International Journal of Engineering Education Vol. 38, No. 4, pp. 1021–1031, 2022
Printed in Great Britain

0949-149X/91 $3.00+0.00

Impact of Grade Penalty in First-Year Foundational Science Courses on Female Engineering Majors*

ALYSA MALESPINA and CHANDRALEKHA SINGH
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, 15260, USA. E-mail: ALM417@pitt.edu, clsingh@pitt.edu

Many frameworks have been put forth for why women continue to be underrepresented in engineering professions. Here, we introduce a framework that posits that grade penalty in first year foundational science courses for engineering majors may be particularly damaging to female students who do not have role models and are questioning whether they have what it takes to excel in an engineering major and career due to pervasive stereotypes. In order to quantify grade penalty, we define Average Grade Anomaly (AGA) as the difference between a student's grade in a course under consideration and their grade point average (GPA) in all other classes thus far. An AGA lower than students' expected grade based on their GPA is a grade penalty and higher than expected grade is a grade bonus. Our framework posits that female engineering majors are more likely to be negatively impacted by a grade penalty in their first-year foundational science courses since their academic self-concept as an engineering major hinges on them securing a certain grade. In the study presented here, we examine AGAs of 6,028 first-year engineering students across a number of required courses. We find that students tend to receive grade bonuses in engineering and English composition courses, and grade penalties in physics, chemistry, and math courses. These courses with grade penalties tend to be large, lecture-based courses. We also find that in physics courses, women have larger grade penalties than men, whereas in chemistry and math, men have larger grade penalties. Thus, physics courses may be most damaging to women out of all of the courses in which they receive grade penalty. We hypothesize that women's decisions to pursue an engineering major and career may be affected more by the grade penalty received in foundational science courses than men's due to societal stereotypes about who can excel in engineering and access to other coping mechanisms that may help to rationalize lower-than-expected grades. Furthermore, the grade penalty measure can be easily computed by the engineering programs concerned with equity. Finally, we provide recommendations for how engineering programs may mitigate grade penalties in the foundational science courses, which may be particularly damaging to women.

Keywords: gender; equity; grades; grade penalty; grade anomaly

1. Introduction and Theoretical Framework

Gender differences in performance and persistence in science, technology, engineering, and mathematics (STEM) are well-studied phenomena, especially in engineering [1–5]. When women leave STEM disciplines, they often do so with higher grades than men who remain in the program [6–8]. Women are more drastically underrepresented in engineering than many other STEM disciplines [1, 2, 5], so focusing on retention is particularly important for this field. If women are leaving engineering programs with grades that meet or exceed minimum requirements [7, 8], it is likely that many students who would succeed in engineering careers will pursue other professional paths.

There are many partial explanations regarding why women who are meeting or exceeding the requirements of their programs leave. These include societal stereotypes and biases about who can excel in these disciplines that discourage women from pursuing STEM careers [9–17], gender discrimination in hiring [18], and differences in STEM motivational beliefs such as self-efficacy and sense of belonging [2, 3, 19–26]. We have been focusing on how to improve equity and inclusion in STEM, with a particular focus on motivational factors [27–32].

Here, we focus on first-year engineering majors and introduce a framework that posits that grade penalty in first year foundational science courses for engineering majors may be particularly damaging to female students who do not have role models and are questioning whether they have what it takes to excel in an engineering major and career. We focus on grade anomaly as a tool to help understand gender differences in first year engineering grades. Grade anomaly is the difference between a student's grade in a course of interest and their GPA in all other courses thus far excluding that course. We divide grade anomalies into "bonuses" and "penalties". A course in which most students earn a lower grade than usual has a grade penalty, while a course in which most students earn a higher grade than usual has a grade bonus.

We propose that grade anomaly is a potential measure of students' academic self-concept which is easy to track, through institutional grade data. Our framework uses grade penalty as a central construct instead of grade because students' self-concept is

* Accepted 24 January 2022.

tied to what type of student *they think they are*. Low academic self-concept can be particularly detrimental to women [33–36], so measuring it may be useful for tracking classeoom equity. Students tend to have a fairly fixed view of what "kind" of student they are: for example, 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" [7, 8]. Grade anomalies may challenge or reinforce students' ideas about what kind of student they are, and if they are capable of succeeding in their chosen major. Many students who leave STEM majors explicitly cite lower grades than they are used to as a reason for doing so [7, 8]. Grade penalties are more common and extreme in STEM disciplines than in humanities or social science departments [8, 37–39], and women are more affected by these grade penalties due to stereotypes and lack of role models, and they are more likely to leave their majors or career aspirations with fewer and smaller grade penalties than men are [8, 37].

In this paper, we use Situated Expectancy Value Theory (SEVT), studies about why students leave STEM, and previous work on grade anomalies to explore if grade anomalies in first-year foundational courses affect male and female engineering majors differently, making grade anomalies an equity issue in engineering. We also posit that grade anomaly may be a better measure of self-concept [36] than raw grades and students are more likely to question whether they should continue in disciplines in which the foundational courses involve grade penalty because it is a unique measure of "within-student" frame of reference (i.e., students are comparing their own grades across different courses as opposed to comparing their grades with others) [34].

