

Test anxiety, self-efficacy, and gender: A quest for equitable assessment practices in physics

Chandralekha Singh and Alysa Malespina

Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA, USA 15260

Prior research shows that gender differences in self-efficacy and test anxiety can predict students' course performance. Although there has been research that uses test anxiety and self-efficacy to predict student grades, no study to our knowledge has investigated this in the context of low- and high-stakes physics assessments. In this work, we used survey data and grade information to compare the predictive power of self-reported self-efficacy and test anxiety on low-stakes (e.g., homework and quizzes) and high-stakes (e.g., traditional exams) assessment scores in a traditionally taught large introductory physics course. We found that there are gender differences in both self-efficacy and test anxiety, as well as in high-stakes assessment outcomes. There were no gender differences in low-stakes assessment scores. Further, we found that models that control for self-efficacy and/or test anxiety eliminate the predictive power of gender for high-stakes assessment scores. Finally, we found that self-efficacy partially mediates the effect of test anxiety on high-stakes assessment outcomes. These trends suggest that teaching methods that alleviate test anxiety have the potential to create a more equitable learning environment.

I. INTRODUCTION

Several prior studies have investigated gender differences in introductory physics course performance [1–3]. Some of the differences found can be attributed to factors that affect women before they enter college classrooms, like societal stereotypes about who belongs in physics [4]. These stereotypes also impact differential encouragement and opportunity for women and men to take math and physics at the pre-college level, which can lead to differences in prior preparation [5]. Gendered performance differences can also be attributed to differences in motivational beliefs about physics between male and female students (which can again at least partly be attributed to societal stereotypes about physics) [6–10, 12, 13]. Two factors that have been linked to gender differences in performance are test anxiety (TA) [12, 13] and self-efficacy (SE) [6–10].

Test anxiety is a phenomenon with three facets: cognitive, affective, and behavioral. The cognitive facet consists of worry and self-preoccupation (e.g., thinking about one's perceived shortcomings instead of the task at hand), as well as intrusive thoughts of failure, all of which limit the time and cognitive resources students with test anxiety can devote to the assessment [14]. The affective facet of TA affects how students feel when they have test anxiety, for example, a fast heartbeat or "butterflies in their stomach" [14]. The behavioral aspect of test anxiety manifests in avoidance techniques, such as procrastination or interacting only with surface-level feedback after the exam (e.g., not examining mistakes closely to make a plan for future improvement) [14, 15].

Test anxiety can affect students in different ways: one student may have strong study skills but be unable to concentrate due to physiological manifestations of test anxiety, another may feel extremely anxious because they lack study skills to succeed in exams, and a third student can have their own unique experience with test anxiety very different from the other two [14]. Ideally, educators will find ways to provide all students the environment they need to succeed in assessments without the burden of test anxiety, but this may require a multifaceted approach to meet different students' needs. Women are more likely to report test anxiety than men [14, 16], so understanding test anxiety is vital to create equitable learning environments.

Self-efficacy is one's belief in their capability to succeed at an activity or subject [16, 17], and has been linked to positive outcomes for physics students [6–10]. Self-efficacy is developed through four mechanisms: mastery experiences, such as overcoming obstacles; social modeling, or people similar to oneself succeeding; social persuasion, encouragement to increase resolve and measure success via personal improvement; and emotional states, such as management of anxiety [17]. Female students tend to have lower self-efficacy than male students in physics [6–11]. Prior research has attributed this to many related factors alluded to earlier, e.g., women are less likely to have the same level of previous mastery experiences as men (because they are less likely to take advanced physics in high school [5]), they have few role models due

to under-representation of women in physics [5, 11], and they are less likely to receive encouragement that they can succeed in physics from family [4], instructors [11, 18], and society at large [19]. Because self-efficacy allows students to develop coping mechanisms that could thwart test anxiety [16], we hypothesize that students with high self-efficacy will also have low test-anxiety.

To explore how test anxiety and self-efficacy predict students' performance in different situations, we compare female and male students' performance on low-stakes and high-stakes assessments. Here, low-stakes assessments are those that make up a small portion of a students' grade. High-stakes assessments are individual assessments that make up a large portion of students' grade (e.g., traditional exams [13]). In this paradigm, ten assessments that each make up five percent of a student's grade would be relatively low-stakes, while a single assessment that makes up fifty percent of a student's grade would be high-stakes, even if the content was identical between the assessments. We hypothesize that test anxiety and self-efficacy will more strongly predict high-stakes than low-stakes assessment outcomes.

