a single course, can be a catalyst for students to leave STEM majors. They found that this does not just include D and F grades or withdrawal from the course, but grades that were high enough to continue the program that did not meet a student's personal expectations (Seymour & Hewitt, 1997; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). They also found that this was a particular issue among high-achieving students, who were more likely to endorse perfectionism and feeling like their identity as "good STEM students" was threatened by B's and C's, or even a low grade on a single exam (Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). At out institution, 16.4% (the portion was very similar for men and women) of students studying Biosciences or Neuroscience leave their major (Maries et al., 2022). Thus, we believe that the courses that have the largest grade penalties, in this case Organic Chemistry 1 and 2, are the courses highly likely to push students out of these disciplines or cause them to question their abilities.

In addition to finding evidence of average grade penalties in some courses overall for all students taken together, we also find evidence of gender differences in average grade penalties in over half of the courses studied, particularly Biology 1 and 2, Organic Chemistry 1 and 2, Chemistry 1, and Physics 1 and 2. In particular, we find Organic Chemistry 1 and 2 concerning, because these courses have the largest average grade penalty women receive, so it is likely to stand out as a unique anomaly, and may impact students' academic self-concept (e.g.,

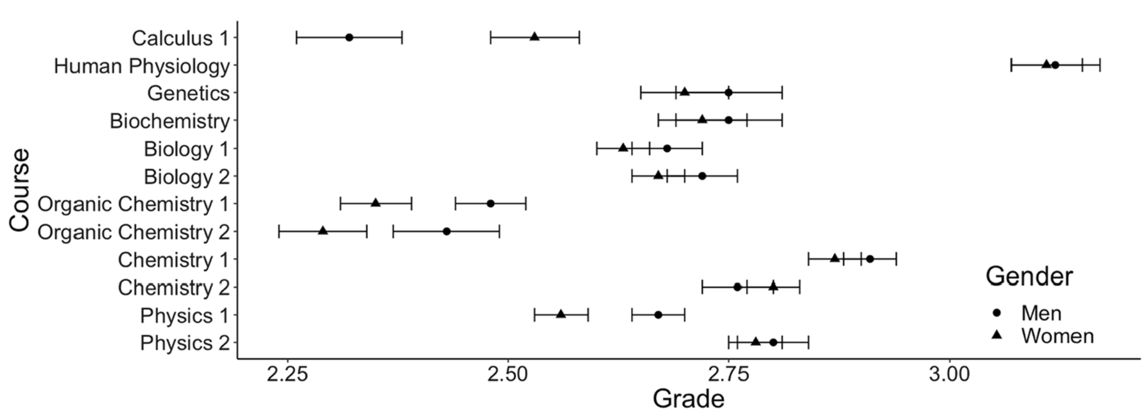

Fig. 3 Average grade anomalies (AGAs) for each course of interest, by gender. The error bars represent standard errors

their long-term expectation of academic success) more than other courses.

In prior studies, women report feeling more demoralized than men when they receive low grades, and cite more worry over not understanding material even if they receive A's, B's, or C's (all of which are grades that allow students to continue in most programs) (Goodman, 2002; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). This trend has been found to be particularly strong among high-achieving women (Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019). Research shows that women often tend to earn higher grades than men with the same standardized test scores (Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019; Voyer & Voyer, 2014). Women in this sample tended to have higher GPAs than men, especially in non-science courses, so a low grade in a science course may be a larger deviation from women's personal norms. Because women are often more accustomed to higher grades, they may have more concern about grades that are lower than what they are accustomed to, or they may compare their relatively low STEM grades and view themselves as less able to succeed in biology than a subject that gives them the recognition for their work that they are accustomed to (Seymour & Hewitt, 1997; Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education, 2019).

We find that grade anomalies and raw grade data do not always reveal the same trends. Some courses have larger gender differences in AGA than in grades, such as Biology 1 and 2, Chemistry 1, and Physics 1. This speaks to the usefulness of tracking both AGA and grades of the students. An instructor may think that a small or negligible grade difference by gender in their course implies that there is gender equity in their classroom based upon grade outcomes alone, but without knowing the gender differences in AGA, the instructor will not recognize the grade penalty and how those grades are perceived by female and male students. Understanding both grades and AGA differences may allow instructors to understand both of these classroom-level inequities. In the study discussed here, there is one course that has larger gender differences in grades than in AGA: Calculus 1. We find this course concerning because of the average grade penalty for all students, particularly because this course is most often taken by first-year students, who are more likely to have an academic self-concept that is fragile (Eccles & Wigfield, 2020).