1.1 Prior Work on Grade Anomaly

Several studies [38–40] have utilized "grade anomaly", the difference between a student's GPA excluding a course of interest and their grade in all courses thus far. Huberth *et al.* [40] developed this measure, but Koester *et al.* [39] conducted the first study we know of that focuses on average grade anomaly (AGA). 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. Further, 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 researchers theorized that large grade penalties and gender differences may be partially attributed to high-stakes assessments [41–44] and stereotype threat [45]. High-stakes assessments like exams are shown to have larger gender differences in grades than low-stakes assessments like problem sets or quizzes [41–44], while stereotype threat (a students' feeling of risk associated with confirming a negative stereotype, for example a woman who fears confirming the stereotype that women are bad at math) takes up cognitive resources of students from underrepresented groups [45]. The Matz *et al.* [38] 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. Their study also raised concerns over high-stakes assessments. They emphasized that large grade anomalies often reflect grading decisions made by instructors, rather than being an accurate measure of student learning.

Thus, past work provides evidence for the existence of grade anomalies in STEM courses, and the existence of gender differences in these anomalies. Here, we present an investigation that focuses on grade anomaly in first-year foundational courses in engineering in which we analyze data to study if these trends hold in a more homogeneous population of first-year engineering students at a single university, rather than combining students across institutions and majors. This focus on first-year engineering students can help control for potential confounding factors. This study is particularly important because first year foundational courses in engineering play a critical role in students' short and long-term professional trajectories.

1.2 Situated Expectancy Value Theory Framework

Expectancy Value Theory (EVT) [33] and Situated Expectancy Value Theory (SEVT) [36] are frameworks to understand student achievement, persistence, and choice of tasks in a domain (e.g., engineering). EVT posits that performance and persistence is determined by someone's expectation of success and the extent to which they value that task. If a student expects they will succeed in a task and believes that task will be valuable to them (for personal interest, as a path to achieve another goal, etc.) they are more likely to pursue that task. If they do not expect to succeed and do not value a task, they are unlikely to attempt it. Here, we will focus primarily on student expectancies, though value is also important to understanding why some students may persist while others do not.

Expectancies are a combination of academic self-concept, expectations for success, and perceptions of task difficulty [33–36]. Academic self-concept is the most stable and the least task and domain-dependent of the three, and it is based primarily on grades and outside (e.g., from parents, peers, and instructors) feedback [33–36, 46]. Grades inform academic self-concept as both an external ("How good at math am I compared to other students?") and internal ("How good am I at math compared to English?") frame of reference [33–35].

Expectations of success are more domain and task specific, and refer to a student's belief in their ability to complete a specific task, which will include considerations such as skill in the subject, time allotted, and experience in a subject ([33–36]. Expectancy for success most closely relates to Bandura's theory of self-efficacy [35, 36, 47]. A student may have a positive academic self-concept in math, but may have low expectancy for success if they take a math test on very new material they have not had adequate time to learn. The third expectancy, perceptions of task difficulty, is more straightforward; most students have less faith in their ability to do well on an exam if their peers have reported it to be particularly difficult [33].

In EVT, the three expectancy concepts were collapsed into one factor. However, the updated framework, SEVT, has called to separate these three concepts [36]. According to Eccles and Wigfield [36], combining academic self-concept, expectancies for success, and perceived task difficulty has led to a lack of understanding of the unique developmental mechanisms of each and how the three concepts relate.

We propose that grade anomalies can act as a proxy for student's internal frame of reference. Additionally, past research has found that during times of transition, the usually-stable academic self-concept becomes more dependent on grade feedback and less dependent on outsider (e.g., parental) feedback [34]. We study first-year engineering students because they are more likely to have an unstable academic self-concept due to the transition from high school to university and can be impacted by their performance in first-year courses in college.

2. Research Questions

We aim to answer the following research questions regarding grade anomalies for first-year engineering students:

RQ1. For which of their first-year courses do engineering students receive a "grade penalty" and for which courses do they receive a "grade bonus"?

RQ2. Do male and female engineering students have different "grade anomalies" in their first-year courses?

RQ3. If there are gender differences in "grade anomalies", do they follow the same trends as gendered grade differences?

3. Methodology

3.1 Participants

This study takes place at the University of Pittsburgh, a large (19,017 degree-seeking undergraduates in 2019), public, urban, predominantly-white institution (78% in 2019) in the northeastern United States [48]. The participants were students enrolled in the School of Engineering, who were in their first or second semester at the university, and took calculus-based physics 1 as well as other mandatory first year courses taken by engineering majors between Spring semester of 2006 and Fall semester of 2019. We excluded courses that were taken during the summer semester. This left us with 6,028 engineering majors who took 48,116 courses during their first and second semesters of college. The sample was 29.9% women and 70.1% men. Students who did not list their gender were excluded from the study as they made up less than 0.1% of the sample. Students identified with the following races/ethnicities: 79% White, 9% Asian, 3% Hispanic/Latinx, 3% multiracial, 5% African American/ Black, and 1% unknown or unspecified. Demographic data were provided through deidentified university records. This research was carried out in accordance with the principles outlined in the University of Pittsburgh Institutional Review Board (IRB) ethical policy.

3.2 Procedures

We chose the courses to include in our investigation by reviewing the engineering first-year curriculum, which is standardized for students at our institution, and confirming that the majority of students took these courses during their first or second semester of college. Chemistry 1 and Chemistry 2 were offered by the Department of Chemistry, but are reserved for engineering students. Physics 1 and Physics 2 were offered by the Department of Physics and Astronomy, and Calculus 1 and Calculus 2 were offered by the Department of Mathematics. All other courses were offered by the school of engineering. Some courses, such as "Composition Seminar", a writing course for first-semester engineering students, have fewer students than other courses. This is a result of changes in the writing part of the curriculum over the thirteen-year data collection period. All curriculum changes resulted in a very general requirement becoming more specific (for example, students were once required to take one general education course in humanities before graduation, but are now required to take "Composition Seminar" during their first semester). In situations involving these cases, we only included the newer and more specific requirement in our analysis.

