Bridging the Gender Gap in Exam Performance through Nudges and Stress Reframing*

Catalina Franco† Marcela Gomez-Ruiz‡

November 20, 2024

Abstract

Why women underperform relative to men in high-stakes exams while excelling under lower stakes remains puzzling. Previous research suggests differential responses to pressure and omitted questions as explanations. We evaluate a unique randomized trial with 4,658 applicants to a coding program in Uruguay, introducing two treatments: a *Nudge* to answer all questions and a *Nudge + Stress* intervention, reframing stress as performance-enhancing. Treated women omit fewer questions, boosting performance by 0.08 SD and 0.18 SD, respectively. There are no effects on men, and 9% more women gain admission. The key insight is that reducing omitted questions narrows gender performance gaps.

JEL CODES: C93, D91, I20, J16, J24

KEYWORDS: Gender, education, entrance exams, omitted questions, stress reappraisal, meditation

*We thank Ceibal in Uruguay (https://ceibal.edu.uy/en/) for generously sharing their administrative data with us, and in particular Guillermina Suárez, Dinorah de León, Gabriel Inchausti, Irina Sánchez, Irene González, Yanedy Pérez and Cecilia Hughes who were part of the Ceibal team designing and implementing the intervention. We obtained valuable feedback from Martin Brun, Alexander Cappelen, Maria Cervini-Plá, Maria Cubel, Yan Chen, Ferrán Elias, Julien Grenet, Nagore Iriberri, Siri Isaksson, Brian Jacob, Marianne Page, Pedro Rey-Biel, Tanya Rosenblat, Perihan Saygin, Hannah Schildberg-Hörisch, Danila Serra, Mikko Silliman, and participants at the Choice Lab coffee meeting, the University of Michigan School of Information BEE lab meeting, the 2023 Nordic Behavioral Economics Conference, the IHE Colloquium at Leibniz University Hannover, the Symposium of the Spanish Economic Association, the Economic Science Association Conference, the LEAD Seminar at UB, the Lunch Seminar at UAB, and the FAIR CELE coffee meeting. Gomez-Ruiz acknowledge financial support from the Spanish Ministry of Science and Innovation (PID2022-137352NB-C41).

†Center for Applied Research (SNF) and FAIR at NHH Norwegian School of Economics, Helleveien 30, 5042 Bergen, Norway. Email: Catalina.Franco@snf.no, corresponding author.

‡Ceibal and Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. Email: Marcela.Gomez@uab.cat.

1 Introduction

Although women now surpass men in educational attainment across most countries, they still face barriers in accessing the most selective academic programs and institutions. A key reason for this disparity is that women often underperform relative to men on high-stakes, competitive exams (e.g., Ors et al., 2013; Azmat et al., 2016; Iriberri and Rey-Biel, 2019; Cai et al., 2019; Arenas and Calsamiglia, 2022), which serve as entry points to these high-quality academic pathways.

One suggested explanation is that exam pressure may have a stronger negative impact on women's performance. For example, Cai et al. (2019) found a significant drop in performance for women—but not for men—from a mock exam to the high-stakes *Gaokao* college entrance exam in China. Another factor potentially underlying gender differences in exam performance is the tendency for women to leave more exam questions unanswered than men (Espinosa and Gardeazabal, 2010; Baldiga, 2014; Pekkarinen, 2015; Riener and Wagner, 2017; Coffman and Klinowski, 2020; Atwater and Saygin, 2020; Iriberri and Rey-Biel, 2021; Karle et al., 2022). Despite the crucial role exam performance plays in shaping future academic and economic opportunities, little is known about which strategies can effectively help level the playing field in accessing high-quality education.

In this paper, we provide the first evidence on how low-cost, brief interventions can improve women's exam performance and chances of admission. We analyze administrative data from an in-exam randomized experiment conducted by a public-private agency in Uruguay for admission to its Coding Program. In the 2023 admissions, the agency had seven exam versions testing 4,658 exam takers in verbal, mathematics, concentration and logic subjects. The exam versions were randomly distributed across exam takers. The versions were equivalent in difficulty level and were assigned to three groups. Two versions were as they typically are (*Control*). Three versions contained a nudge prompting exam-takers to answer all exam questions and reminding them that wrong answers are not penalized (*Nudge*). One exam version contained the nudge plus a stress management intervention focused on highlighting how stress can have positive effects on performance (*Nudge + Stress*).[1] In this version, students read a paragraph about how to interpret stress, answered a question about the paragraph (stress reappraisal exercise), and then were prompted to perform a short meditation midway through the exam.

We find that both the *Nudge* and *Nudge + Stress* treatments substantially increase the performance of women via fewer omitted questions and a higher likelihood of obtaining correct answers in

[1] The Behavioral Science Lab at Ceibal designed the interventions with the aim of raising the share of admitted female applicants into the program. The stress prompts are based on the paper by Harris et al. (2019).

attempted questions. The *Nudge + Stress* intervention is particularly effective in reducing the gender gap in admissions to the Coding Program by helping women score above the admission threshold due to the larger performance effects of this combined intervention. Our main contribution is to demonstrate how simple nudges and reframing the perception of stress can serve as powerful, low-cost strategies to expand women's access to high-quality education.

Our three main findings are as follows. First, compared to students in the *Control* group, who only received standard exam instructions and questions, women's overall exam scores increased by 0.08 standard deviations (SD) in the *Nudge* group and by 0.18 SD in the *Nudge + Stress* group. These score increases for women are driven by improved performance in both the verbal (0.11 SD) and mathematics (0.095 SD) sections in *Nudge*, and in verbal (0.23 SD), mathematics (0.175 SD), and concentration (0.21 SD) in *Nudge + Stress*. The effect sizes in the combined *Nudge + Stress* treatment are nearly double those in the *Nudge* treatment alone, suggesting that the stress management component—which guides participants to reinterpret stress as a positive, performance-ready state—provides additional benefits. Moreover, in *Nudge + Stress*, a brief 30-second meditation break after the math section appears to improve performance in the subsequent concentration section.

Second, we find that women are less likely to omit questions as a result of the treatments. Again, the effects are stronger in *Nudge + Stress* than in *Nudge* given that omitted questions go down for every exam subject in the former and only for verbal and logic in the latter. Importantly, accuracy (i.e., correct/attempted) increases slightly or remains the same as in *Control* indicating that women attempting more questions does not reduce the likelihood that the answers are correct. This evidence suggests that women are indeed leaving valuable points on the table by not attempting all questions, and that simple reminders to answer every question—along with clarifications that incorrect answers carry no penalties—can significantly boost performance. Overall, *Nudge + Stress* increases the fraction of the exam completed by 5.5 percentage points (pp) compared to a baseline of 83% in the *Control* group. Consequently, the gender gap in exam completion, originally at 6.7 pp, is significantly reduced.

Finally, 9% more women are admitted to the Coding Program as a result of *Nudge + Stress*. This represents a reduction of over two thirds of the gender gap in admissions in the *Control* group estimated at 8.5 pp. Importantly, the women who are admitted as a result of the intervention are of no lower quality than admitted women in the *Control* group as their continuation rates based on teachers' assessments after phase 1 of the program are the same. There is no effect of *Nudge* on the gender gap in admissions and there are virtually no negative effects of any of the interventions on men.

The effectiveness of our interventions appears to be driven by increased effort, reflected in both a higher rate of attempted questions and greater effectiveness at solving attempted questions. The interventions have a larger impact on women's behavior than men's, likely because men were already answering most exam questions, leaving less room for improvement.[2] In contrast, women are more inclined to leave questions unanswered, potentially due to factors like higher risk aversion and lower self-confidence, as suggested by prior literature (Atwater and Saygin, 2020; Iriberri and Rey-Biel, 2021; Balart et al., 2022; Karle et al., 2022). The *Nudge* intervention may work by bringing the "attempt all questions" mindset to the forefront, prompting more engagement with every single question. Meanwhile, the *Nudge + Stress* intervention could bolster confidence, particularly for students who might otherwise second-guess their answers.[3] The theory in psychology (Beilock, 2011) posits that stress influences performance through working memory, as cognitive resources must be allocated to manage worry. Worry is expressed by way of intrusive thoughts such as the consequences of failing the test, which divert attention from problem-solving. By reducing the cognitive load associated with stress and worry, our interventions likely free up cognitive capacity for solving exam questions, leading to the performance improvements observed.

Since we cannot directly validate the proposed theories, we consider alternative mechanisms and assess their plausibility. First, a potential confounder is baseline covariate differences between men and women. However, our regressions control for a comprehensive set of covariates, accounting for these differences in the main results. Second, it is possible that women engage more seriously with the intervention, but both genders write a similar amount of words in the stress reappraisal exercise and overall spend comparable time on the exam, suggesting equal engagement. Third, women's improved performance may stem from being in a "friendly environment" rather than from better stress management, and the meditation prompt may have functioned more as a break than a stress relief tool. Although we cannot definitively rule out these alternatives, we believe the positive effects of the *Nudge + Stress* intervention on performance are valuable regardless of the specific mechanism. One way to interpret our findings is that less threatening testing environments—whether by reducing stress or offering brief accommodations—can significantly enhance women's performance.

[2] Our null result for men aligns with the findings of De Paola and Gioia (2016) on performance under time pressure and Cavatorta et al. (2021) on exam anxiety.

[3] The OECD (2015) report points out that girls are more fearful of making mistakes, and report greater anxiety towards mathematics and lower levels of self-efficacy and self-concept. High-achieving girls, in particular, suffer more from high levels of anxiety than high-achieving boys, even when they have greater intrinsic motivation to learn mathematics.

We contribute to the literature on the gender gap in academic performance by highlighting nudges and stress management as low-cost tools to boost effort and exam performance. Specifically, we add to the emerging research on the effects of incorporating mindfulness techniques into educational settings and daily life (Cassar et al., 2022; Shreekumar and Vautrey, 2022). We also contribute by linking two seemingly disparate strands of literature: gender differences in exam performance and gender differences in omitting exam questions. Although these factors appear intuitively related, to the best of our knowledge, no one has formally established this link. Taken together, our results suggest that nudges and short stress management exercises can go a long way in increasing the share of women in high-quality education with potential downstream effects on gender earnings gaps and diversity within these fields.

2 Setting and Research Design

2.1 Education in Uruguay and the Potential Demand for the Coding Program

Uruguay pioneered the implementation of the *One Laptop Per Child* initiative worldwide through Ceibal, a public-private agency that ensures connectivity and access to educational content for all students in the public education system. Despite this progress, Uruguay faces persistent educational challenges, including having one of the lowest secondary and tertiary graduation rates in the region (USAID, 2022).

In the face of high dropout rates and low graduation rates, short-term programs providing useful skills for the labor market can be attractive to students who are looking for opportunities to improve their skill set. The Coding Program satisfies this need by providing free training in a high-demand field for people between 18 and 30 years of age who have completed at least third grade of high school.[4] Given the relevance of the program and the strong connections with private sector companies hiring program graduates, it is also attractive to youths who have not necessarily dropped out from the formal education system. The demand for the program has increased over the years[5] and in 2023 over 9,000 people expressed interest in the Coding Program.

[4]High school in Uruguay comprises 6 years. Students are expected to finish high school at age 17. Usually, third grade of high school corresponds to age 15.

[5]Since 2017, the program has trained more than 4,100 youths.

2.2 The Coding Program Design and Entrance Exam

The program aims to train participants in a coding language, English, and soft skills related to preparation for the labor market (CV and interview preparation). The program is organized in three phases taking place over the course of one year. In phase 1, from March to June, students take basic courses asynchronously in an online format. In phase 2, from July to December, those who approve phase 1 based on teachers' assessments and homework submission, gain deeper knowledge in a coding language of their own choosing (e.g., testing, web development) in small synchronous online classes. In phase 3, students are offered work placement assistance in the technology sector.