These findings are similar to past work on GA in a variety of ways. For example, past work which included multiple universities in their sample also found that most STEM lecture courses had grade penalties (Koester et al., 2016; Matz et al., 2017). Math and chemistry courses often had the larger grade penalties than physics and biology, which is similar to our study (Matz et al., 2017). Measuring AGA in addition to grades may be a useful way to find inequities in the learning environment. Measuring grades and gendered grade differences is both valuable for and accessible to individual instructors, but grade anomalies may be useful to departments concerned about students' retention over longer periods and finding which courses may be particularly discouraging to students.

Conclusion

In this investigation, we found that grade penalties exist for all of the courses we studied for undergraduate biosciences majors or those with interest in health-related professions. Furthermore, seven courses had an AGA (larger grade bonuses or smaller grade penalties) that favored men over women, while five courses had an AGA that did not favor either gender. This gender inequity in AGA raises particular concern about the need for an equitable learning environment.

These findings are also valuable, because they provide evidence that courses in STEM departments in general tend to have grade penalties. This supports prior research on grade anomalies in a more homogeneous population, bioscience majors at one university.

Limitations and future research

Although we have evidence of grade penalties in the studied courses as well as gendered GA differences, we did not have access to syllabi or other information about individual courses offered over the 13-year period of data collection. Therefore, we are not able to pinpoint specific practices that may lead to grade penalties, grade bonuses, or gender inequities at our institution. However, based upon the structure of most courses currently offered that were discussed here, it is reasonable to assume that most of these courses have focused on teaching in a traditional, lecture-based, and exam-reliant format.

In addition, this research is based at a primarily white, large, public university in the US. While our results are likely to generalize to similar institutions, we do not know what patterns of grade anomalies and potential gender differences in them exist at smaller liberal arts colleges, minority-serving institutions, or community colleges in the US. In addition, conducting research at a diverse range of institutions in different countries, as well as a focus on how GA affects students from a variety of underrepresented groups, will help us more fully understand how grade anomalies differ for a range of students.

Finally, grade anomalies have the potential to be an easily accessible and useful tool for universities to pinpoint courses that may be particularly discouraging to students or effect retention. One way to realize this potential may be future studies that focus on institutional awareness and potential amelioration of grade penalties.

Abbreviations

AGA	Average grade anomaly
EVT	Expectancy value theory
GA	Grade anomaly
SD	Standard deviation
SEVT	Situated expectancy value theory
STEM	Science, technology, engineering, and mathematics

Acknowledgements
The authors would like to thank Robert Devaty for critical feedback on the manuscript.

Author contributions
AM contributed to analysis and interpretation of data, as well as writing and revision of manuscript. CS contributed to the conception and design of research, acquisition and interpretation of data, and revision of the manuscript. All authors read and approved the final manuscript.

Funding
This research was funded by National Science Foundation Award No. DUE- 1524575.

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

Declarations

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

Received: 30 July 2022 Accepted: 11 January 2023
Published online: 14 February 2023