3.3 Measures

3.3.1 Course Grade

Course grades were based on the 0–4 scale used at our university, with A = 4, B = 3, C = 2, D = 1, F = 0 or W (late withdrawal), where the suffixes '+' and '–', respectively, add or subtract 0.25 grade points (e.g., B– = 2.75 and B+ = 3.25), except for the A+, which is also reported as 4. Due to the wide variety of courses sampled and the time frame of the study, 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.

3.3.2 Grade Anomaly (GA)

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

$$GPA_{exc} = \frac{(GPA_c \times Units_c) - (Grade \times Units)}{Units_c - Units} \quad (1)$$

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

$$GA = GPA_{exc} - Grade. \quad (2)$$

Thus, if a student has a negative GA, that means they received a lower grade in the course of interest than in their other classes. We call this a "grade penalty". If a student has a positive GA, that means they received a higher grade in the course of interest than in their other classes. We call this a "grade bonus".

3.4 Analysis

To characterize both average grade anomaly (AGA) and grades, we found the sample size, mean, standard deviation, and standard error of each measure 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 grade anomaly, 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:

$$d = \frac{\mu_1 - \mu_2}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}} \quad (3)$$

where μ_1 and μ_2 are the means of the two groups σ_1 and σ_2 are the standard deviations [49]. Cohen's d is considered small if $d \sim 0.2$, medium if $d \sim 0.5$, and large if $d \sim 0.8$ [50]. We used a significance level of 0.05 in the t-tests and as a balance between Type I (falsely rejecting a null hypothesis) and Type II (falsely accepting a null hypothesis) errors [49]. All analysis was conducted using R [51], using the package plotrix [52] for descriptive statistics, lsr [53] for effect sizes, and ggplot2 [54] to create plots.

4. Results

4.1 For which of their first-year courses do Engineering Students receive a "Grade Penalty" and for which do they receive a "Grade Bonus"?

To answer RQ1, we calculated average grade anomaly (AGA) for each course engineering students are required to take during their first and second semester at the university. We show the descriptive statistics for both grades and AGA in Table 1 and Fig. 1. The largest student sample can be found for Physics 1, because this is the class we used to select

Table 1. Grades and AGA for each course of interest, including the department that offered it and semester in which it was offered. The following words are abbreviated: Seminar (Sem.), Computing (Comp.), Communication (Com.), Professional (Prof.), and Standard Deviation (SD)

				Grade		AGA	
Course	**Semester**	**Department**	**N**	**Mean**	**SD**	**Mean**	**SD**
Composition Sem.	Fall	English	787	3.51	0.72	0.48	0.70
Intro to Analysis	Fall	Engineering	5352	3.52	0.53	0.55	0.54
Chemistry 1	Fall	Engineering	4185	2.55	0.99	–0.62	0.85
Physics 1	Fall	Physics	6022	2.73	0.79	–0.47	0.72
Calculus 1	Fall	Mathematics	4381	2.81	1.01	–0.23	1.08
Intro to Comp.	Spring	Engineering	4672	3.29	0.73	0.27	0.62
Prof. Com.	Spring	English	338	3.81	0.31	0.63	0.55
Chemistry 2	Spring	Chemistry	3385	2.52	0.91	–0.59	0.68
Physics 2	Spring	Physics	4762	2.58	0.91	–0.58	0.70
Calculus 2	Spring	Mathematics	4601	2.63	1.13	–0.48	0.99

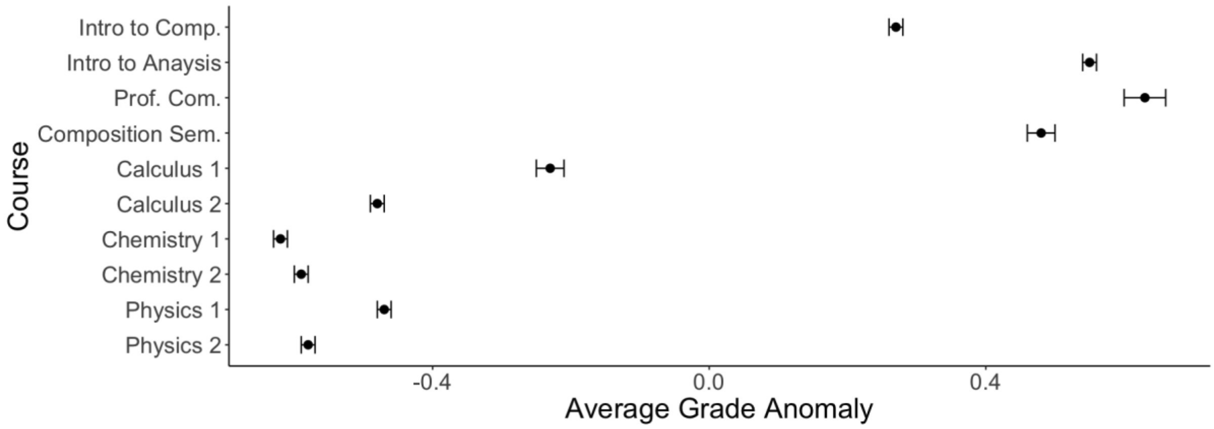

Fig. 1. Average grade anomaly (AGA) of all students by course. The following words are abbreviated: Seminar (Sem.), Computing (Comp.), Communication (Com.), and Professional (Prof.). Ranges represent standard error of the mean.

engineering students for our sample. We find that students generally receive grade penalties in the courses offered by the departments of Chemistry, Mathematics, and Physics, while students receive grade bonuses in first-year courses offered by the department of English and School of Engineering. The courses in which students receive a grade penalty are (in order from largest to smallest penalty): Chemistry 1, Chemistry 2 and Physics 2 (tie), Calculus 2 and Physics 1 (tie), and Calculus 1. The courses that students receive a grade bonus are (in order from smallest to largest bonus): Introduction to Computing, Composition Seminar, Introduction to Analysis, and Engineering Communication in a Professional Context.