Besides the eligibility requirements of having completed at least third grade of high school and being at least 18 years old, applicants must score at least 50% in an entrance exam administered online in January.[6] The entrance exam is a 64 question multiple-choice test evaluating four subjects in the following order: verbal (21 questions), math (20 questions), concentration (9 questions) and logic (14 questions).[7] Each correct answer is worth one point, and there is no penalty for wrong answers. Given that applicants take the exam online, to minimize cheating, the agency randomly assigns applicants to take different exam versions. The exam is graded by a computer, so there is no scope for manipulation of the scores. In 2023 there were seven exam versions, all designed to have a similar level of difficulty (see details in Appendix B).[8] There is a time limit of 180 minutes to respond the exam, however, the session does not expire and some applicants stay longer in the exam, which disqualifies them automatically.[9]

2.3 Data, Randomization and Analytical Sample

The data are administrative records from the 2023 admissions, enrollment in phase 1 and continuation into phase 2. The dataset contains exam-related data: individual baseline characteristics collected at exam registration, overall and subject-specific performance, admission decisions, total time spent solving the exam, and for the applicants in *Nudge + Stress*, self-reported stress and responses to the stress reappraisal exercise. We also have response-level data for each of the 64 questions, indicating whether applicants answered and whether the answers were correct. Post-exam,

[6] This cutoff is not communicated to applicants. They are encouraged to perform as best as possible.

[7] Concentration questions generally involve counting the number of times a letter or number appears in a sequence and similar (tedious) tasks as those frequently used in real-effort experiments. They do not require ex-ante knowledge but rather measure effort.

[8] Some questions may be repeated across versions or if questions are different, they are of similar difficulty. We know of no other measures that the agency uses to prevent cheating.

[9] This disqualification rule is not clearly communicated. However, we see that less than 3% of applicants exceed the time limit. The program website informs applicants that the exam takes 2.5 hours.

we include teachers' assessments of performance after phase 1 and students' decisions to proceed to phase 2 of the program.

To construct the treatment and control groups we use the different exam versions administered by the program. In 2023, 7 different exam versions were available. Applicants assigned to versions 1, 6 and 7 were in the *Nudge* treatment, while applicants in version 4 were in the *Nudge + Stress* treatment. Applicants in versions 2 and 3 were in the *Control* treatment.[10] From a total of 4,943 exam takers across treatments and control exam versions, we exclude 160 (3.0%) who do not have information on gender, 1 observation that does not have information on level of education, and 149 (3.2%) who take longer than 4 hours to complete the exam.[11] The final sample is 4,658, with 1,530 in *Control*, 2,417 in *Nudge* and 711 in *Nudge + Stress*.

2.4 Validity of the Design

Table 1 shows, for a list of 21 baseline covariates, the balance between control and treated applicants (the treatment group includes both *Nudge* and*Nudge + Stress*) in columns 1-3, and gender differences in these covariates in columns 4-6. We find that covariate means across treatment and control are not statistically different.[12] The gender differences in the sample are substantial and interesting. Women constitute about 45% of the applicants to the program, so our context is different to others where very few women intend a STEM program, for example in the setting described by Carlana and Fort (2022). In general, women appear to more often come from more disadvantaged backgrounds. For example, they are more likely than men to be from a low SES household, to be a parent and to be unemployed and actively seeking a job. Moreover, women are less likely than men to own a computer, to have a parent with tertiary education, to have been in a STEM track in prior education, and to have prior knowledge of coding. However, they may be more motivated or positively selected in unobservables than men since they are more likely to be working toward or already have a university degree. Given these gender differences, we present our main estimates controlling for the full set of baseline covariates and provide additional evidence to understand whether our results are driven by differences in covariates in Section 5.

Additionally, since the randomization was conducted across exam versions and not within,

[10] We exclude version 5 because we found it to be slightly more difficult than the rest using the data from years 2021 and 2022 (see details in Appendix B).

[11] We think they did not take the exam seriously since the average completion time is less than 110 minutes.

[12] The table does not impute missing values to present the covariate means. However, to control for baseline covariates in the regressions while keeping the sample size intact, we impute the missing values and add indicators for missings. Testing for balance using that larger set of covariates, we find that one of 21 variables is not balanced. The F-statistic of the joint balance test is 1.14 with a p-value equal to 0.267.

we are careful to assess whether the treatment and control versions are indeed equivalent. For example, differences in the difficulty of exam versions could confound positive treatment effects with the treatment being coincidentally assigned to easier exam versions. We use the identical 7 exam versions implemented in 2021 and 2022 to evaluate differences in difficulty by running a regression on the likelihood of obtaining a correct answer on indicators for the different versions using question-by-question data (see Table A9). The point estimates are small and insignificant except for version 5, which seems harder than the rest and which we exclude from the analysis for that reason. We also present descriptive evidence that there is no differential selection of applicants into different versions by showing that each version contains similar fractions of applicants who are female, have college or higher education and come from a low socioeconomic background (see Table A10). In sum, we do not find evidence that control versions are different from the treatment versions (see details in Appendix B).

3 Intervention Details and Empirical Strategy

The *Nudge* and *Nudge + Stress* treatments are designed to address different types of constraints that may differentially affect women's performance. At the core of these interventions is the idea that some exam takers may be performing below their potential due to factors unrelated to their underlying academic ability (e.g., Duquennois, 2022; Franco and Povea, 2023).[13] One such factor could be risk aversion or low self-confidence, which may prevent more risk-averse or underconfident applicants from answering every question if they are unsure of the answer. The *Nudge* treatment, which encourages applicants to "complete as many questions as they can," may partially alleviate this barrier. Similarly, exam takers who believe that the stress they experience during the exam negatively affects their performance may underperform due to diminished cognitive resources. Psychological theory posits that stress, through intrusive thoughts about potential failure, can deplete working memory resources that would otherwise be devoted to solving exam questions (Beilock, 2011). The *Nudge + Stress* intervention may mitigate this effect by offering an alternative interpretation of stress responses, in addition to encouraging participants to answer all possible questions.

[13] These two papers examine "external" constraints that can impact students' performance, such as exam design features like math questions framed around monetary themes and variations in the positioning of correct answers in multiple-choice tests. The *Nudge* and *Nudge + Stress* treatments are more related to "internal" constraints, which refer to how the exam taker is feeling while taking the exam.

3.1 *Nudge* Treatment

Exam versions assigned to this treatment (versions 1, 6, and 7) add three sentences right before starting with the first exam subject. The English version of the text reads as follows: "*The test will begin now. Remember that incorrect answers do not subtract points. Try to complete as many questions as you can!*" The remaining of the exam looks exactly as the *Control* versions, where applicants simply go through different screens displaying the questions.

3.2 *Nudge + Stress* Treatment

Applicants assigned to this treatment (version 4) follow these steps: First, they enter the online platform where the exam is provided, and answer a question about "how anxious they feel at this moment." Second, they read a paragraph explaining stress responses and how to choose an interpretation of stress and its physiological signals. Third, they answer a reflection question related to the paragraph they just read. Fourth, they read the three sentences included in the *Nudge* treatment. Fifth, they begin the exam. Finally, after completing the verbal and math sections, they see a short text reminding them of the stress responses text they read earlier, are informed that it may help to use techniques to reduce anxiety, and are prompted to practice one of three briefly explained techniques for 30 seconds.

We provide a summary of the main points from the stress treatment below, with full English translations available in Appendix A. Including all the text applicants read and their response to the question, the intervention takes up no more than one page.

Stress reappraisal text. The text describes stress responses as follows: "... *When our bodies experience a stress response, our minds also produce an emotional response. In this way, the body and mind work together. But the emotional response we have depends in large part on how we choose to interpret stress and arousal. If we interpret the state of physiological arousal as negative, we experience negative emotions such as fear and threat. Instead, if we interpret physiological arousal as positive, then we experience positive emotions such as arousal and anticipation. People who respond really well to stressful situations are those who interpret their body's physiological arousal in a positive way: they get excited because their body is ready for peak performance during a test, a game, or a presentation.*" The text then prompts students to write: "Explain in 1 or 2 sentences why the following statement is true: '*The body's response to stress is an adaptation: it leads to a better physiological state.*'" This text and question were not framed as a separate part from the test. Since there was a question to respond, applicants may have thought that the question was part of the exam itself.

Meditation prompt. After applicants solve the verbal and math questions, they read: "*Remember that in a previous assignment it was argued that people experience a physiological stress response in many situations, e.g., taking a course exam. That stress response is necessary for increased alertness and responsiveness.*" Then they see some examples of techniques to reduce or attenuate anxiety: Taking full, deep breaths; visualizing in your mind a place that produces calm; progressive muscle relaxation. The text is followed by the prompt: "*Of these three, choose a technique and spend the next 30 seconds simply breathing deeply (you can try inhaling in 4 times and exhaling in 6), visualizing a place of calm or relaxing your body. Try doing it by closing your eyes.*" Subsequently, the applicants continue working on the concentration and logic exam questions.[14]

3.3 Econometric Specification

Our econometric specification involves estimating the gender gap for outcomes such as exam performance, omitted questions and admissions:

$$\begin{aligned} y_i = \beta_0 + \beta_1 \text{Nudge}_i + \beta_2 \text{Nudge+Stress}_i + \beta_3 \text{Male}_i + \beta_4 \text{Nudge}_i \times \text{Male}_i + \\ \beta_5 \text{Nudge+Stress}_i \times \text{Male}_i + X_i \gamma + \varepsilon_i \end{aligned} \tag{1}$$

The coefficient β_1 provides the effect of the nudge on women, β_2 provides the effect of the *Nudge + Stress* intervention on women, β_3 represents the gender gap in the control group, β_4 is the difference-in-differences (DID) coefficient measuring the change in the gender gap generated by the nudge, and β_5 is the DID coefficient measuring the change in the gender gap generated by the combined nudge and stress intervention. We control for the full set of baseline covariates available (X_i) since we observe large differences in observables between men and women and the results without controls seem to be relatively upward biased (see Tables A4 and A3). At the bottom of each results table we provide the p-value testing whether the point estimates from the two treatments are different and the mean outcome for women in the control group.

[14] While possible, it is not easy to go back to questions since applicants would need to click the back button as many times as necessary to go back to a previous question.

4 Results

4.1 Descriptive Evidence on Gender Differences in Exam Performance

We begin by examining the performance gender gap in the entrance exam for the Coding Program in the absence of the treatments. Figure 1 Panel (a) shows the distribution of exam performance by gender in the control group. While most applicants score above the 50% cutoff, the distribution of women's scores shows a bigger left tail and less density above the cutoff compared to men's scores. Overall, larger shares of women than men have low scores in the entrance exam. On average, women in the control group score 57% while men score 20.8 pp higher (Table 2, Column 1).

4.2 Effects on Exam Performance

Figure 1, Panels (b) to (e) show histograms of raw performance by treatment for women and men, respectively. In Panels (b) and (d), we observe fewer women with very low scores due to the treatment, especially under the *Nudge + Stress* intervention, where the distribution appears to shift from the left to the right of the cutoff. This suggests that the treatment benefits women across the left tail of the score distribution, not just those near the cutoff. In contrast, the interventions have no effect on the score distributions of men, as shown in Panels (c) and (e), since fewer men scored very low initially. Instead, the *Nudge + Stress* treatment appears to boost the scores of men who were already performing at higher levels.

Table 2 adds baseline covariates and presents the results for the overall score and standardized scores for the individual exam subjects. Women's overall scores (Column 1) increase substantially with both treatments: by 8.2 percentage points in *Nudge* and 18.2 percentage points in *Nudge + Stress*, relative to the control group mean of 57%. This indicates that the combined intervention has an effect size twice as large as the nudge alone (p-value = 0.044). The DID coefficients show negative and slightly smaller effect sizes than the main effects for women, suggesting that the gains observed for women are not present for men. The coefficients' slightly smaller magnitude and with opposite sign for men suggest that women's gains are nearly offset by a corresponding decrease for men.