In this study, we investigate the relationship between gender, low- and high-stakes assessment outcomes, test anxiety, and self-efficacy. Specifically, we aim to answer the following questions:

- RQ1. Are there gender differences in students' prior preparation, self-efficacy, or test anxiety?
- RQ2. Are there gender differences in students' low- and high-stakes assessment scores?
- RQ3. Are self-efficacy and test anxiety independent predictors of assessment scores?
- RQ4. Does gender predict scores in low- or high-stakes assessments when controlling for self-efficacy or test anxiety?

II. METHODOLOGY

This study took place at a large research University in the United States. We administered a motivational survey to students in an introductory calculus-based physics I course during the first and last week of the semester. We call the first and final data sets "pre" and "post", respectively. For analysis, we longitudinally matched students ($N = 176$), including only students who completed both surveys. The traditional lecture-based course was taught by four instructors and primarily covered Newtonian mechanics. Most students were engineering and physical science majors. Students were either given extra credit or a participation grade for taking the survey, depending on the instructor.

Demographic data and prior academic information - for example, Scholastic Achievement Test (SAT) scores and high school grade point averages (HS GPAs) - were acquired from de-identified university records via an honest broker. American College Testing (ACT) scores were converted into approximate SAT scores [20]. Demographic data indicated

our sample was 37% women. We recognize that gender is not binary, but are unable to reflect this in our work with the available data. Students identified with the following races/ethnicities: 72% White, 14% Asian, 7% Hispanic/Latinx, 4% multiracial, 2% African American/Black, and 1% unspecified. The majority (80%) of students in the sample were first-semester first year students, and 60% of them were engineering majors. Outside of engineering, most students were undeclared or studying a physical science.

Academic performance measures were provided by instructors. These measures included assessment scores for homework, quizzes, midterm exams, and final exams. For analysis, we divided these into two groups: "low-" and "high-stakes" assessments. The low-stakes assessment measure combined quiz and homework grades, while high-stakes assessment measure combined midterm and final exam grades. Homework grades were only available for 74% of students, but other assessment scores were available for the whole sample. Before averaging, all assignments were normalized to a 10-point scale.

The TA survey questions were adapted from the previously validated [21, 22] Motivated Strategies for Learning Questionnaire [23]. To ensure we were measuring domain-specific mindset, we explicitly mentioned physics in the survey items, as seen in Table I. For example "I have an uneasy, upset feeling when I take an exam," becomes "I have an uneasy, upset feeling when I take a physics test". SE survey questions were constructed from other surveys and previously validated [6, 7]. We further validated the survey through twenty one-hour student interviews to ensure that students interpreted questions as intended. Additionally, we performed confirmatory factor analysis (CFA) using the students in this study as a check for continued validity. For both pre and post-surveys, the Comparative Fit Index (CFI) and Tucker Lewis Index (TLI) were ≥ 0.95 [24], and the Root Mean Square Error of Approximation (RMSEA) was ≤ 0.08 [25], and the Standardized Root Mean Square Residual (SRMR) was ≤ 0.06 [24]. Standardized factor loadings ranged from 0.62-0.88 [24]. Chronbach's α was 0.75 and 0.81 for self-efficacy pre and post; while α was 0.88 and 0.90 for test anxiety pre and post. We found Pearson correlations between all variables. The weakest correlations were non-significant, while the strongest correlation was $r = 0.60$ ($p < 0.001$), between post test anxiety and post self-efficacy.

Test anxiety items were on a five-point Likert scale (1-Not at all true, 2-A little true, 3-Somewhat true, 4-Mostly true, 5-Completely true) and self-efficacy questions were on a four-point Likert scale (1-NO!, 2-no, 3-yes, 4-YES!). A higher score indicates the student has more test anxiety or higher self-efficacy.

Before conducting any analysis, SAT scores, HS GPAs, assessment scores, test anxiety and self-efficacy scores were winsorized to two standard deviations from the mean (in order to maintain the direction of outliers while eliminating extreme values [26]). To determine if there were gender differences in pre and post- TA, as well as assessment scores and prior aca-

TABLE I: The physics test anxiety (TA) and self-efficacy (SE) items on the survey.