4.2 Do Male and Female Engineering Students have different "Grade Anomalies" in their first-year courses?

To find if there are differences in grade anomalies for men and women, we grouped students by their self-reported gender and calculated the average grade anomaly for both groups for each course of interest. We then calculated Cohen's *d* as a measure of effect size between the two groups [49], which can be seen in Table 2. Women had indistinguishable or favorable AGA outcomes (e.g., smaller grade penalties or larger grade bonuses) compared to men, with three exceptions (see Fig. 2). For Intro to Computing, Physics 1, and Physics 2, men have smaller grade penalties or larger grade bonuses than women.

For both men and women, Professional Communication, a writing course, provided the largest grade bonus. For men, the courses that provided the largest grade penalties are Chemistry 1 and Chemistry 2, with grade penalties of –0.62 and –0.64, respectively. This means that men tended to receive over half a letter grade lower in this course than in their other courses. For women, the course that provided the largest grade penalty is Physics 2, with an AGA of –0.71.

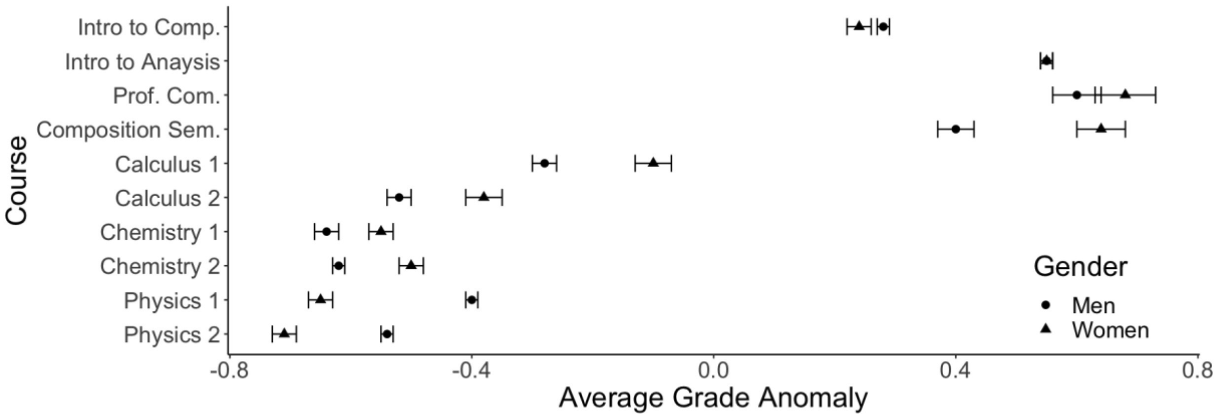

Fig. 2. Comparison of average grade anomaly (AGA) between men and women. The following words are abbreviated: Seminar (Sem.), Computing (Comp.), Communication (Com.), Professional (Prof.), and Standard deviation (SD). Ranges represent standard error of the mean.

Table 2. Comparison of grades and AGA between men and women. If the effect size given by Cohen's *d* is positive for grade, women had higher grades than men. If *d* is positive for AGA, women had a larger grade bonus or smaller grade penalty than men in that course. The following words are abbreviated: Seminar (Sem.), Computing (Comp.),Communication (Com.), Professional (Prof.), and Standard deviation (SD). c = $p < 0.05$, b = $p < 0.01$, and a = $p < 0.001$

	Women					Men						
		Grade		AGA			Grade		AGA		Cohen's *d*	
Course	**N**	**Mean**	**SD**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Mean**	**SD**	**Grade**	**AGA**
Composition Sem.	277	3.70	0.48	0.64	0.65	510	3.41	0.80	0.40	0.71	0.41[a]	0.34[a]
Intro to Comp.	1352	3.34	0.70	0.24	0.57	3320	3.26	0.75	0.28	0.63	0.10[a]	–0.05
Intro to Analysis	1580	3.60	0.49	0.55	0.47	3772	3.49	0.55	0.55	0.57	0.20[a]	0.00
Prof. Com.	129	3.86	0.35	0.68	0.52	209	3.77	0.34	0.60	0.57	0.26[a]	–0.10
Chemistry 1	1103	2.68	0.90	–0.55	0.81	3082	2.50	0.98	–0.64	0.86	0.19[b]	0.10[b]
Chemistry 2	899	2.67	0.84	–0.50	0.63	2486	2.46	0.93	–0.62	0.69	0.22[a]	0.17[a]
Physics 1	1783	2.67	0.73	–0.65	0.68	4239	2.76	0.81	–0.40	0.72	–0.11[a]	–0.36[a]
Physics 2	1342	2.54	0.82	–0.71	0.64	3420	2.59	0.94	–0.54	0.71	–0.06[c]	–0.25[a]
Calculus 1	1222	2.98	0.90	–0.10	0.99	3159	2.75	1.04	–0.28	1.11	0.23[a]	0.16[a]
Calculus 2	1288	2.79	1.06	–0.38	0.93	3313	2.63	1.13	–0.48	0.99	0.20[a]	0.15[a]

4.3 If there are Gender Differences in "Grade Anomalies", do they follow the same trends as Gendered Grade Differences?

There are some classes that show similar trends for grades and AGA, which can be seen in Table 2. We define similar trends as having a similar effect size (small, medium, or large) and a similar *p*-value. These include

Composition Seminar, Chemistry 1, Chemistry 2, Calculus 1, and Calculus 2. There are courses with a larger gender difference in grade than AGA. These include Intro to Computing, Intro to Analysis, and Engineering Communication in a Professional Context. There are courses with a larger gender difference in AGA than grade. These include Physics 1 and Physics 2.