By exam subject (Columns 2-5 of Table 2), we find large performance gaps in the control group, with women scoring around 0.2 SD lower than men in all subjects except verbal, where women underperform by 0.076 SD (significant at the 10% level). The *Nudge* treatment substantially boosts women's performance in verbal (0.11 SD) and math (0.095 SD) but has no effect on concentration and logic. The DID coefficients are negative and similar in size to the positive effect for women in math, again indicating smaller gains for men. In the *Nudge + Stress* treatment, women experience notably higher performance gains across nearly all subjects: verbal (0.226 SD), math (0.175 SD), and concentration (0.208 SD), while the effect on logic is smaller and not statistically significant. The DID coefficients are statistically significant and almost equal to the positive treatment effect for women, except in concentration, where the DID coefficient is smaller and not significant, suggesting that gains in concentration are comparable for both men and women.

We offer a few observations on our results. First, the relatively small gender gap favoring men in verbal is completely reversed in both treatments, consistent with findings that women often outperform men in verbal subjects, a domain stereotypically associated with females (Coffman, 2014). Second, the interventions have the strongest effect on the first exam subject, which immediately follows the treatment-related information. Third, it is not surprising that there are no large effects in the last exam subject as the ability to sustain effortful mental activity declines as an exam progresses (Brown et al., 2022). Fourth, while the effects diminish as the exam progresses, they reappear in the concentration section—the second-to-last subject—following the meditation prompt in the *Nudge + Stress* treatment. This result could suggest that the meditation prompt reinforces the main stress reappraisal intervention or acts as a beneficial break. Both interpretations offer valuable insights for exam design.

To contextualize our findings, we turn to the broader education literature, as we are unaware of previous interventions at the moment when students take an exam. The effect size of the *Nudge + Stress* treatment on overall performance is comparable to that of interventions like providing incentives to teachers to take daily photos with their students to reduce absenteeism (Duflo et al., 2012). Our effects are also substantially larger than those of costly programs, such as the One Laptop per Child initiative, which has shown null effects across various countries (Cristia et al., 2017; Yanguas, 2020; Falck et al., 2018). Next, we look into how these gains were achieved by looking at omitted questions and response accuracy.

4.3 Effects on Omitted Questions and Accuracy

Overall exam scores depend on both the number of questions answered correctly and the number attempted, with unattempted (omitted) questions typically counted as incorrect, as is the case in the Coding Program entrance exam. To understand the performance effects, we draw on the economics literature showing that women are more likely than men to omit questions (Espinosa and Gardeazabal, 2010; Baldiga, 2014; Pekkarinen, 2015; Riener and Wagner, 2017; Coffman and Klinowski, 2020;

Atwater and Saygin, 2020), even when there are no penalties for incorrect answers (Iriberri and Rey-Biel, 2021; Karle et al., 2022). In this section, we examine which aspects of performance the intervention influenced to raise scores among female applicants, focusing on omitted questions and accuracy rates (correct/attempted), as both directly impact the total number of correct answers.

Table 3, Panel A presents the results for omitted questions using the main specification. On average, women in the control group omit 16.9% of the 64 questions (equivalent to 10.8 questions). The point estimate for control men is -6.7 pp, indicating that men omit only 6.5 questions on average. The *Nudge* treatment slightly reduces the fraction of omitted questions among women by 2.5 percentage points, while the *Nudge + Stress* treatment reduces it by 5.5 percentage points, nearly closing the gender gap in omitted questions seen in the control group. The DID coefficient, with the opposite sign and similar effect size, suggests that the intervention does not reduce men's omitted questions.

By reducing omitted questions, the *Nudge + Stress* treatment also decreases the proportion of women leaving the exam entirely blank and increases their overall exam completion rate (Table 4, Columns 2 and 3). In the control group, 2.4% of women leave the exam blank, a rate twice as high as for men. Control women complete 83.1% of the exam on average, and the *Nudge + Stress* treatment raises this by 5.5 percentage points, mirroring the reduction in omitted questions. As a result, it is now clearer whether an incorrect answer reflects a genuine error rather than an omission, allowing male and female applicants to be evaluated on more equal grounds.

Similar patterns of omitted questions and treatment effects are present in each individual exam subject (Table 3, Columns 2-5). Control women omit 7.3% of the verbal questions, 18.3% of the math questions, 23.4% of the concentration questions, and 25.2% of the logic questions. The point estimate for men is always negative and significant at the 1% level, indicating that men tend to omit fewer questions in every subject of the exam. The effects of the *Nudge + Stress* treatment are -3.3 pp in verbal, -6.5 pp in math, -5.9 pp in concentration, and -7.3 pp in logic. These effects reduce the gender gap in omitted questions in the control group by at least half.

Turning to accuracy rates, attempting more questions may not result in higher exam scores if the newly attempted questions are incorrect. However, even if the accuracy rate remains the same as in the control group, attempting more questions will increase the fraction correct and overall score. In Table 3, Panel B, we see that the overall accuracy rate for control women is 66.6% and only 1.6 pp higher for control men (68.2%). We do not find strong effects of the treatments on overall accuracy (Column 1), but only in a subset of the exam subjects. For example, the *Nudge* treatment increases accuracy for women by 1.3 pp in verbal and 2 pp in Math. The *Nudge + Stress* treatment increases accuracy for women by 2.7 pp in verbal and 4.3 pp in concentration, although not statistically significantly so in math. The DID coefficients are not significant and not always have negative signs, suggesting that men's accuracy rates may have also slightly improved with the intervention, especially in concentration. All in all, there are some small gains in accuracy rates, but this is not the main margin that the intervention is affecting.

Finally, we explore the dynamics of omitted questions, accuracy rates and fraction correct in Appendix Figure A2. Constructing question deciles as in Brown et al. (2022),[15] we observe that the *Nudge + Stress* substantially reduces the fraction of omitted questions among women across all question deciles (Panel (a)), but not among men (Panel (b)). Accuracy rates remain similar across treatments (Panels (c) and (d)), and the fraction correct again shows gains for women in *Nudge + Stress* across all question deciles, while there is no effect for men. The findings for women suggest that the effects are not only an artifact of trying harder in subjects in which women may feel more confident, but rather that the increase in effort is sustained along the whole exam.

The results on omitted questions and accuracy rates provide the key insight from our research: Strategies encouraging women to omit fewer questions can meaningfully reduce gender gaps in exam performance. Our findings suggest that helping women attempt more questions can significantly boost their overall scores, even if accuracy rates remain the same. This implies that women may be "leaving money on the table" by omitting questions in the absence of penalties for incorrect answers. This insight is particularly encouraging, as gender differences in omitting questions and willingness to guess are well-documented, yet few studies have effectively identified ways to bridge these gaps (Iriberri and Rey-Biel, 2021).

4.4 Effects on Admissions, Program Continuation and Graduation

Table 4 presents results regarding admission into the program, continuation to phases 1 and 2, and graduation. Column 1 shows the gender gap in admissions equal to 8.5 pp in the control group. While 74% of male applicants in the control are admitted to the program, 65.5% of female applicants gain admission. The *Nudge* treatment does not affect admissions for women, but the *Nudge + Stress* condition increases the admission rate of women by 5.9 pp. Consequently, the gender gap in admissions is reduced by almost 70%. The DID coefficient equal to -9.1 pp, similar in size and of

[15]The 64 questions are divided in 10 equally sized groups. We compute the mean of the variable in each decile and overlay a kernel-weighted local polynomial regression to more clearly see the patterns across the exam.

opposite sign as the coefficient on Male, indicates that male applicants' admission does not improve to the same extent as women's as a result of the *Nudge + Stress* treatment. In sum, the intervention, which nudges exam takers to attempt all questions and provides guidelines for understanding stress responses, benefits women disproportionately, resulting in a 9% increase in female admissions to the Coding Program.

The effects on continuation from being admitted to phase 1 of the program and from phase 1 to phase 2 are in Table 4, Columns 4 to 6. Non-enrollment occurs either because applicants fail to score above the 50% cutoff or because about 4% of those who qualify choose not to enroll. There are no statistically significant differences by treatment, except for the DID coefficient in the *Nudge + Stress* treatment, which indicates that men are equally likely to enroll in the program in the treatment or the control group.

Between 23-27% of all applicants approve phase 1 by submitting all homework and performing well in the course (Column 5).[16] Similarly, 21-27% of all applicants enroll in phase 2, indicating that most students who approve phase 1 proceed to phase 2 (Column 6). Regardless of treatment assignment, men are more likely to progress through the different phases of the program than women. Reassuringly, women who gained access to the program through the intervention are of no lower quality than those admitted without intervention, as they are equally likely to approve phase 1 and continue to phase 2 compared to control women.

The last outcome related to performance in the program we analyze is graduation (Table 4, Column 7). Of the initial 4,658 applicants, only 13.6% of women and 18.9% of men graduate from the program. There are no effects of the treatments on graduation, suggesting that the women who were admitted as a result of the interventions are of similar quality as the admitted women in the control. Higher progression and graduation rates among men may reflect, for instance, men's higher intrinsic interest in the topics of the program (see Table 1) or uncertain employment prospects available to female programmers in a male-dominated field.

4.5 Heterogeneity Analysis

We perform two heterogeneity exercises in Table A2 using the variables level of education and socio-economic status (SES). We draw four main conclusions for this analysis. First, the largest gender gaps are among the least educated applicants. Men in the control with less than college

[16]The approval rate among those who enroll in phase 1 is about 40%. Some students leave the program after phase 1 due to reasons such as their preferred coding language or a suitable time slot not being available.

perform much better and are much more likely to gain admission than their female counterparts. However, control women with some college or more seem to have higher scores and admission rates than men with some college or more. Second, the interventions differentially benefit women with lower education as the positive effects are concentrated in the Nudge and Nudge + Stress coefficients that capture the effects among treated women relative to control women. In particular, the *Nudge + Stress* treatment increases the admission rate of women with less than some college by 10.5 pp. Third, among low-educated applicants, the interventions do not seem to help men as the interaction between Nudge and Male and Nudge + Stress and Male are negative and of similar size as the positive effect for women, although not always statistically significant. Fourth, in general, there do not seem to be strong differential effects by SES. The only relatively salient result is that low-SES women perform worse than high-SES applicants and there is no differential treatment effect for low-SES applicants. It is worth noting that SES information is missing for 20% of the observations, potentially reflecting non-random patterns in the data. Overall, we conclude that the intervention disproportionately benefits women with lower education while not negatively affecting men of any education level.

5 Underlying mechanisms

In this section, we provide details about theories that can explain why the interventions helped women improve their performance and why it affects women primarily and not men. We discuss alternative stress-unrelated mechanisms that may be behind the results, and provide some direct evidence on how a similar intervention to *Nudge + Stress* affects the perceptions of stress during an exam.

Although the *Nudge* intervention is less effective than the *Nudge + Stress* intervention, it provides valuable insights into gender gaps in omitted questions. By emphasizing the importance of attempting all exam items, the *Nudge* intervention encourages exam takers to engage more deliberately with each question, potentially motivating them to attempt questions they might otherwise skip. This aligns with responses to an open-ended survey question from the 2024 Coding Program entrance exam, where applicants (women only in 2024) were asked why they left questions unanswered. The survey reveals that a majority of women (67%) omitted questions because they felt they did not know the answer, supporting the idea that emphasizing effort through the *Nudge* intervention could encourage more attempts. Additionally, 14% of respondents cited stress or a lack of confidence as reasons for omissions, reinforcing the notion that stress or self-doubt can inhibit guessing, as noted in the literature (Baldiga, 2014).