Item No.	Item Text
TA 1	I am so nervous during a physics test that I cannot remember what I have learned.
TA 2	I have an uneasy, upset feeling when I take a physics test.
TA 3	I worry a great deal about physics tests.
TA 4	When I take a physics test, I think about how poorly I am doing.
SE 1	I am able to help my classmates with physics in the laboratory or in recitation.
SE 2	I understand concepts I have studied in physics.
SE 3	If I study, I will do well on a physics test.
SE 4	If I encounter a setback in a physics exam, I can overcome it.

ademic preparation, we performed unpaired t -tests and calculated the Cohen's d between groups. Cohen's d is a measure of effect size, and we used the following standards: small, $d \sim 0.2$; medium, $d \sim 0.5$; and large, $d \sim 0.8$ [27].

Next, we conducted a mediation analysis of test anxiety and high-stakes assessment grades, using self-efficacy as a mediator. This model was chosen because self-efficacy had the largest Pearson correlation with test anxiety compared to all other variables. All involved variables were z-scored so the regression weights were standardized. Mediation was conducted in R using the bootstrap method with 1000 simulations [28].

Finally, to explore the predictive relationships between test anxiety and assessment outcomes, we used multiple regression analysis. For each regression model, we report the standardized β coefficients, sample size, and R-squared. Standardized coefficients were used because they are in units of standard deviation and allow for direct comparison of effects [27]. We initially used gender, SAT math scores, and HS GPA as predictors for low- and high-stakes assessment scores. After establishing baseline models, we introduced pre-TA or the average TA as predictor variables. Average test anxiety is the mean of pre and post test anxiety, and was used as a proxy for students' test anxiety while they were taking the course. During regression analysis, we used combined assessment categories (e.g., low- and high-stakes assessments), but results were similar when the categories were separated. For example, the regression models predicting low-stakes assessment scores were similar to both the models predicting quiz grades and those predicting homework grades.

III. RESULTS AND DISCUSSION

Are there gender differences in students' prior preparation, self-efficacy, or test anxiety? Both measures of prior preparation had gender differences (see Table II). SAT math scores were higher for men, while high school GPAs were

higher for women. This suggests that in this group of students, women generally had more developed general academic skills and self-regulatory behavior, while men, on average, had higher levels of math problem-solving skills at the time of admission [29].

Table II also shows that women reported higher levels of test anxiety than men, but lower levels of self-efficacy. For self-efficacy, the gender differences grew from small to medium over the semester; the test anxiety gender gap also increased, but maintained a large effect size. This is consistent with other studies that find gender differences in self-efficacy [7, 10] and test anxiety [12] in the physics context.

TABLE II: Means (M) and standard deviations (SD) of academic and motivational factors by gender. Cohen’s d is negative when men have a higher score than women. * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$.

Variable	Women		Men		d
	M	SD	M	SD	
SAT Math (200-800)	691	55	718	51	-0.51**
HS GPA (0-5)	4.22	0.33	4.18	0.33	0.13
TA Pre (1-4)	2.83	1.02	2.11	0.79	0.82***
TA Post (1-4)	3.15	0.99	2.29	0.89	0.93***
SE Pre (1-4)	3.03	0.44	3.24	0.43	-0.49**
SE Post (1-4)	2.82	0.48	3.18	0.47	-0.75***
Homework (0-10)	9.16	0.52	9.04	0.59	0.19
Quiz (0-10)	9.36	0.58	9.48	0.46	-0.22
Midterm Avg. (0-10)	7.30	1.13	8.01	1.18	-0.61***
Final Exam (0-10)	5.50	1.73	6.41	1.75	-0.53***

Are there gender differences in students’ low- and high-stakes assessment scores? From Table II, homework and quiz grades show no significant gender difference, while midterm and final exams have significant medium differences. Although grade schemes differed by instructor, all courses had grades based primarily on midterm and final exam scores. Because research suggests women are more likely than men to leave STEM fields due to concerns about grades (even if they have an A or B average) [30], women’s significantly lower exam scores may contribute to the loss of women from majors, like engineering, that require introductory calculus-based physics.

Are self-efficacy and test anxiety independent predictors of assessment scores? The relationship between average self-efficacy and test anxiety is explored in a mediation model (see Figure 1). This model tests if average self-efficacy mediates the relationship between test anxiety and high stakes assessment scores, and was statistically significant. The Average Mediation Effect (AME) was -0.19 ($p < 0.001$), with a confidence interval of [-0.29,-0.09]. The average direct effect (ADE) was -0.24, $p = 0.006$ and the total direct effect (TDE) was -0.43, $p < 0.001$. Mediation models were similar for men and women when tested separately, so we combined them to maximize our sample size.