5. Discussion

Our results show that there are grade penalties in all science and math courses studied, and bonuses in all engineering and English courses. We note that, similar to other studies that focus on AGA [38, 39], science and math courses have large grade penalties, while humanities courses have grade bonuses. Our results differ from past studies because first year courses offered by the engineering school have grade bonuses as opposed to penalties. In this section, we discuss: the potential harms of grade anomalies, what gender differences in grade anomalies can reveal about course equity, how grade anomaly related to academic self-concept; concerns about unequal access to coping mechanisms (methods that students use to persist in an environment with large grade penalties) regarding grade penalties, and why grade anomalies are a useful measure above and beyond raw grades.

First, we discuss why grade anomalies can be harmful. Lower than expected grades, even in a single course, can be a catalyst for students to leave STEM majors [7, 8]. 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 [7, 8]. This was a particular issue among high-achieving students, who were more likely to endorse perfectionism and feeling that their identity as "good STEM students" was threatened by B's and C's, or even a low grade on a single exam [8].

One way that students report coping with these unexpectedly low grades is by relying on others (such as friends, professors, or mentors) for support, which often comes as reassurance that low grades are normal in these difficult classes [8]. A second way students report coping is by accepting the harsher grading standards (such as curved grading or very low class exam averages) of STEM courses and decoupling their self-concept as STEM students with their grades [8]. Both of these coping mechanisms raise equity concerns, which are discussed in the next section.

It is important to note that these grading standards are a choice made by STEM departments and instructors. Requiring students to accept harsh grading standards and separate their identities as STEM students from their grades in order to successfully complete their degrees can distract students from their coursework. This, combined with evidence that shows that many high-achieving students leave these majors due to grade-related concerns, should lead instructors to question if their standards actually improve the education they offer students, or if they are simply pushing away all students except those who are capable of maintaining their academic self-concept divorced from grades.

In addition to seeing evidence of grade penalties in some courses for first-year engineering students, especially large-enrollment introductory courses, we also see some gender differences in grade anomalies. In particular, there were larger grade penalties for women in Physics 1 and Physics 2. Because women leave majors with higher grades than men who remain both in [6, 8] and outside [37] of STEM, this raises serious equity concerns. Past work suggests that women tend to have lower motivational beliefs that relate to academic self-concept, such as self-efficacy and sense of belonging [3, 20, 22, 30, 41, 55–57]. 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) [2, 8]. This trend has been found to be particularly strong among high-achieving women [8].

We suggest that women may be more likely to have a low academic-self-concept than men at similar performance levels for two reasons. First, prior research also suggests that women are less likely to separate their academic self-concept from their grades which is one of the clearest types of recognition in a domain [7, 8, 37]. In particular, grades are the resource that women have the most access to. Academic self-concept is formed through grades and feedback from outsiders. Because women are less likely to receive recognition as someone with potential in STEM from their parents [11, 58, 59], society at large [10, 60], and their instructors [12, 18, 61], they are more likely to rely on grade information to develop their academic self-concept. Next, women often tend to earn higher grades than men with the same standardized test scores [8, 62]. 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 (especially if there are stereotypes about who can excel in those domains), or they may compare their relatively-low STEM grades and leave for another subject that gives them the recognition for their work that they are accustomed to [7, 8].

Next, interview-based studies [8] suggest that women are less likely to have access to the coping mechanisms (i.e., support from peers or mentors and resources to decouple their self-identity as STEM students with their grades [8]) that students often use to continue even if they receive lower-than-expected grades in foundational courses. For example, women are less likely to receive advice that low grades are acceptable from peers and mentors because they are less likely to have peers and mentors they can relate to due to the underrepresentation of women in many STEM fields, such as engineering [5]. Further, if women are less likely to continue in their field of interest due to low grades [2, 8], the women who remain in or complete programs are more likely to be high-achieving in the field, and would thus be less likely to give advice that is useful to the average student (for example "I, like many others, received a C in this class but was still able to complete my program"). This second coping mechanism is essentially separating grades from academic self-concept. As stated earlier, because women are less likely to have positive feedback from outsiders (e.g., parents and instructors), they have no other way to form academic self-concept unless they can find a support system that can provide that outside recognition, though this sort of system may not be available to every student who seeks it out. These same arguments about lack of access to coping mechanism may apply to students from other underrepresented groups, who are also more likely to leave their programs due to lower-than-expected grades, i.e., due to grade penalty, than students from groups that are not underrepresented. Other groups that may be more affected by grade anomalies are racial and ethnic minority students [8, 63] and first-generation students [64].

Finally, 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 Physics 1 and Physics 2. This speaks to the usefulness of tracking both AGA and grades of the students. An instructor may see a small grade difference and understand that there is gender inequity in their classroom, but without knowing the gender differences in AGA, the instructor will not understand how those grades are perceived by female and male students. Understanding both grades and AGA differences may allow instructors to understand both classroom-level inequities and the extent to which their course may be pushing underrepresented groups, such as women, out of STEM fields.