The *Nudge + Stress* intervention appears to alleviate the mental burden associated with stress and second-guessing. By reframing stress as a potential ally in performance rather than an obstacle, the intervention can help exam takers redirect their focus toward solving exam problems rather than ruminating about potential failure. This aligns with psychological research, including the seminal work of Beilock (2011), which shows that under pressure, individuals often divert working memory to managing worry—such as concerns about failing the exam—leaving fewer cognitive resources available for the task at hand (Ramirez and Beilock, 2011; Jamieson et al., 2018; Schillinger et al., 2021).

Additional support for the idea that the intervention changes perceptions of stress comes from the 2024 survey administered at the end of the exam. Applicants in the *Nudge + Stress* intervention were 13-19% less likely to report that stress reduced their performance and 46-56% more likely to report that stress helped enhance their performance.[17] The question about why women may benefit more from stress or anxiety reduction interventions may relate to the fact that women tend to report higher levels of anxiety than men (Remes et al., 2016; OECD, 2015) and perform worse in competitions after being induced to be stressed (Cahlíková et al., 2020). Cavatorta et al. (2021) find that women who are more anxious at baseline benefit more from training sessions aimed at reducing the attention given to negative stimuli. Similar to our findings, they observe that treated women, but not men, attempt to solve more questions in a cognitive task in the lab.

The following subsections explore several of the main alternative mechanisms that could explain the results.

5.1 Gender Differences in Covariates

Table 1 shows that men and women in the sample differ in many observable dimensions, which suggests that they may differ in unobservables as well. Because of the gender differences in covariates and the fact that the estimates without covariates tend to be upward biased (Tables A3 and A4), we suspect that observables may play an important role in our setting. Nevertheless, the trial is internally valid within gender as we do not find any systematic differences in baseline covariates (Table A5).

[17] An alternative explanation is that respondents simply echoed the messaging of the stress reappraisal exercise rather than genuinely feeling that stress improved their performance. See Appendix C for details on the 2024 data.

The reason why we present our main estimates including controls is to address the potential issue raised by gender differences in covariates. Nevertheless, we conduct an additional robustness exercise in Tables A6 and A7, where we reweigh observations using inverse probability weighting (IPW) by giving more weight to men who are more similar to women in terms of baseline covariates.[18] Given that we see that the treatment effects are similar with different ways of controlling for observable characteristics, we conclude that the effects are not entirely driven by gender differences in these variables.

5.2 Gender Differences in Engagement

Given an initial level of stress, if the intervention reduces stress for everybody and lower stress increases performance, we may see differential effects for women if they are more likely to take up the intervention. For example, Shreekumar and Vautrey (2022) observe a strong selection by gender into their study, with men being only 15% of the pool of people interested in receiving access to a meditation app. Hence, it may be the case that men do not believe in mindfulness techniques, do not think that they can benefit from them or are simply uninterested.

Using several proxies for engagement, and considering that applicants may have believed the writing exercise was part of the exam due to the absence of instructions indicating otherwise, we conclude that men took the stress reappraisal exercise as seriously as women. First, applicants of both genders wrote an average of 35 words in the stress reappraisal exercise, with similar distributions of word counts (see Figure A3). Second, while the data does not reveal whether applicants followed the meditation instructions, took a break, or proceeded directly with the exam, we observe that treated applicants of both genders spent more time on the exam compared to control applicants (Figure A4). This suggests that participants of both genders engaged with the instructions. Finally, the treatment effects are large and significant for both men and women immediately following the meditation prompt, indicating that differences in engagement are unlikely to explain why the intervention primarily benefits women.

[18]Specifically, we first estimate a logit model to predict which characteristics are more predictive of "being a woman" and calculate the propensity score. Then we check whether there is enough overlap in the distribution of the propensity scores by gender. Finally, we obtained the results described above weighting the observations by $\frac{1}{pscore}$ for women and $\frac{1}{1-pscore}$ for men. We show that the covariates do not differ statistically by gender after applying these weights in Table A8.

5.3 Gender Differences in Responses to "Friendly Environments"

Another plausible mechanism is that the interventions helped create a more "friendly" or supportive testing environment, particularly for women. For example, the meditation prompt in the *Nudge + Stress* treatment may have functioned less as a tool for managing anxiety and more as a brief pause or reset, allowing students to approach the remainder of the exam with renewed focus. While this "environmental adjustment" differs from stress reappraisal, it remains a plausible explanation for the observed effects.

Research on stereotype threat suggests that negative stereotypes about a group's performance—such as women in mathematics—can impair outcomes by creating a cognitive burden and distracting individuals from the task (Steele and Aronson, 1995; Spencer et al., 1999). The *Nudge + Stress* intervention may mitigate these effects by signaling that stress is a normal and even beneficial part of test-taking, reducing the fear of failure and the pressure to disprove stereotypes. Alternatively, the intervention may have created a testing environment where women felt less constrained by external expectations, contributing to improved performance.

Evidence from other contexts supports this idea. For instance, studies on mixed-sex versus same-sex competition show that women perform better in same-sex competitions in the lab (Gneezy et al., 2003) and are slower in mixed-sex speedboat races compared to all-women races, while men perform faster in mixed-sex settings (Booth and Yamamura, 2018). Similarly, our 2024 admission data for the Coding Program, where only women were invited to apply to increase female participation, aligns with this pattern. In this setting, we repeated the *Nudge + Stress* intervention and found no significant effects on performance. Women had similar rates of exam completion and omitted questions as men in 2023, suggesting that in the absence of men, women's performance was unaffected by the intervention[19], much like men's performance in 2023. However, due to numerous changes in the 2024 testing environment compared to 2023, we report these results in Appendix C and focus on gender differences in the main body of the paper.

Drawing from prior research and the 2024 data, we conclude that women's perception of the testing environment may significantly influence their performance. While we cannot definitively rule out that the effects of the *Nudge + Stress* intervention stemmed from changes in stress responses, it is equally plausible that the intervention's impact came from creating a more supportive and less intimidating environment. Regardless of the mechanism, conveying that the test or testing environment is not threatening appears crucial in reducing gender gaps in exam performance.

[19] Gomez-Ruiz et al. (2024) found that women performed better on this particular admission test when men were absent.

6 Discussion and Conclusion

Our study proposes that exams may not only capture academic ability or preparation but also reflect other factors such as the impact of stress on performance. In the context of the entrance exam for a popular coding program in Uruguay, we investigate the effects of two interventions: one encouraging students to attempt all questions and another combining the nudge to answer all questions with reframing stress as performance-enhancing. Our results show that women, but not men, benefit significantly from the combined intervention particularly, which reduces existing gender gaps in performance and admissions.

We highlight the importance of the intervention in reducing the number of omitted questions among women. While many papers document a gender gap in omitted questions, even when there are no penalties for wrong answers, how to reduce this gap has proven elusive. We believe our results are very encouraging in this respect and pave the way forward in terms of how to level the playing field in exam performance.

Our findings align with a growing body of research in psychology and economics on the role of stress and anxiety on economic outcomes. Unlike previous studies that have primarily focused on long-term interventions such as mindfulness training (Cassar et al., 2022), our study highlights the potential of real-time, low-cost interventions implemented during exams. These results highlight the importance of designing testing environments that account for psychological barriers and optimize conditions for all test-takers.

The implications of our findings extend beyond this specific trial. Accommodations such as stress reframing prompts, brief meditation exercises, or other targeted interventions can help reduce cognitive and emotional distractions during exams. These measures are both practical and scalable, offering a pathway to create more equitable assessment practices without requiring significant changes to exam structures or administration.

Ultimately, our study highlights the importance of addressing external pressures and psychological barriers to improve the performance of underrepresented groups. By fostering environments that enable individuals to perform to the best of their abilities, we can make meaningful progress in reducing gender disparities in academic and career outcomes.

References

Arenas, A. and C. Calsamiglia (2022). Gender differences in high-stakes performance and college admission policies. Technical report, IZA Discussion Papers.

Atwater, A. and P. O. Saygin (2020). Gender differences in willingness to guess on high-stakes standardized tests. *Mimeo*.

Azmat, G., C. Calsamiglia, and N. Iriberri (2016). Gender differences in response to big stakes. *Journal of the European Economic Association 14*(6), 1372–1400.

Balart, P., L. Ezquerra, and I. Hernandez-Arenaz (2022). Framing effects on risk-taking behavior: evidence from a field experiment in multiple-choice tests. *Experimental Economics 25*(4), 1268–1297.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science 60*(2), 434–448.

Beilock, S. (2011). *Choke*. Hachette UK.

Booth, A. and E. Yamamura (2018). Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in japan. *Review of Economics and Statistics 100*(4), 581–593.

Brown, C. L., S. Kaur, G. Kingdon, and H. Schofield (2022). Cognitive endurance as human capital. Technical report, National Bureau of Economic Research.

Cahlíková, J., L. Cingl, and I. Levely (2020). How stress affects performance and competitiveness across gender. *Management Science 66*(8), 3295–3310.

Cai, X., Y. Lu, J. Pan, and S. Zhong (2019). Gender gap under pressure: evidence from China's National College entrance examination. *Review of Economics and Statistics 101*(2), 249–263.

Carlana, M. and M. Fort (2022). Hacking gender stereotypes: Girls' participation in coding clubs. In *AEA Papers and Proceedings*, Volume 112, pp. 583–87.

Cassar, L., M. Fischer, and V. Valero (2022). Keep calm and carry on: The short-vs. long-run effects of mindfulness meditation on (academic) performance. Technical report, IZA Discussion Papers.

Cavatorta, E., S. Grassi, and M. Lambiris (2021). Digital antianxiety treatment and cognitive performance: An experimental study. *European Economic Review 132*, 103636.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics 129*(4), 1625–1660.

Coffman, K. B. and D. Klinowski (2020). The impact of penalties for wrong answers on the gender gap in test scores. *Proceedings of the National Academy of Sciences 117*(16), 8794–8803.

Cristia, J., P. Ibarrarán, S. Cueto, A. Santiago, and E. Severín (2017). Technology and child development: Evidence from the one laptop per child program. *American Economic Journal: Applied Economics 9*(3), 295–320.

De Paola, M. and F. Gioia (2016). Who performs better under time pressure? Results from a field experiment. *Journal of Economic Psychology 53*, 37–53.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. *American economic review 102*(4), 1241–1278.

Duquennois, C. (2022). Fictional money, real costs: Impacts of financial salience on disadvantaged students. *American Economic Review 112*(3), 798–826.

Espinosa, M. P. and J. Gardeazabal (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical psychology 54*(5), 415–425.

Falck, O., C. Mang, and L. Woessmann (2018). Virtually no effect? different uses of classroom computers and their effect on student achievement. *Oxford Bulletin of Economics and Statistics 80*(1), 1–38.

Franco, C. and E. Povea (2023). Innocuous exam features? the impact of answer placement on high-stakes test performance and college admissions. Technical report, Mimeo.

Geraldes, D. (2020). Women dislike competing against men. *Available at SSRN 3741649*.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The quarterly journal of economics 118*(3), 1049–1074.

Gomez-Ruiz, M., M. Cervini-Plá, and X. Ramos (2024). Do women fare worse when men are around? quasi-experimental evidence.

Harris, R. B., D. Z. Grunspan, M. A. Pelch, G. Fernandes, G. Ramirez, and S. Freeman (2019). Can test anxiety interventions alleviate a gender gap in an undergraduate STEM course? *CBE—Life Sciences Education 18*(3), ar35.

Iriberri, N. and P. Rey-Biel (2019). Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *The Economic Journal 129*(620), 1863–1893.

Iriberri, N. and P. Rey-Biel (2021). Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment. *European Economic Review 131*, 103603.