The model in Figure 1 suggests that test anxiety is par-

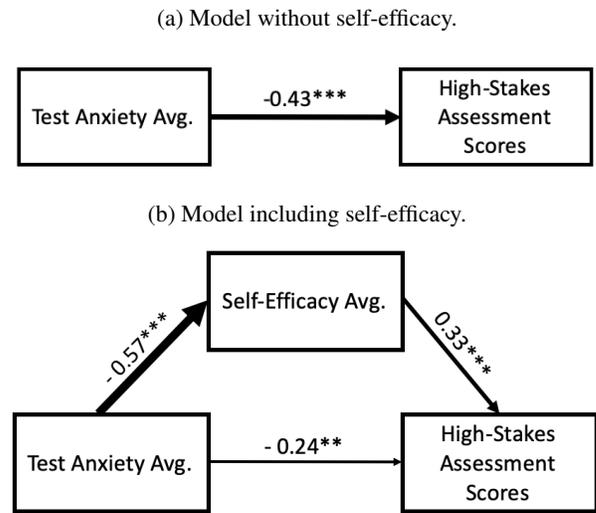


FIG. 1: Mediation model results: average self-efficacy mediates the relationship between average test anxiety and high-stakes assessment scores. Line thickness corresponds to the standardized regression coefficient size. $N = 181$.

tially mediated by self-efficacy. Direct effects of test-anxiety on high-stakes assessment grades may be because test anxiety affects test performance directly by taking up students’ cognitive resources during assessments [31]. Additionally, Bandura [16] theorized that poor performance and anxiety are "co-effects" of low self-efficacy, which may explain the mediation effect. Heightened self-efficacy and reduced test anxiety likely form a virtuous cycle wherein students’ development of coping mechanisms (for example, stress-reduction techniques and explicitly rehearsing strategies for academic challenges) increase their self-efficacy, and increased self-efficacy frees students’ cognitive resources to focus on the task at hand and better implement coping strategies [16, 32].

Importantly, the relationship between test anxiety, self-efficacy and high-stakes assessments is similar for women and men. Prior work has shown similar results: anxiety affects academic self-efficacy similarly for men and women [33], but women tend to have lower self-efficacy than men regarding physics [6–10].

Does gender predict scores in low- or high-stakes assessments when controlling for self-efficacy or test anxiety? Tables III and IV show the results of our eight regression models predicting low- and high-stakes assessments, respectively. Each of the models use the variables in the far left column to predict the outcome variable (e.g., low- or high-stakes assessment scores). The strength of each predictor, controlling for all other predictors in the model, is given by the standardized β coefficient [27].

In Models 1-4 (see Table III, gender is not a significant predictor of low-stakes assessment scores, so there is no difference in motivational factors to account for. Only HS GPA

and and pre-SE were significant predictors of low-stakes assessment outcomes. However, pre-SE was only a predictor when pre-TA was also included, as in Model 4 (as opposed to Models 2-3, which include TA and SE separately). However, SE is unlikely to have real-world effects on students' low-stakes assessment outcomes; all of the models in Table III explain only small amounts of the variance. Average TA and SE were not significant predictors (alone or together).

TABLE III: Standardized β coefficients of regression models predicting low-stakes assessment scores. $N = 131$, "F" and "M" refer to female and male students. Significant predictors are bold. * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$.

Predictor	Regression Coefficient (β)			
	Model 1	Model 2	Model 3	Model 4
Gender (F=1, M=0)	0.01	-0.03	0.04	0.00
SAT Math	0.19*	0.19*	0.19*	0.20*
HS GPA	0.11	0.11	0.12	0.12
Test Anxiety Pre		0.07		0.16
Self-Efficacy Pre			0.16	0.22*
R^2	0.03	0.03	0.05	0.06

Regression models predicting high-stakes assessments are found in Table IV. Model 5 shows that women have lower scores than men on high-stakes assessments when controlling for SAT math scores and HS GPA. Therefore, more factors are needed to account for this discrepancy. Pre SE and TA were not significant predictors, so models that included them did not account for the gender difference. However, including average TA (as in Model 6) or average SE (as in Model 7) renders the gender effect non-significant, suggesting that both TA and SE can explain some of the gendered performance differences in this class. If both average TA and SE are included, as in Model 8, TA becomes non-significant.

Prior work has found similar results: both TA and SE are significant, but TA does not explain more variance than SE alone, and TA becomes non-significant when SE is included [12]. These similarities suggest that differences in both test anxiety and self-efficacy are widespread in physics.