There are also some courses that have larger gender differences in grades than in AGA, such as Intro to Computing, Intro to Analysis, and Engineering Communication in a Professional Context. However, we do not find these differences as concerning, because all of these courses are, on average, offering grade bonuses to all students, so they are less likely to decrease students' academic self-concept.

From our results, we make several recommendations to instructors and departments. First, measuring grade anomaly 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 discouraging students from underrepresented groups to leave a major.

Next, we encourage instructors and departments to evaluate their goals when developing grading practices. If a department aims to create a diverse and welcoming environment that attracts students, while also maximizing student learning, there are several productive practices to consider. Frequent, low-stakes assessment gives ample opportunity for instructor feedback and can minimize gender inequities in STEM classrooms [41, 42, 47]. This includes offering many types of assessment, such as quizzes, clicker questions, and projects in addition to or instead of homework problem sets and exams [41, 44, 65]. Collaborative and active learning approaches in equitable learning environments may also improve learning outcomes and grades while eliminating gender performance differences [41, 42, 66, 67]. Finally, we recommend instructors avoid curved grading and very low class averages: these practices do not reflect student performance, but they do discourage students and often contribute to students' reasons for leaving a field [7, 8].

The results presented in this study are very important because they provide evidence that courses in STEM departments (particularly large, mandatory, introductory courses) tend to result in grade penalties for students. This allows us to pinpoint departments and courses that may have grading practices that are inequitable or unrepresentative of student learning, as well as those that have more equitable and representative grading, so that practices may be shared across disciplines to improve learning environments in all disciplines. The relatively new measure of AGA may also act as a measure of academic self-concept that is easy for institutions to access. This can also be useful to researchers as they develop separate measurements for academic self-concept and expectancies for success. Because most research on measures of self-concept is relatively new [36], qualitative work in this area may help further clarify the connection between grade anomalies and academic self-concept, as well as reveal how they both affect retention. Further, grade anomaly may correlate with multiple factors, not just self-concept (for example, student self-efficacy, interest, course engagement, and impact of teaching methods), and qualitative or survey data may reveal more nuanced impact for each of these factors on grade penalty.

Although we have strong evidence of grade penalties in chemistry, mathematics and physics courses for first-year engineering students as well as gendered grade anomaly differences, we did not have access to syllabi or other information about individual courses for every course offered over the thirteen-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. Instead, we assume that, like the courses currently offered, most of these large, introductory courses are taught in a traditional, lecture-based, and exam-reliant format.

6. Conclusion

In this work we found that grade anomalies exist for all first-year engineering courses at our institution. Engineering and English Composition courses offered grade bonuses while Physics, Math, and Chemistry courses had grade penalties. Further, all courses had a grade anomaly (larger grade bonuses or smaller grade paneities) that favored women over men except for both Physics courses and Introduction to Computing. This raises particular concern about physics classes and the need for an equitable learning environment for engineering students. We also note that grade anomalies and raw grades do not reveal the same gender difference trends. Thus, both grade anomaly and raw grades should be tracked when determining if a course is equitable.

Finally, this research is based at a primarily white, large, public university, and while our results may generalize to similar institutions, we do not know what patterns of grade anomalies exist at liberal arts colleges, minority-serving institutions, or community colleges. Conducting research at a diverse range of institutions in different countries, as well as a stronger focus on how grade anomaly affects students from a variety of underrepresented groups, will help us more fully understand how grade anomalies differ for a range of students.

References

1. H. B. Gonzalez and J. J. Kuenzi, Science, Technology, Engineering, and Mathematics (STEM) Education: A Primer, Library of Congress Congressional Research Service, CRS Report No. R42530, 2012.
2. I. F. Goodman, Final Report of the Women's Experiences in College Engineering (WECE) Project, 2002.
3. K. M. Whitcomb, Z. Y. Kalender, T. J. Nokes-Malach, C. D. Schunn and C. Singh, Comparison of self-efficacy and performance of engineering undergraduate women and men, *International Journal of Engineering Education*, **36**(6), pp. 1996–2014, 2020.