References

Aguillon, S. M., Siegmund, G.-F., Petipas, R. H., Drake, A. G., Cotner, S., & Ballen, C. J. (2020). Gender differences in student participation in an active-learning classroom. *CBE Life Sciences Education, 19*(2), 1–10. https://doi.org/10.1187/cbe.19-03-0048
Bachelor's Degrees in Physics and STEM Earned by Women. (2020). http://www.aps.org/programs/education/statistics/womenstem.cfm
Ballen, C. J., Salehi, S., & Cotner, S. (2017a). Exams disadvantage women in introductory biology. *PLoS ONE, 12*(10), e0186419–e0186419. https://doi.org/10.1371/journal.pone.0186419
Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017b). Enhancing diversity in undergraduate science: self-efficacy drives performance gains with active learning. *CBE Life Sciences Education, 16*(4), ar56. https://doi.org/10.1187/cbe.16-12-0344
Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.
Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science, 355*(6323), 389–391. https://doi.org/10.1126/science.aah6524
Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96*(1), 97–109. https://doi.org/10.1037/0022-0663.96.1.97
Cavallo, A. M. L., Potter, W. H., & Rozman, M. (2004). Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors. *School Science and Mathematics, 104*(6), 288–300. https://doi.org/10.1111/j.1949-8594.2004.tb18000.x
Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates.
Cwik, S., & Singh, C. (2022). Students' sense of belonging in introductory physics course for bioscience majors predicts their grade. *Physical Review Physics Education Research, 18*(1), 010139. https://doi.org/10.1103/PhysRevPhysEducRes.18.010139
Dandar, V. M., & Lautenberger, D. M. (2021). *Exploring Faculty Salary Equity at U.S. Medical Schools by Gender and Race/Ethnicity*. A. o. A. M. Colleges.
Eaton, A. A., Saunders, J. F., Jacobson, R. K., & West, K. (2020). How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates. *Sex Roles, 82*(3–4), 127–141. https://doi.org/10.1007/s11199-019-01052-w
Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859
Eddy, S. L., Brownell, S. E., & Wenderoth, M. P. (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE Life Sciences Education, 13*(3), 478. https://doi.org/10.1187/cbe.13-10-0204
Elliott, S. L. (2016). From the editor-in-chief: questions of gender equity in the undergraduate biology classroom. *Journal of Microbiology & Biology Education, 17*(2), 186–188. https://doi.org/10.1128/jmbe.v17i2.1136
2021 FACTS: Applicants and Matriculants Data. (2021). Association of American Medical Colleges. Retrieved June 28 from https://www.aamc.org/data-reports/students-residents/interactive-data/2021-facts-applicants-and-matriculants-data
Frey, B. B. (Ed.). (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc.
Ganley, C. M., George, C. E., Cimpian, J. R., & Makowski, M. B. (2018). Gender equity in college majors: Looking beyond the STEM/Non-STEM dichotomy for answers regarding female participation. *American Educational Research Journal, 55*(3), 453–487. https://doi.org/10.3102/0002831217740221
Gniewosz, B., Eccles, J. S., & Noack, P. (2015). Early adolescents' development of academic self-concept and intrinsic task value: The role of contextual feedback. *Journal of Research on Adolescence, 25*(3), 459–473. https://doi.org/10.1111/jora.12140
Gonzalez, H. B., & Kuenzi, J. J. (2012). *Science, Technology, Engineering, and Mathematics (STEM) Education A Primer* (CRS Report No. R42530).
Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700. https://doi.org/10.1037/a0026659
Goodman, I. F. (2002). Final Report of the Women's Experiences in College Engineering (WECE) Project. *Online Submission*.
Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLoS ONE, 11*(2), e0148405. https://doi.org/10.1371/journal.pone.0148405
Haak, D. C., Hillerislambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science (american Association for the Advancement of Science), 332*(6034), 1213–1216. https://doi.org/10.1126/science.1204820
Henderson, R., Sawtelle, V., & Nissen, J. M. (2020). Gender & self-efficacy: A call to physics educators. *The Physics Teacher, 58*(5), 345–348. https://doi.org/10.1119/1.5145533
Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? *PLOS Biology, 16*(4), e2004956. https://doi.org/10.1371/journal.pbio.2004956
Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences, 117*(9), 4609–4616. https://doi.org/10.1073/pnas.1914221117
Husemann, M., Rogers, R., Meyer, S., & Habel, J. C. (2017). "Publicationism" and scientists' satisfaction depend on gender, career stage and the wider