TABLE IV: Standardized β coefficients of regression models predicting high-stakes assessment scores. $N = 176$, "F" and "M" refer to female and male students. Significant predictors are bold. * = $p < 0.05$, ** = $p < 0.01$, and *** = $p < 0.001$.

Predictor	Regression Coefficient (β)			
	Model 5	Model 6	Model 7	Model 8
Gender (F=1, M=0)	-0.20**	-0.09	-0.08	-0.05
SAT Math	0.46***	0.42***	0.43***	0.41***
HS GPA	0.23***	0.23***	0.25***	0.24***
Test Anxiety Avg		-0.26**		-0.10
Self-Efficacy Avg			0.37**	0.32**
R^2	0.38	0.43	0.49	0.49

IV. CONCLUSION AND FUTURE DIRECTIONS

Our results show that there are gender differences in students' self-efficacy and test anxiety. When controlling only for prior preparation, there is also a gender gap in high-stakes assessment scores, which becomes non-significant when controlling for self efficacy and/or test anxiety. Lastly, self-efficacy mediates the relationship between test anxiety and high stakes assessment scores. From these findings we conclude that physics classrooms have the potential to become more equitable for women if instructors focus on giving students the support they need to enter a cycle of increasing self-efficacy and decreasing test anxiety.

Educators should maintain high standards while providing students the tools and scaffolding support they need to become independent learners, utilizing frequent feedback [16]. The student sample in this study was 80% first-semester college students, who may be new to exams that make up the majority of their course grade. Implementing frequent, low-stakes assessment (for example, weekly or biweekly exams) can give students many attempts to practice test-taking and study skills and gives ample opportunity for feedback [34, 35]. It also can space the practice (instead of cramming right before the exams, students will study more uniformly) which can lead to better retention of content and better skill development [36]. This can create a more equitable classroom environment by minimizing fears that come with test anxiety (e.g., receiving a low course grade due to one bad exam score) but also helps students develop the skills they will need if and when they encounter high-stakes exams in the future. In addition, instructors should decrease the importance of traditional exams in final course grades, while increasing the importance of clicker questions, homework, projects, and other assessments. A broad range of assessments should be implemented throughout the course, both to give students frequent feedback and opportunities to master content [35, 37].

This study mirrors results seen in other disciplines and institutions [12, 13]. As we collect more data, we want to include intersectional analysis to understand the relationship between gender, race, test anxiety, and self-efficacy. Additionally, we are interested in how test anxiety affects students in other contexts, e.g., at minority-serving institutions, community colleges, and institutions with small class sizes or in courses in which there is explicit attempt to make the learning environment equitable and inclusive.

ACKNOWLEDGMENTS

We thank the National Science Foundation for awards DUE-1524575 and PHY-1806691.