4. K. M. Whitcomb, A. Maries and C. Singh, Examining gender differences in a mechanical engineering and materials science curriculum, *International Journal of Engineering Education*, **37**(5), pp. 1261–1273, 2021.
5. [National Center for Science and Engineering Statistics, https://ncses.nsf.gov/pubs/nsf19304/digest/field-of-degree-women#physical-sciences, Assessed 28 September 2021.
6. K. M. Whitcomb and C. Singh, Gender inequities throughout STEM: Women with higher grades drop STEM majors while men persist, *arXiv:2004.00560*, 2020.
7. E. Seymour, N. M. Hewitt and C. M. Friend, *Talking About Leaving: Why Undergraduates Leave the Sciences*, Westview Press, Boulder, CO, 1997.
8. E. Seymour, A.-B. Hunter, H. Thiry, T. J. Weston, R. P, Harper, D. G. Holland, A. K. Koch and B. M. Drake, *Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*, Cham: Springer International Publishing AG, 2019.
9. S. J. Leslie, A. Cimpian, M. Meyer and E. Freeland, Expectations of brilliance underlie gender distributions across academic disciplines, *Science*, **347**(6219), pp. 262–265, 2015.
10. L. Bian, S. J. Leslie and A. Cimpian, Gender stereotypes about intellectual ability emerge early and influence children's interests, *Science*, **355**(6323), pp. 389–391, 2017.
11. M. M. Bleeker and J. E. Jacobs, Achievement in math and science: Do mothers' beliefs matter 12 years later?, *Journal of Educational Psychology*, **96**(1), pp. 97–109, 2004.
12. A. A. Eaton, J. F. Saunders, R. K. Jacobson and K. West, How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates, *Sex Roles*, **82**(3), pp. 127–141, 2020.
13. C. M. Ganley, C. E. George, J. R. Cimpian and M. B. Makowski, Gender equity in college majors: Looking beyond the STEM/non-STEM dichotomy for answers regarding female participation, *American Educational Research Journal*, **55**(3), pp. 453–487, 2018.
14. S. J. Correll, Gender and the career choice process: The role of biased self-assessments, *American Journal of Sociology*, **106**(6)**,** pp. 1691–1730, 2001.
15. S. J. Correll, Constraints into preferences: Gender, status, and emerging career aspirations, *American Sociological Review*, **69**(1), pp. 93–113, 2004.
16. Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach and C. Singh, Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way, *Physical Review Physics Education Research*, **15**(2), p. 020148, 2019.
17. A. Maries, N. I. Karim and C. Singh, Is agreeing with a gender stereotype correlated with the performance of female students in introductory physics?, *Physical Review Physics Education Research*, **14**(2), p. 020119, 11/15/ 2018.
18. C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham and J. Handelsman, Science faculty's subtle gender biases favor male students, *Proceedings of the National Academy of Sciences of the United States of America*, **109**(41), p. 16474, 2012.
19. T. Bouffard-Bouchard, S. Parent and S. Larivee, Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students, *International Journal of Behavioral Development*, **14**(2), pp. 153–164, 1991.
20. V. Sawtelle, E. Brewe and L. H. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, *Journal of Research in Science Teaching*, **49**(9), pp. 1096–1121, 2012.
21. A. Cavallo, W. H. Potter and M. Rozman, Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors, *School Science and Mathematics*, **104**(6), pp. 288–300, 2004.
22. E. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn and C. Singh, Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?, *Physical Review Physics Education Research*, **14**(2), 020123, 2018.
23. J. M. Nissen and J. T. Shemwell, Gender, experience, and self-efficacy in introductory physics, *Physical Review Physics Education Research*, **12**(2), 020105, 2016.
24. P. R. Pintrich and E. V. De Groot, Motivational and self-regulated learning components of classroom academic performance, *Journal of Educational Psychology*, **82**(1), pp. 33–40, 1990.
25. B. J. Zimmerman, Self-efficacy: An essential motive to learn, *Contemporary Educational Psychology*, **25**(1), pp. 82–91, 2000.
26. J. A. Raelin, M. B. Bailey, J. Hamann, L. K. Pendleton, R. Reisberg and D. L. Whitman, The gendered effect of cooperative education, contextual support, and self-efficacy on undergraduate retention, *Journal of Engineering Education*, **103**(4), pp. 599–624, 2014.
27. S. Cwik and C. Singh, Damage caused by societal stereotypes: Women have lower physics self-efficacy controlling for grade even in courses in which they outnumber men, *Physical Review Physics Education Research*, **17**(2), p. 020138, 2021.
28. Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach and C. Singh, Gendered patterns in the construction of physics identity from motivational factors, *Physical Review Physics Education Research*, **15**(2), p. 020119, 2019.
29. M. Good, A. Maries and C. Singh, Impact of traditional or evidence-based active-engagement instruction on introductory female and male students' attitudes and approaches to physics problem solving, *Physical Review Physics Education Research*, **15**(2), p. 020129, 2019.
30. Y. Li and C. Singh, Effect of gender, self-efficacy, and interest on perception of the learning environment and outcomes in calculus-based introductory physics courses, *Physical Review Physics Education Research*, **17**(1), p. 010143, 2021.
31. Z. Y. Kalender, E. Marshman, C. D. Schunn, T. J. Nokes-Malach and C. Singh, Damage caused by women's lower self-efficacy on physics learning, *Physical Review Physics Education Research*,**16**(1), p. 010118, 2020.
32. S. Cwik and C. Singh, How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses, *Physical Review Physics Education Research*, **17**(2), p. 020143, 2021.
33. J. T. Spence, *Achievement and Achievement Motives: Psychological and Sociological Approaches*, San Francisco: W. H. Freeman, 1983.
34. B. Gniewosz, J. S. Eccles and P. Noack, Early adolescents' development of academic self-concept and intrinsic task value: The role of contextual feedback, *Journal of Research on Adolescence*, **25**(3), pp. 459–473, 2015.
35. A. Wigfield and J. S. Eccles, *Development of Achievement Motivation*, San Diego: Academic Press, 2002.
36. J. S. Eccles and A. Wigfield, From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation, *Contemporary Educational Psychology*, **61**, 101859, 2020.