Jamieson, J. P., A. J. Crum, J. P. Goyer, M. E. Marotta, and M. Akinola (2018). Optimizing stress responses with reappraisal and mindset interventions: An integrated model. *Anxiety, Stress, & Coping 31*(3), 245–261.

Karle, H., D. Engelmann, and M. Peitz (2022). Student performance and loss aversion. *The Scandinavian Journal of Economics 124*(2), 420–456.

OECD (2015). The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. `http://dx.doi.org/10.1787/9789264229945-en`.

Ors, E., F. Palomino, and E. Peyrache (2013). Performance gender gap: does competition matter? *Journal of Labor Economics 31*(3), 443–499.

Pekkarinen, T. (2015). Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. *Journal of Economic Behavior & Organization 115*, 94–110.

Ramirez, G. and S. L. Beilock (2011). Writing about testing worries boosts exam performance in the classroom. *Science 331*(6014), 211–213.

Remes, O., C. Brayne, R. Van Der Linde, and L. Lafortune (2016). A systematic review of reviews on the prevalence of anxiety disorders in adult populations. *Brain and Behavior 6*(7), e00497.

Riener, G. and V. Wagner (2017). Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions. *Economics of Education Review 59*, 43–62.

Schillinger, F. L., J. A. Mosbacher, C. Brunner, S. E. Vogel, and R. H. Grabner (2021). Revisiting the role of worries in explaining the link between test anxiety and test performance. *Educational Psychology Review 33*, 1887–1906.

Shreekumar, A. and P.-L. Vautrey (2022). Managing emotions: The effects of online mindfulness meditation on mental health and economic behavior. Technical report, Tech. Rep., MIT.

Spencer, S. J., C. M. Steele, and D. M. Quinn (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology 35*(1), 4–28.

Steele, C. M. and J. Aronson (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology 69*(5), 797.

USAID (2022). A summary analysis of education trends in Latin America and the Caribbean: 2022 update. Technical report.

Yanguas, M. L. (2020). Technology and educational choices: Evidence from a one-laptop-per-child program. *Economics of Education Review 76*, 101984.

7 Figures

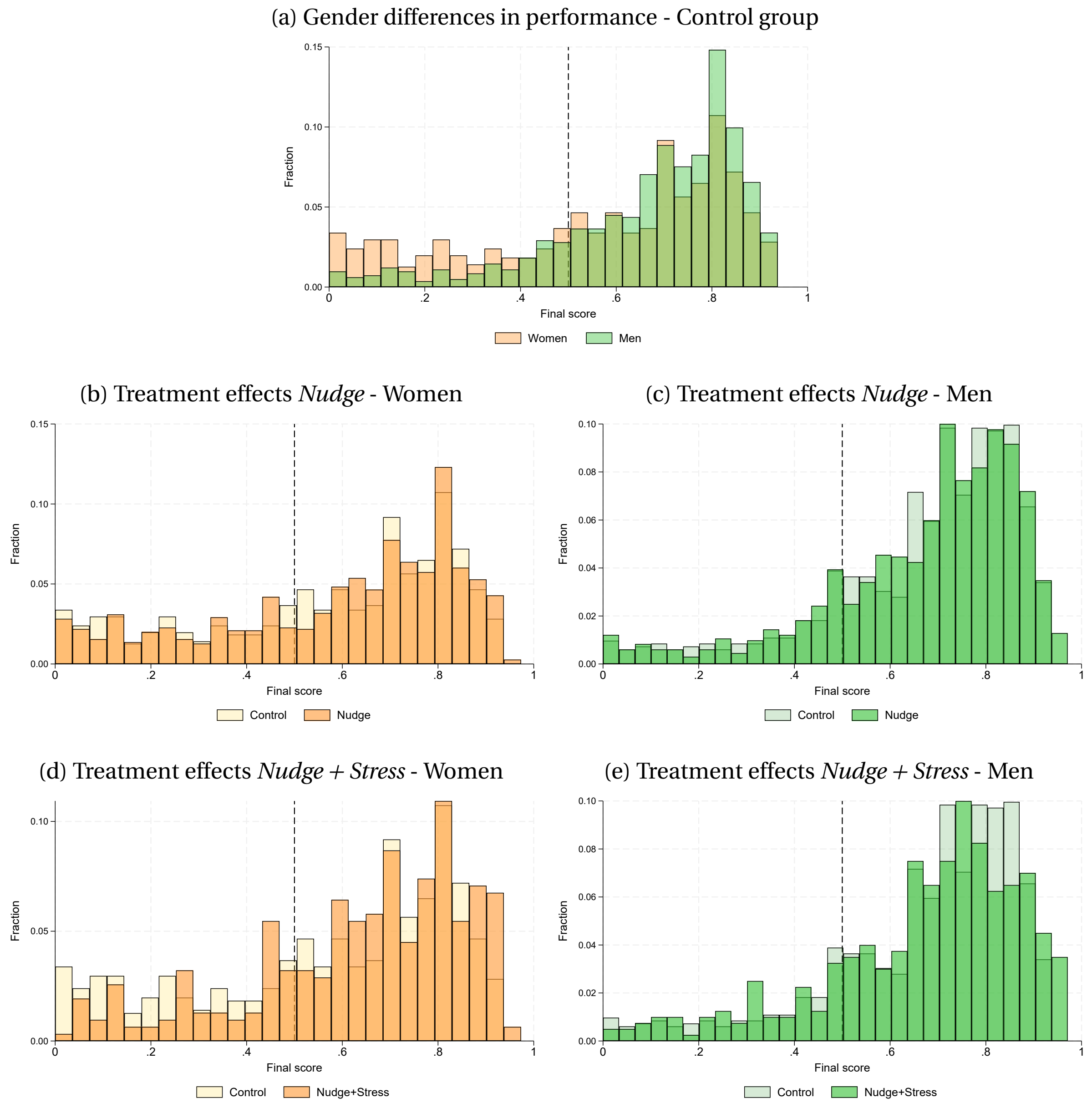


Figure 1: Gender differences in control group performance and treatment effects across the exam performance distribution by gender

Notes: Panel (a) shows final score (raw) for the control group by gender. Panels (b) to (d) plot overall exam performance for treated and control women and men, for the *Nudge* and *Nudge+ Stress* treatments in rows 2 and 3, respectively. The red vertical line represents the cutoff of 50% granting admission to the Coding Program.

8 Tables

Table 1: Covariate balance and differences by gender

	(1) Control	(2) Nudge and Stress	(3) Diff. (1)-(2)	(4) Men	(5) Women	(6) Diff. (4)-(5)
Sociodemographics and applicant education						
Female	0.538	0.550	−0.012	0.000	1.000	−1.000
	(0.499)	(0.498)	(0.016)	(0.000)	(0.000)	(0.000)
Age	23.767	23.750	0.018	24.104	23.466	0.638***
	(3.434)	(3.448)	(0.108)	(3.350)	(3.493)	(0.101)
Secondary or lower	0.561	0.573	−0.011	0.514	0.615	−0.100***
	(0.496)	(0.495)	(0.015)	(0.500)	(0.487)	(0.015)
Some college or higher	0.308	0.302	0.006	0.349	0.267	0.083***
	(0.462)	(0.459)	(0.014)	(0.477)	(0.442)	(0.014)
Other type of education	0.130	0.125	0.005	0.136	0.119	0.017*
	(0.336)	(0.331)	(0.010)	(0.343)	(0.324)	(0.010)
Attended public education inst.	0.883	0.908	−0.024**	0.917	0.885	0.032***
	(0.321)	(0.290)	(0.010)	(0.276)	(0.319)	(0.009)
STEM track	0.214	0.197	0.017	0.121	0.270	−0.149***
	(0.410)	(0.398)	(0.013)	(0.326)	(0.444)	(0.011)
Plan to study something else	0.771	0.767	0.004	0.783	0.756	0.027**
	(0.420)	(0.423)	(0.014)	(0.412)	(0.429)	(0.013)
Prior knowledge of coding	0.211	0.199	0.012	0.121	0.270	−0.149***
	(0.408)	(0.399)	(0.013)	(0.326)	(0.444)	(0.012)
High English level	0.539	0.520	0.019	0.483	0.562	−0.079***
	(0.499)	(0.500)	(0.016)	(0.500)	(0.496)	(0.015)
Household and Socioedemographic characteristics						
Low SES	0.420	0.418	0.002	0.466	0.380	0.087***
	(0.494)	(0.493)	(0.017)	(0.499)	(0.485)	(0.016)
Residing in capital city	0.524	0.538	−0.014	0.526	0.540	−0.014
	(0.500)	(0.499)	(0.016)	(0.499)	(0.499)	(0.015)
Household size	3.126	3.039	0.087	3.037	3.092	−0.055
	(2.120)	(1.728)	(0.062)	(1.796)	(1.922)	(0.055)
Head of household	0.259	0.276	−0.017	0.247	0.290	−0.042***
	(0.438)	(0.447)	(0.014)	(0.432)	(0.454)	(0.013)
Has children	0.138	0.136	0.002	0.205	0.080	0.125***
	(0.345)	(0.343)	(0.011)	(0.404)	(0.271)	(0.010)
Parent with tertiary education	0.334	0.318	0.017	0.295	0.347	−0.052***
	(0.472)	(0.466)	(0.015)	(0.456)	(0.476)	(0.014)
More than 50 books at home	0.283	0.264	0.019	0.289	0.255	0.035***
	(0.451)	(0.441)	(0.014)	(0.454)	(0.436)	(0.013)
Owns computer	0.911	0.906	0.005	0.866	0.942	−0.075***
	(0.285)	(0.292)	(0.009)	(0.340)	(0.235)	(0.009)
Access to internet	0.864	0.873	−0.009	0.830	0.903	−0.073***
	(0.343)	(0.333)	(0.011)	(0.376)	(0.296)	(0.010)
Not working and looking for a job	0.459	0.436	0.023	0.467	0.424	0.043***
	(0.499)	(0.496)	(0.016)	(0.499)	(0.494)	(0.015)
Has private health insurance	0.635	0.647	−0.012	0.627	0.656	−0.028**
	(0.482)	(0.478)	(0.015)	(0.484)	(0.475)	(0.014)
Obs.	1,530	3,128	4,658	2,115	2,543	4,658

Notes: Columns 1 and 2 show baseline covariate means by control and treatment, respectively. Column 3 computes the difference between columns 1 and 2 and shows whether the difference is statistically significant. Columns 4 and 5 show the baseline covariate means by gender, irrespective of treatment assignment. Column 6 tests whether the gender differences are significant. Variable definitions are in Appendix **??**. Standard deviations below the means and standard errors below the differences in parentheses. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$. We run two separate regressions: the first regressing the full set of covariates on treatment status, and the second on gender. In the first regression, we fail to reject the null hypothesis that all covariates are jointly equal to zero (F-statistic = 1.15, p-value =0.254). In the second regression, we reject the null hypothesis (F-statistic = 36.96, p-value = 0.000).