- [1] S. Salehi, E. Burkholder, P. Lepage, S. Pollock, and C. Wieman, Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics, *Phys. Rev. Phys. Educ. Res.* **15**, 020114 (2019).
- [2] P. Sadler and R. Tai, Success in introductory college physics: The role of high school preparation, *Sci. Educ.* **85**, 111 (2001).
- [3] A. Madsen, S. McKagan, and E. Sayre, Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gaps?, *Phys. Rev. Phys. Educ. Res.* **9**, 020121 (2013).
- [4] M. Bleeker and J. Jacobs, Achievement in math and science: Do mothers' beliefs matter 12 years later?, *J. Educ. Psychol.* **96**, 1 (2004).
- [5] A. Porter and R. Ivie, Statistical Research Center of the American Institute of Physics, Women in Physics and Astronomy, 2019.
- [6] E. Marshman, Z. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, Female students with A's have similar physics self-efficacy as male students with C's in introductory courses: A cause for alarm?, *Phys. Rev. Phys. Educ. Res.* **14**, 020123 (2018).
- [7] E. Marshman, Z. Kalender, C. Schunn, T. Nokes-Malach, and C. Singh, A longitudinal analysis of students' motivational characteristics in introductory physics courses: Gender differences, *Can. J. Phys.* **96**, 391 (2017).
- [8] V. Sawtelle, E. Brewere, and L. Kramer, Exploring the relationship between self-efficacy and retention in introductory physics, *J. Res. Sci. Teach.* **49**, 1096 (2012).
- [9] A. Cavallo, W. 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, *Sch. Sci. Math* **104**, 288 (2004).
- [10] J. Nissen and J. Shemwell, Gender, experience, and self-efficacy in introductory physics, *Phys. Rev. Phys. Educ. Res.* **12**, 020105 (2016).
- [11] D. Doucette and C. Singh, Why Are There So Few Women in Physics? Reflections on the Experiences of Two Women, *Phys. Teach.* **58**, 297-300 (2020).
- [12] J. Stang, E. Altieri, J. Ives, and P. Dubois, Exploring the contributions of self-efficacy and test anxiety to gender differences in assessments, in *Physics Education Research Conference 2020*, (Virtual Conference, 2020).
- [13] C. Ballen, S. Salehi, and S. Cotner, Exams disadvantage women in introductory biology, *PLoS ONE* **12**, 10 (2017).
- [14] M. Zeidner, *Test Anxiety: The State of the Art* (Springer, 1998).
- [15] L. Solomon and E. Rothblum, Academic procrastination: Frequency and cognitive-behavioral correlates, *J. Couns. Psychol.* **31**, 4 (1984).
- [16] A. Bandura, *Self-efficacy: The Exercise of Control* (Macmillan, 1997).
- [17] A. Bandura, On the functional properties of perceived self-efficacy revisited, *J. Manage.* **38**, 1 (2012).
- [18] Y. Li, K. Whitcomb, and C. Singh, How perception of being recognized or not recognized by instructors as a "physics person" impacts male and female students' self-efficacy and performance, *Phys. Teach.* **58**, 484-487 (2020).
- [19] E. Reuben, P. Sapienza, and L. Zingales, How stereotypes impair women's careers in science, in *Proceedings of the National Academy of Sciences* **111**, 12 (2014).
- [20] The College Board and ACT, Inc, Guide to the 2018 ACT/SAT Concordance, 2018.
- [21] M. Credé and L. Phillips, A meta-analytic review of the motivated strategies for learning questionnaire, *Learn. Individ. Differ.* **21**, 4 (2011).
- [22] P. Pintrich, D. Smith, T. Garcia, and W. Mckeachie, Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ), *Educ. Psychol. Meas.* **53**, 3 (1993).
- [23] P. Pintrich, *A Manual for the Use of the Motivated Strategies for Learning Questionnaire* (University of Michigan, 1991).
- [24] L. Hu and P. Bentler, Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification, *Psychol. Methods* **3**, 4 (1998).
- [25] M. Browne and R. Cudeck, Alternative ways of assessing model fit, *Sociol. Methods Res.* **21**, 2 (1992).
- [26] B. Frey, editor, *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, (SAGE Publications, 2018).
- [27] J. Cohen, P. Cohen, S. West, and L. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, (Lawrence Erlbaum, 2003).
- [28] D. Tingley, T. Yamamoto, K. Hirose, L. Keele, and K. Imai, Mediation: R Package for casual mediation analysis. R package version 4.4.2 (2013). Retrieved from <http://CRAN.R-project.org/package=mediation>.
- [29] B. Galla, E. Shulman, B. Plummer, M. Gardner, S. Hutt, J. Goyer, S. D'Mello, A. Finn, and A. Duckworth, Why high school grades are better predictors of on-time college graduation than are admissions test scores: The roles of self-regulation and cognitive ability, *Am. Educ. Res. J.* **56**, 6 (2019).
- [30] I. Goodman, *Final report of the women's experiences in college engineering project*, (Goodman Research Group, 2002).
- [31] S. Beilock, R. Rydell and A. McConnell, Stereotype threat and working memory: Mechanisms, alleviations, and spillover, *J. Exp. Psychol. Gen.* **136**, 256 (2007).
- [32] R. Smith, Effects of coping skills training on generalized self-efficacy and locus of control, *J. Pers. and Social Psychology* **56**, 2 (1989).
- [33] J. Meece, A. Wigfield, and J. Eccles, Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics, *J. Educ. Psychol.* **82**, 1 (1990).
- [34] J. Laverty, W. Bauer, G. Kortemeyer, and G. Westfall, Want to reduce guessing and cheating while making students happier? Give more exams!, *Phys. Teach.* **50**, 540 (2012).
- [35] D. Haak, J. Hillerislanders, E. Pitre, and S. Freeman, Increased structure and active learning reduce the achievement gap in introductory biology, *Science* **332**, 6034 (2011).
- [36] J. Anderson, *Learning and Memory: An Integrated Approach 2nd ed.* (Wiley, 2000).
- [37] S. Freeman, S. Eddy, M. McDonough, M. Smith, N. Okoroafor, H. Jordt, and M. Wenderoth, Active learning boosts performance in STEM courses, in *Proceedings of the National Academy of Sciences* **111**, 23 (2014).