37. K. Rask, Attrition in STEM fields at a liberal arts college: The importance of grades and pre-collegiate preferences, *Economics of Education Review*, **29**(6), pp. 892–900, 2010.
38. R. L. Matz, B.P. Koester, S. Fiorini, G. Grom, L. Shepard, C. Stangor, B. Weiner and T. A. McKay, Patterns of gendered performance differences in large introductory courses at five research universities, *AERA Open*, **3**(4), 2332858417743754, 2017.
39. B. P. Koester, G. Grom and T. A. McKay, Patterns of gendered performance difference in introductory STEM courses, *arXiv:1608.07565*, 2016.
40. M. Huberth, P. Chen, J. Tritz and T. A. McKay, Computer-Tailored Student Support in Introductory Physics, *PloS one*, **10**(9), e0137001, 2015..
41. C. J. Ballen, S. Salehi and S. Cotner, Exams disadvantage women in introductory biology, *PloS One*, **12**(10), pp. e0186419, 2017.
42. D. C. Haak, J. Hillerislambers, E. Pitre and S. Freeman, Increased structure and active learning reduce the achievement gap in introductory biology, *Science*, **332**(6034), pp. 1213–1216, 2011.
43. J. E. Bickel, L. M. Bunnell and T. Heus, Utilizing peer teaching and reflection on low-stakes quizzes to improve concept learning outcomes in introductory calculus-based physics classes, *European Journal of Physics*, **42**(5), 55701, 2021.
44. C. Singh and A. Malespina, Test anxiety, self-efficacy, and gender: A quest for equitable assessment practices in physics, *Physics Education Research Conference 2021*, Virtual, pp. 390–395, 2021. doi: 10.1119/perc.2021.pr.Singh.
45. C. M. Steele and J. Aronson, Stereotype threat and the intellectual test performance of African Americans, *Journal of Personality and Social Psychology,* **69**(5), pp. 797–811, 1995.
46. A. Wigfield and J. Cambria, Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes, *Developmental Review*, **30**(1), pp. 1–35, 2010.
47. A. Bandura, W. Freeman, and R. Lightsey, Self-efficacy: The exercise of control, *Journal of Cognitive Psychotherapy*, **13**(2), pp. 158–166, 1999.
48. Integrated Postsecondary Education Data System (IPEDS), https://nces.ed.gov/ipeds/datacenter/institutionprofile.aspx?unitId=215293&goToReportId=6, Assessed 25 September 2021.
49. B. B. Frey, Ed. *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, SAGE Publications, Inc., Thousand Oaks, CA, 2018.
50. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, L. Erlbaum Associates, Hillsdale, N.J., 1988.
51. R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020.
52. J. Lemon, Plotrix: a package in the red light district of R, *R-News*, **6**(4), p. 12, 2006.
53. D. Navarro, *Learning statistics with R: A tutorial for psychology students and other beginners*, University of New South Wales, Sydney, Australia, 2015.
54. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York, 2016.
55. A. L. Zeldin, S. L. Britner and F. Pajares, A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers, *Journal of Research in Science Teaching*, **45**(9), pp. 1036–1058, 2008.
56. R. A. Louis and J. M. Mistele, The differences in scores and self-efficacy by student gender in mathematics and science, *International Journal of Science and Mathematics Education*, **10**(5), pp. 1163–1190, 2012.
57. Z. Hazari, D. Chari, G. Potvin and E. Brewe, The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates, *Journal of Research in Science Teaching*, **57**(10), pp. 1583–1607, 2020.
58. K. Crowley, M. A. Callanan, H. R. Tenenbaum and E. Allen, Parents explain more often to boys than to girls during shared scientific thinking, *Psychological Science*, **12**(3), pp. 258–261, 2001.
59. J. E. Jacobs and J. S. Eccles, The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions, *Journal of Personality and Social Psychology*, **63**(6), pp. 932–944, 1992.
60. E. Reuben, P. Sapienza and L. Zingales, How stereotypes impair women's careers in science, *Proceedings of the National Academy of Sciences*, **111**(12), pp. 4403–4408, 2014.
61. J. Wang and Z. Hazari, Promoting high school students' physics identity through explicit and implicit recognition, *Physical Review Physics Education Research*, **14**(2), 020111, 2018.
62. D. Voyer and S. D. Voyer, Gender differences in scholastic achievement: A meta-analysis, *Psychological Bulletin*, **140**(4), pp. 1174–1204, 2014.
63. K. M. Whitcomb and C. Singh, Underrepresented minority students receive lower grades and have higher rates of attrition across STEM disciplines: A sign of inequity?, *International Journal of Science Education*, **43**(7), pp. 1054–1089, 2021.
64. M. E. Thompson, Grade expectations: The role of first-year grades in predicting the pursuit of STEM majors for first- and continuing-generation students, *The Journal of Higher Education*, **92**(6), pp. 961–985, 2021.
65. J. T. Laverty, W. Bauer, G. Kortemeyer and G. Westfall, Want to reduce guessing and cheating while making students happier? Give more exams!, *The Physics Teacher*, **50**(9), pp. 540–543, 2012.
66. S. M. Aguillon, G.-F. Siegmund, R. H. Petipas, A. G. Drake, S. Cotner and C. J. Ballen, Gender differences in student participation in an active-learning classroom, *CBE Life Sciences Education*, **19**(2), 2020.
67. N. I. Karim, A. Maries and C. Singh, Impact of evidence-based flipped or active-engagement non-flipped courses on student performance in introductory physics, *Canadian Journal of Physics*, **96**(4), pp. 411–419, 2018.

Alysa "Ly" Malespina is a PhD candidate in the Department of Physics and Astronomy at the University of Pittsburgh. She obtained her BA in Physics and Philosophy from Rollins College. Her research focuses on using statistical methods to study motivational characteristics and equity in physics education.

Chandralekha Singh is a distinguished professor in the Department of Physics and Astronomy and the Director of the Discipline-based Science Education Research Center at the University of Pittsburgh. She obtained her PhD in theoretical condensed matter physics from the University of California Santa Barbara and was a postdoctoral fellow at the University of Illinois Urbana Champaign, before joining the University of Pittsburgh. She has been conducting research in physics education for more than two decades. She held the Chair-line of the American Physical Society Forum on Education from 2009–2013 and was the chair of the editorial board of Physical Review Special Topics Physics Education Research from 2010–2013. She was the co-organizer of the first and third conferences on graduate education in physics and chaired the second conference on graduate education in physics in 2013. She is a Fellow of the American Physical Society, American Association of Physics Teachers and the American Association for the Advancement of Science.