Table 2: Effects on performance

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total score	Verbal	Math	Concentration	Logic
Nudge	0.082**	0.113***	0.095**	0.053	0.031
	(0.039)	(0.041)	(0.041)	(0.042)	(0.040)
Nudge+Stress	0.182***	0.226***	0.175***	0.208***	0.079
	(0.053)	(0.053)	(0.055)	(0.058)	(0.056)
Male	0.208***	0.076*	0.244***	0.167***	0.221***
	(0.040)	(0.041)	(0.040)	(0.044)	(0.043)
Nudge × Male	−0.049	−0.006	−0.081	−0.008	−0.052
	(0.050)	(0.052)	(0.051)	(0.056)	(0.052)
Nudge+Stress × Male	−0.158**	−0.162**	−0.175**	−0.070	−0.129*
	(0.068)	(0.071)	(0.070)	(0.076)	(0.072)
Constant	0.104**	0.086*	0.082*	0.106**	0.110**
	(0.045)	(0.047)	(0.046)	(0.049)	(0.047)
Pval Diff. Nudge-Stress	0.044	0.024	0.120	0.005	0.350
Mean dep.var (women)	0.57	12.63	11.89	4.38	7.64
SD dep.var (raw)	0.266	4.177	6.488	2.945	5.165
Questions	64	21	20	9	14
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	4,658	4,658	4,658	4,658	4,658

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1. At the bottom of the table we report the point estimate and p-value of the treatment effect on women, along with the mean and SD for the outcome before standardization, and the total number of exam questions considered in each outcome. All standardized outcomes are standardized based on the mean and SD of men in the control group. Column 1 displays the estimates for the total score obtained in the entrance exam. Columns 2 to 5 presents the estimates for each exam subject. Verbal and math appeared after the stress reappraisal exercise, and concentration and logical reasoning appeared after the meditation exercise. Robust standard errors in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table 3: Omitted questions and accuracy rate

	(1) Total	(2) Verbal	(3) Math	(4) Concentration	(5) Logic
Panel A: Omitted questions					
Nudge	-0.025*	-0.018**	-0.024	-0.020	-0.038**
	(0.013)	(0.009)	(0.015)	(0.017)	(0.017)
Nudge+Stress	-0.055***	-0.033***	-0.065***	-0.059***	-0.073***
	(0.015)	(0.010)	(0.019)	(0.022)	(0.023)
Male	-0.067***	-0.029***	-0.073***	-0.097***	-0.094***
	(0.012)	(0.008)	(0.014)	(0.016)	(0.017)
Nudge × Male	0.022	0.017	0.023	0.027	0.024
	(0.015)	(0.010)	(0.018)	(0.020)	(0.021)
Nudge+Stress × Male	0.043**	0.029**	0.051**	0.057**	0.042
	(0.019)	(0.012)	(0.023)	(0.026)	(0.027)
Constant	0.152***	0.067***	0.160***	0.205***	0.234***
	(0.013)	(0.010)	(0.015)	(0.018)	(0.018)
Pval Diff. Nudge-Stress	0.028	0.066	0.018	0.050	0.100
Mean dep.var (women)	0.169	0.073	0.183	0.234	0.252
Obs.	4,658	4,658	4,658	4,658	4,658
Panel B: Accuracy rates					
Nudge	0.009	0.013**	0.020**	0.008	-0.016
	(0.007)	(0.006)	(0.010)	(0.012)	(0.013)
Nudge+Stress	0.018*	0.027***	0.020	0.043**	-0.023
	(0.010)	(0.009)	(0.014)	(0.017)	(0.017)
Male	0.016**	-0.002	0.037***	-0.003	0.021
	(0.008)	(0.007)	(0.010)	(0.013)	(0.013)
Nudge × Male	0.002	0.009	-0.015	0.014	-0.001
	(0.010)	(0.009)	(0.013)	(0.016)	(0.016)
Nudge+Stress × Male	-0.016	-0.015	-0.028	0.009	-0.020
	(0.014)	(0.012)	(0.018)	(0.022)	(0.022)
Constant	0.683***	0.656***	0.717***	0.638***	0.741***
	(0.008)	(0.007)	(0.011)	(0.014)	(0.014)
Pval Diff. Nudge-Stress	0.332	0.105	0.990	0.029	0.668
Mean dep.var (women)	0.666	0.643	0.720	0.635	0.725
Questions	64	21	20	9	14
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	4,603	4,603	4,291	4,167	4,130

Notes: The table presents estimates for each outcome variable in the panel and column headers following Equation 1. At the bottom of each panel table we report the point estimate and p-value of the treatment effect on women. Accuracy is defined as the ratio of the number of correct answers over the total questions attempted. Robust standard errors in parentheses. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table 4: Effects on admission, exam completion and program continuation

	Admitted	Exam completed		Continuation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Above cutoff	None	Fraction	Enroll 1	Approved	Enroll 2	Graduated
Nudge	0.018	−0.003	0.025*	0.002	−0.017	−0.002	−0.007
	(0.020)	(0.007)	(0.013)	(0.021)	(0.019)	(0.019)	(0.016)
Nudge+Stress	0.059**	−0.020***	0.055***	0.048	−0.018	−0.007	−0.012
	(0.028)	(0.006)	(0.015)	(0.030)	(0.028)	(0.027)	(0.023)
Male	0.085***	−0.013**	0.067***	0.084***	0.043*	0.057**	0.053***
	(0.021)	(0.006)	(0.012)	(0.022)	(0.023)	(0.022)	(0.020)
Nudge × Male	−0.014	0.004	−0.022	−0.001	−0.006	−0.017	−0.020
	(0.026)	(0.008)	(0.015)	(0.028)	(0.028)	(0.027)	(0.024)
Nudge+Stress × Male	−0.091**	0.020***	−0.043**	−0.094**	−0.034	−0.042	−0.020
	(0.037)	(0.007)	(0.019)	(0.039)	(0.039)	(0.038)	(0.033)
Constant	0.698***	0.026***	0.848***	0.673***	0.254***	0.241***	0.153***
	(0.023)	(0.008)	(0.013)	(0.024)	(0.025)	(0.025)	(0.022)
Pval Diff. Nudge-Stress	0.124	0.000	0.028	0.103	0.960	0.843	0.829
Mean control women	0.655	0.024	0.831	0.629	0.231	0.208	0.136
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	4,658	4,658	4,658	4,658	4,658	4,658	4,658

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1. At the bottom of the table we report the point estimate and p-value of the treatment effect on women. Column 1 displays the estimates of the probability of program admission. Column 2 presents the estimates of the likelihood of answering zero questions. Column 3 reports the estimates of the fraction of the exam completed. Column 4 displays the estimates of taking longer than 180 minutes to complete the exam. Column 5 reports the estimates of the probability of enrollment in Phase 1 for those who are admitted. Column 6 reports the estimates of the likelihood of approving Phase 1. Column 7 reports the estimates of the probability of enrollment in Phase 2 for those who approved Phase 1. 792 students did not make the entrance exam cutoff. Robust standard errors in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

A Stress Management Exercises Prompts

Below are the English translations of the prompts used in the mindfulness exercises. The intervention was designed by the agency based on the prompts outlined in Harris et al. (2019). The highlighting is ours to help the time-constrained reader skim.

A.1 Stress reappraisal prompt

There are many situations (for example, a music recital, an athletic competition, a course exam, or a job interview) in which people experience a physiological stress response. This stress response is necessary to increase alertness and responsiveness. Humans can respond with peak performance in stressful situations because we become in a state of physiological arousal, which puts our body in a state of alertness, ready for action. When our bodies experience a stress response, our minds also produce an emotional response. In this way, the body and mind work together. But the emotional response we have depends in large part on how we choose to interpret stress and arousal. If we interpret the state of physiological arousal as negative, we experience negative emotions such as fear and threat. Instead, if we interpret physiological arousal as positive, then we experience positive emotions such as arousal and anticipation. People who respond really well to stressful situations are those who interpret their body's physiological arousal in a positive way: they get excited because their body is ready for peak performance during a test, a game, or a presentation.

Explain in 1 or 2 sentences why the following statement is true: *"The body's response to stress is an adaptation: it leads to a better physiological state"*

A.2 Meditation prompt

Before continuing with the test...

Remember that in a previous assignment it was argued that people experience a physiological stress response in many situations, e.g., taking a course exam. That stress response is necessary for increased alertness and responsiveness. It has been observed that for some people it is beneficial to perform some techniques to reduce or attenuate anxiety:

1. Deep, full breaths, with exhalations longer than inhalations, are helpful in calming the mind.
2. Visualize in your mind a place that produces calm: it can be a silent beach, a forest full of green trees and flowers, or floating in the sea without any worries.

3. Progressive muscle relaxation, which consists of bringing one-to-one attention to each muscle in the body, contracting it first, and then relaxing it completely.

Of these three, choose a technique and spend the next 30 seconds simply breathing deeply (you can try inhaling in 4 times and exhaling in 6), visualizing a place of calm or relaxing your body. Try doing it by closing your eyes.

B Equivalence and Selection of Exam Versions in Analytical Sample

The randomization of the treatment took place across exams due to ease of implementation by the agency. The agency administers seven different exam versions to avoid cheating, since the exam is administered online and without any camera or device that monitors students. Naturally, if the exam versions are not calibrated to have the same level of difficulty, we may confound positive performance and admission effects in the treatment group with students facing harder exams in the control group. To provide evidence that this is not the case, we perform two design validity exercises to demonstrate that the exam versions included in the study are indeed equivalent.

The first exercise uses data from applicants in 2021 and 2022 who faced the exact same versions and questions as in 2023. With question-level data, we assess the overall likelihood of answering a question correctly. We regress the likelihood of answering correctly a given question in the exam on indicators for the exam version and interactions of exam version with the female indicator for the years 2021 and 2022. The aim is to compare the exam versions not incorporating stress reappraisal exercises with version 4, which includes these exercises in 2023. We add question fixed effects in all regressions. Table A9 presents the results of this analysis for 2021 in Columns 1 and 3 and for 2022 in Columns 3 and 4. We find that for all exam versions except version 5, the likelihood of answering correctly is not significant. Version 5 (part of the control in 2023) seems to be slightly harder than version 4 and all other exam versions. We therefore exclude version 5 from the analysis, since including it would overestimate our results by having lower scores in the control group.

In the second exercise, we present some characteristics by exam version in Table A10. We including all exam versions and not only the ones that are part of the analytical sample. Versions 1, 4, 6 and 7, the treatment versions, do not stand out relative to the control versions in terms of the fraction of applicants who are female, have university or higher education, and are from a low SES background.

C 2024 Intervention

In 2024, Ceibal, the agency in charge of the Coding Program, intended to replicate the 2023 intervention and collect a new wave of data. Following our advice, the intervention was randomized within version. Ceibal also wanted to test the impact of the order of the exam subjects. For example, whether the effects are similar when verbal questions appear first than when math questions appear first. We pre-registered the trial and a pre-analysis plan in the AER RCT registry (AEARCTR-0012720). However, due to a budget cut, Ceibal decided to only accept female applicants in 2024. Hence, the applicants faced substantially different conditions from the regular admissions. In what follows, we describe the details of the design, briefly describe the results, and offer some explanations for why the conditions were different than the typical admissions, which we believe help explain why the 2023 results do not replicate in 2024.

In 2024, due to the exclusively female applicant pool, only four of the seven exam versions were administered (versions 1, 4, 6, and 7). The 2024 intervention included three randomly assigned treatment arms within these four exam versions. The intervention consisted of two treatment groups and one control group.

The first treatment arm (T1) replicated the structure used in 2023, including a stress reappraisal exercise followed by sections on verbal skills, math, meditation, concentration, and logic. The second treatment arm (T2) maintained the same stress reappraisal exercise as T1 but altered the order of sections before the meditation exercise, with math preceding verbal skills. The control group received the exam without any stress management exercises, although the nudge to attempt all questions was included, and followed the same section order as T1.

To test whether the intervention had an effect on the performance, omitted questions and admissions outcomes, we estimate the average treatment according to the following equation:

$$y_i = \beta_0 + \beta_1 \mathrm{T1}_i + \beta_2 \mathrm{T2}_i + \gamma_j + \varepsilon_i \tag{2}$$

where the coefficient β_1 provides the treatment effect for T1 (verbal before math), β_2 represents the treatment effect for T2 (math before verbal), and γ_j are exam version fixed effects. We use robust standard errors in all result tables. The balance of covariates for the 2024 sample is in Table A11.