academic system. *Palgrave Communications, 3*, 17032. https://doi.org/10.1057/palcomms.2017.32
Koester, B. P., Grom, G., & McKay, T. A. (2016). Patterns of Gendered Performance Difference in Introductory STEM Courses. https://arxiv.org/pdf/1608.07565.pdf
Lemon, J. (2006). Plotrix: A package in the red light district of R. *R-News, 6*(4), 12.
Leslie, S.-J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science, 347*(6219), 262–265. https://doi.org/10.1126/science.1261375
Ly, D. P., Seabury, S. A., & Jena, A. B. (2016). Differences in incomes of physicians in the United States by race and sex: observational study. *BMJ, 353*, i2923. https://doi.org/10.1136/bmj.i2923
Malespina, A., & Singh, C. (2022a). Gender differences in grades versus grade penalties: Are grade anomalies more detrimental for female physics majors? *Physical Review Physics Education Research, 18*(2), 020127. https://doi.org/10.1103/PhysRevPhysEducRes.18.020127
Malespina, A., & Singh, C. (2022b). Impact of grade penalty in first-year foundational science courses on female engineering majors. *International Journal of Engineering Education, 38*(4), 1021. https://www.ijee.ie/latestissues/Vol38-4/13_ijee4218.pdf.
Maries, A., Whitcomb, K., & Singh, C. (2022). Gender inequities throughout STEM. *Journal of College Science Teaching, 51*(3), 27–36. https://www.nsta.org/journal-college-science-teaching/journal-college-science-teaching-januaryfebruary-2022/gender
Marshman, E., Kalender, Z. Y., Schunn, C., Nokes-Malach, T., & Singh, C. (2018). A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences. *Canadian Journal of Physics, 96*(4), 391–405. https://doi.org/10.1139/cjp-2017-0185.
Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five research universities. *AERA Open, 3*(4), 2332858417743754. https://doi.org/10.1177/2332858417743754
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 of the United States of America, 109*(41), 16474. https://doi.org/10.1073/pnas.1211286109
Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners* University of New South Wales. https://learningstatisticswithr.com
Nissen, J. M. (2019). Gender differences in self-efficacy states in high school physics. *Physical Review Physics Education Research, 15*(1), 013102. https://doi.org/10.1103/PhysRevPhysEducRes.15.013102
Nissen, J. M., & Shemwell, J. T. (2016). Gender, experience, and self-efficacy in introductory physics. *Physical Review Physics Education Research, 12*(2), 020105. https://doi.org/10.1103/PhysRevPhysEducRes.12.020105
Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33
R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Raelin, J. A., Bailey, M. B., Hamann, J., Pendleton, L. K., Reisberg, R., & Whitman, D. L. (2014). The gendered effect of cooperative education, contextual support, and self-efficacy on undergraduate retention. *Journal of Engineering Education, 103*(4), 599–624. https://doi.org/10.1002/jee.20060
Rask, K. (2010). Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review, 29*(6), 892–900. https://doi.org/10.1016/j.econedurev.2010.06.013
Sawtelle, V., Brewe, E., & Kramer, L. H. (2012). Exploring the relationship between self-efficacy and retention in introductory physics. *Journal of Research in Science Teaching, 49*(9), 1096–1121. https://doi.org/10.1002/tea.21050
Schudson, Z. C. (2021). Psychology's Stewardship of Gender/Sex. *Perspectives on Psychological Science, 16*(6), 1105–1112. https://doi.org/10.1177/17456916211018462
Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: why undergraduates leave the sciences*. Westview.
Spence, J. T. (1983). *Achievement and achievement motives: Psychological and sociological approaches*. Freeman.
Talking about Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education. (2019). E. Seymour & A.-B. Hunter (Eds). Springer.
Tomkin, J. H., & West, M. (2022). STEM courses are harder: evaluating inter-course grading disparities with a calibrated GPA model. *International Journal of STEM Education, 9*(1), 27. https://doi.org/10.1186/s40594-022-00343-1
Traxler, A. L., Cid, X. C., Blue, J., & Barthelemy, R. (2016). Enriching gender in physics education research: A binary past and a complex future. *Physical Review Physics Education Research, 12*(2), 020114. https://doi.org/10.1103/PhysRevPhysEducRes.12.020114
Van Dusen, B., & Nissen, J. (2020). Equity in college physics student learning: A critical quantitative intersectionality investigation. *Journal of Research in Science Teaching, 57*(1), 33–57. https://doi.org/10.1002/tea.21584
Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204. https://doi.org/10.1037/a0036620
Whitcomb, K. M., Kalender, Z. Y., Nokes-Malach, T. J., Schunn, C. D., & Singh, C. (2020). Comparison of self-efficacy and performance of engineering undergraduate women and men. *International Journal of Engineering Education, 36*(6), 1996–2014. https://www.ijee.ie/1atestissues/Vol36-6/24_ijee4004.pdf.
Whitcomb, K. M., & Singh, C. (2020). For physics majors, gender differences in introductory physics do not inform future physics performance. *European Journal of Physics, 41*(6), 065701. https://doi.org/10.1088/1361-6404/ab9f1d
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer. https://ggplot2.tidyverse.org
Witteveen, D., & Attewell, P. (2020). The STEM grading penalty: An alternative to the "leaky pipeline" hypothesis. *Science Education, 104*(4), 714–735. https://doi.org/10.1002/sce.21580

Publisher's Note