Table A12 presents the results on performance, exam completion and admission. The intervention did not improve the total exam score overall. However, there was a substantial increase in math scores (0.18 SD) in T2, where math was the first section following the stress reappraisal exercise. This increase in math was offset by a significant decrease in verbal scores, leaving the overall score unchanged compared to the control group. In T1, where verbal preceded math as in 2023, no significant effects were observed except for a small improvement in concentration, the subject following the meditation exercise. Consequently, the intervention did not significantly increase the likelihood of admission to the program.

The intervention did, however, impact the likelihood of leaving the exam completely blank. Specifically, applicants were more likely to leave the exam blank when it began with the math section, an effect not observed when verbal was the first section, where almost no one left the exam blank. This suggests that the subject order influences the likelihood of engaging with the exam and completing more questions. In particular, it is interesting that beginning the exam with math, a subject where women may feel less confident, makes a substantial amount of women give up. We think this finding provides evidence that subject order may have first order effects on exam performance.

One of the main takeaways from the 2024 intervention is that women in 2024 exhibited behavior similar to men in 2023 under identical interventions (*Nudge + Stress* in 2023 and T1 in 2024). For example, the fraction of applicants leaving the exam blank was near zero in 2024, comparable to the fraction of men leaving it blank in 2023, whereas in 2023 women were twice as likely to leave the exam blank as men. Additionally, the fraction of the exam completed in 2024 was 88%, closely aligning with the 90% completion rate for men in 2023 and significantly higher than the 83% completion rate for women in 2023.

We do not report the results from 2024 in the main text because the setting differs across several dimensions from the 2023 intervention. These differences may explain the lack of observed effects, but it is not possible to determine precisely what drives them. The female-only setting altered two key features of the admissions process.

First, the nature of the competition for slots changed substantially, as there were fewer slots but also fewer applicants. For example, in 2023, the admission rate for women was 66%, while in 2024, it was 72%. The agency's decision to exclude men aimed to reduce the number of students in the program due to budget cuts. However, this adjustment likely made the program less competitive, potentially lowering the stakes for applicants.

Second, the competition shifted to a single-sex environment, where applicants competed only against other women. Evidence from 2019, another year when only women participated, suggests that women's performance improved under such conditions (Gomez-Ruiz et al., 2024).[20] More broadly, research suggests that women respond differently to single-sex versus mixed-gender competition. For example, the seminal study by Gneezy et al. (2003) found that women increase their performance when competing against other women but not when competing against men in laboratory settings. Similarly, Booth and Yamamura (2018) observed this effect in the field, showing that women performed better in women-only speedboat races in Japan, while Geraldes (2020) highlights that women may actively dislike competing against men.

Overall, the unique conditions of the 2024 admissions process likely influenced applicant behavior, making it difficult to draw definitive conclusions about gender gaps in exam performance from the 2024 data.

[20] The improved performance in 2019 has been attributed to the deactivation of gender stereotypes in women-only environments.

D Appendix Figures

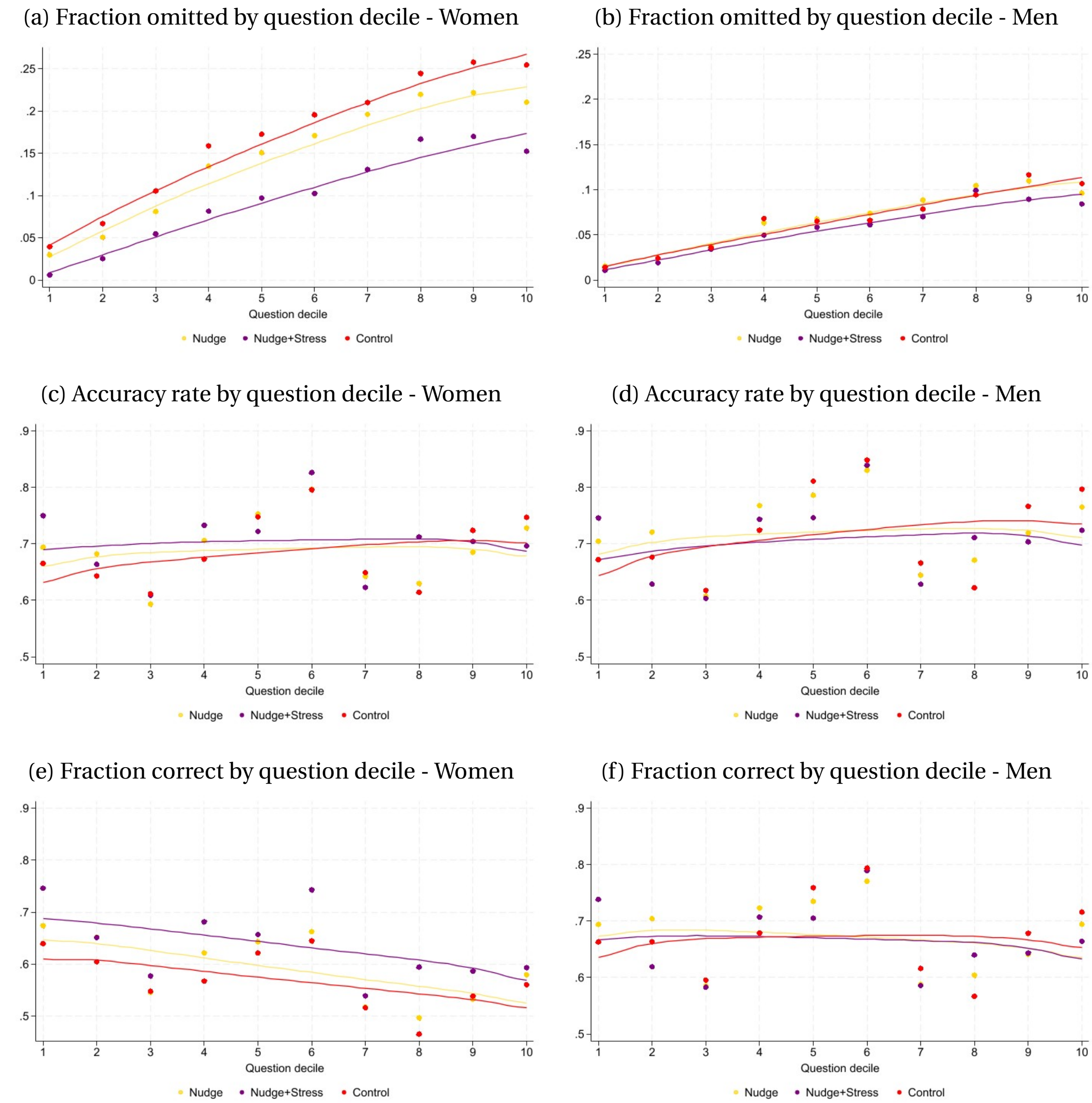


Figure A2: Fraction omitted, correct and accuracy rates by question decile and gender

Notes: Question deciles computed using the 64 questions in the exam. The question order is the same across all exam versions, but not all exams contain identical questions (see Appendix B). All plots show, for treatment and control applicants separately, the mean fraction of omitted, correct questions and accuracy rates by question decile. We overlay a kernel-weighted local polynomial regression, with the width of the smoothing window around each point equal to 3. Panels (a) and (b) show the fraction of omitted questions by decile. Panels (c) and (d) show the accuracy rate defined as correct over attempted. Panels (e) and (f) show the fraction of correct answers by decile, counting omitted questions as incorrect answers.

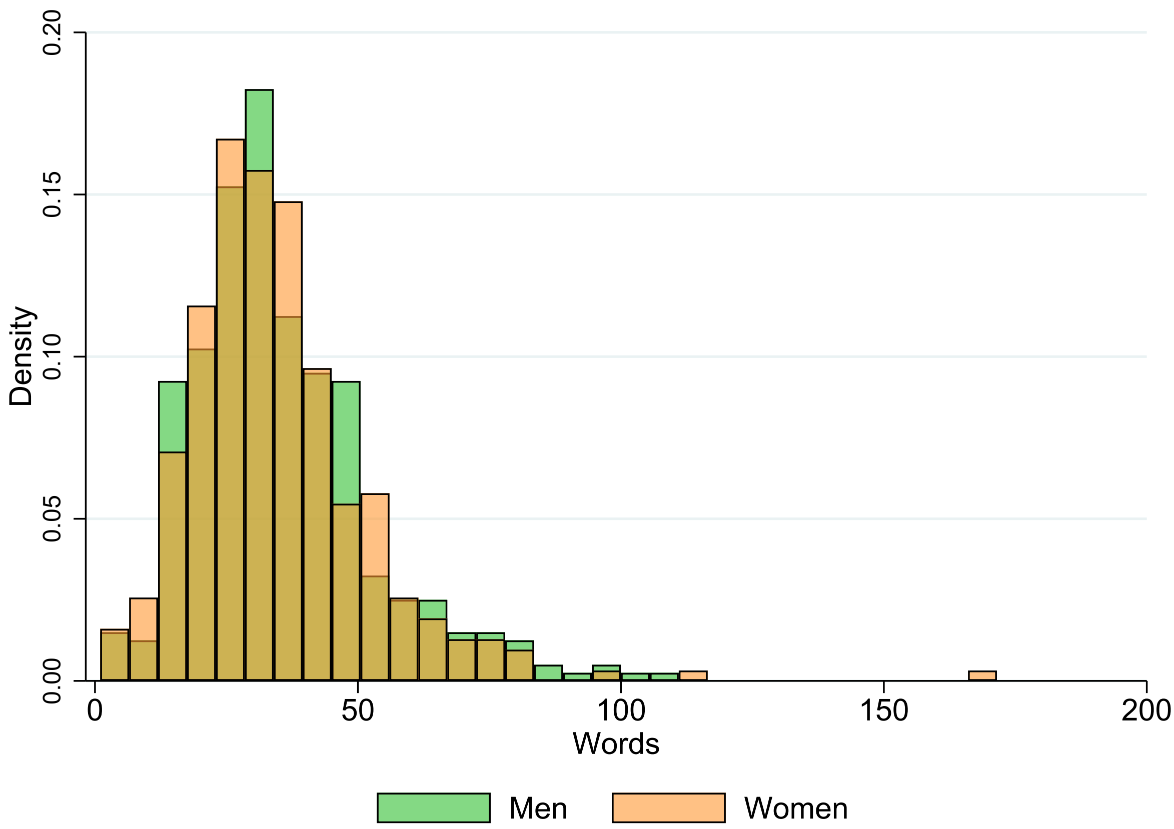
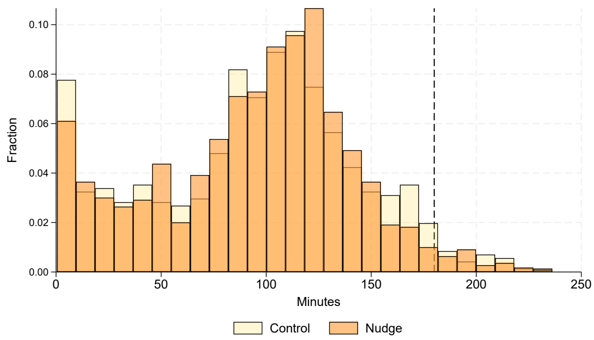


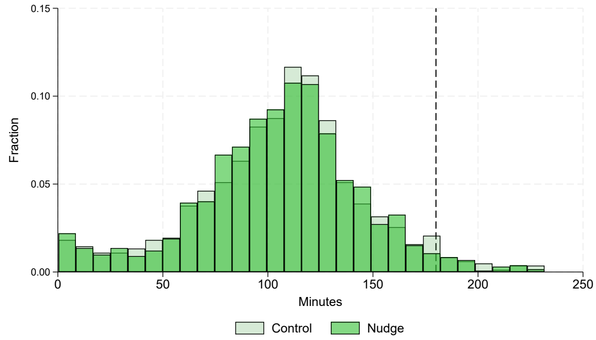
Figure A3: Gender differences in number of words written (treatment group only)

Notes: The graph shows the distribution of the number of words written after the stress reappraisal prompt by gender. This question is only for the *Nudge + Stress* treatment group.

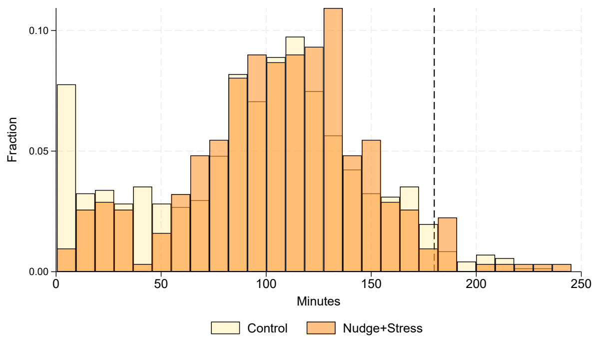
(a) Time spent in *Nudge* - Women



(b) Time spent in *Nudge* - Men



(c) Time spent in *Nudge + Stress* - Women



(d) Time spent in *Nudge + Stress* - Men

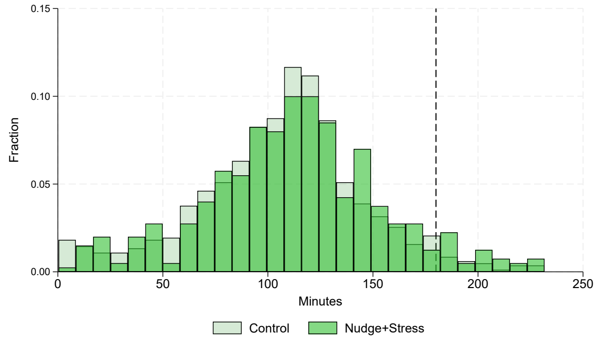


Figure A4: Gender differences in time spent answering the exam by treatment and gender

Notes: The histograms overlay time spent in bins for control and treatment applicants. Control applicants are represented in light-shaded bars, while treated applicants are in dark-shaded bars. Panels (a) and (b) show time spent in the *Nudge* treatment for women and men, respectively. Panels (c) and (d) show time spent in the *Nudge + Stress* treatment for women and men, respectively. The dotted vertical line represents the time limit beyond which applicants are disqualified from the admission process. Although the online platform allows them to continue answering, their responses are not considered for admission.

E Appendix Tables

Table A1: Covariates definition

Covariate	Definition
Sociodemographics and applicant education	
Age	Candidate's age (cont. variable)
Secondary or lower	Takes the value 1 for candidates with secondary or lower education.
Some college or higher	Takes the value 1 for candidates with university education.
Other type of education	Takes the value 1 for candidates with tertiary, non-university education, such as school teachers.
Attended public education inst.	Takes the value 1 for candidates who attended a public education institution.
STEM track	Takes the value 1 for candidates who have attended scientific or technological education.
Plans to study something else	Takes the value 1 for candidates who intend to pursue further education, including university, secondary, or other courses.
Prior knowledge of Coding	Takes the value 1 for candidates with prior coding knowledge.
High English level	Takes the value 1 for candidates with intermediate-advanced English proficiency.
Household and sociodemographic characteristics	
Low SES	Takes the value 1 for candidates from low-income families.
Residing in the capital city	Takes the value 1 for candidates living in the capital city.
Household size	Number of members in the household (cont. variable)
Head of household	Takes the value 1 for candidates who are the head of the household.
Has children	Takes the value 1 for candidates with children.
Parent with tertiary education	Takes the value 1 for candidates whose main reference (either father or mother) has tertiary or university education
More than 50 books at home	Takes the value 1 candidates with more than 50 books in their home.
Owns computer	Takes the value 1 for candidates with a personal or desktop computer at home.
Access to Internet	Takes the value 1 for candidates with a Wi-Fi Internet connection and 0 for those with only mobile phone Internet.
Not working and looking for a job	Takes the value 1 for candidates who are unemployed and actively seeking employment.
Has a private health insurance	Takes the value 1 for candidates with private health insurance.

Table A2: Heterogeneity by education and socio-economic status

	Exam performance					Continuation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total	Verbal	Math	Concentration	Logic	Admitted	Graduated
Panel A: Educational level							
Nudge	0.142**	0.152**	0.136**	0.147***	0.090	0.040	-0.003
	(0.058)	(0.065)	(0.061)	(0.057)	(0.057)	(0.030)	(0.019)
Male	0.307***	0.177***	0.318***	0.262***	0.319***	0.123***	0.075***
	(0.058)	(0.063)	(0.060)	(0.061)	(0.061)	(0.031)	(0.024)
Nudge × Male	-0.097	-0.034	-0.097	-0.088	-0.122*	-0.020	-0.034
	(0.072)	(0.079)	(0.074)	(0.075)	(0.073)	(0.038)	(0.029)
Some college or higher	0.448***	0.384***	0.413***	0.447***	0.391***	0.181***	0.066**
	(0.062)	(0.063)	(0.065)	(0.066)	(0.066)	(0.033)	(0.026)
Nudge × Some college or higher	-0.124	-0.082	-0.085	-0.195**	-0.123	-0.045	-0.009
	(0.078)	(0.080)	(0.081)	(0.083)	(0.081)	(0.040)	(0.032)
Male × Some college or higher	-0.211***	-0.218***	-0.157**	-0.198**	-0.210**	-0.081**	-0.052
	(0.077)	(0.079)	(0.079)	(0.087)	(0.085)	(0.041)	(0.039)
Nudge × Male × Some college or higher	0.091	0.046	0.018	0.158	0.148	0.003	0.030
	(0.099)	(0.102)	(0.101)	(0.112)	(0.105)	(0.052)	(0.049)
Nudge+Stress	0.294***	0.329***	0.324***	0.276***	0.138*	0.105**	0.024
	(0.081)	(0.083)	(0.085)	(0.083)	(0.082)	(0.044)	(0.030)
Nudge+Stress × Male	-0.260***	-0.298***	-0.323***	-0.094	-0.157	-0.155***	-0.087**
	(0.099)	(0.104)	(0.103)	(0.104)	(0.101)	(0.055)	(0.041)
Nudge+Stress × Some college or higher	-0.224**	-0.206*	-0.293***	-0.142	-0.124	-0.092	-0.070
	(0.106)	(0.107)	(0.110)	(0.116)	(0.111)	(0.057)	(0.046)
Nudge+Stress × Male × Some college or higher	0.196	0.295**	0.290**	0.019	0.031	0.143*	0.153**
	(0.138)	(0.141)	(0.141)	(0.155)	(0.147)	(0.073)	(0.069)
Obs.	4,658	4,658	4,658	4,658	4,658	4,658	4,658
Panel B: Low SES							
Nudge	0.039	0.091	0.043	-0.025	0.015	0.037	-0.045*
	(0.059)	(0.056)	(0.060)	(0.065)	(0.062)	(0.029)	(0.027)
Male	0.117**	0.027	0.129**	0.066	0.163***	0.065**	0.038
	(0.058)	(0.055)	(0.058)	(0.066)	(0.063)	(0.029)	(0.031)
Nudge × Male	-0.010	0.005	-0.024	0.092	-0.059	-0.042	0.017
	(0.073)	(0.072)	(0.074)	(0.083)	(0.078)	(0.036)	(0.037)
Low SES	-0.198***	-0.093	-0.215***	-0.205***	-0.189**	-0.079**	-0.028
	(0.071)	(0.072)	(0.073)	(0.075)	(0.075)	(0.037)	(0.030)
Nudge × Low SES	0.060	-0.004	0.075	0.118	0.037	-0.037	0.053
	(0.088)	(0.091)	(0.091)	(0.094)	(0.091)	(0.046)	(0.036)
Male × Low SES	0.148	0.076	0.166*	0.194**	0.107	0.043	0.006
	(0.090)	(0.093)	(0.092)	(0.099)	(0.096)	(0.048)	(0.043)
Nudge × Male × Low SES	-0.067	0.027	-0.075	-0.225*	-0.020	0.036	-0.064
	(0.114)	(0.119)	(0.118)	(0.126)	(0.119)	(0.061)	(0.053)
Nudge+Stress	0.084	0.162**	0.049	0.158*	-0.006	0.005	-0.046
	(0.080)	(0.079)	(0.081)	(0.091)	(0.084)	(0.041)	(0.038)
Nudge+Stress × Male	-0.086	-0.085	-0.055	-0.044	-0.122	-0.078	0.001
	(0.100)	(0.102)	(0.100)	(0.114)	(0.107)	(0.053)	(0.051)
Nudge+Stress × Low SES	0.078	0.051	0.121	0.011	0.057	0.055	0.018
	(0.121)	(0.124)	(0.128)	(0.130)	(0.125)	(0.065)	(0.050)
Nudge+Stress × Male × Low SES	-0.026	-0.060	-0.112	-0.030	0.122	0.008	-0.012
	(0.159)	(0.164)	(0.165)	(0.173)	(0.165)	(0.086)	(0.073)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	3,737	3,737	3,737	3,737	3,737	3,737	3,737

Notes: The table presents estimates for each outcome variable in the column headers following adding an additional interaction with the variables some college or higher and low SES in Equation 1. All columns control for baseline covariates. We report the coefficients relevant to see if there are any heterogeneity effects by these two variables.The sample size in the low SES regressions is smaller than in the main sample because there are missing values in this variable. Robust standard errors in parentheses. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table A3: Effects on performance (without controls)

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total score	Verbal	Math	Concentration	Logic
Nudge	0.082*	0.111**	0.095**	0.054	0.030
	(0.048)	(0.048)	(0.048)	(0.048)	(0.047)
Nudge+Stress	0.255***	0.301***	0.240***	0.271***	0.141**
	(0.061)	(0.062)	(0.061)	(0.065)	(0.062)
Male	0.380***	0.220***	0.411***	0.306***	0.385***
	(0.046)	(0.046)	(0.046)	(0.048)	(0.047)
Nudge × Male	−0.077	−0.029	−0.105*	−0.035	−0.078
	(0.058)	(0.059)	(0.058)	(0.062)	(0.059)
Nudge+Stress × Male	−0.267***	−0.267***	−0.272***	−0.171**	−0.223***
	(0.077)	(0.079)	(0.077)	(0.083)	(0.079)
Constant	−0.000	−0.000	0.000	−0.000	−0.000
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Pval Diff. Nudge-Stress	0.002	0.001	0.010	0.000	0.051
Mean dep.var (women)	0.57	12.63	11.89	4.38	7.63
SD dep.var (raw)	0.266	4.179	6.492	2.946	5.165
Questions	64	21	20	9	14
Controls	No	No	No	No	No
Obs.	4,658	4,658	4,658	4,658	4,658

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1. At the bottom of the table we report the point estimate and p-value of the treatment effect on women, along with the mean and SD for the outcome before standardization, and the total number of exam questions considered in each outcome. All standardized outcomes are standardized based on the mean and SD of men in the control group. Column 1 displays the estimates for the total score obtained in the entrance exam. Columns 2 to 5 presents the estimates for each exam subject. Verbal and math appeared after the stress reappraisal exercise, and concentration and logical reasoning appeared after the meditation exercise. Robust standard errors in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table A4: Effects on admission, exam completion and program continuation (without controls)

	Admitted	Exam completed		Continuation			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Above cutoff	None	Fraction	Enroll 1	Approved	Enroll 2	Graduated
Nudge	0.018	−0.003	0.024*	0.002	−0.019	−0.005	−0.011
	(0.023)	(0.007)	(0.014)	(0.023)	(0.020)	(0.020)	(0.016)
Nudge+Stress	0.088***	−0.024***	0.071***	0.075**	−0.002	0.008	−0.001
	(0.031)	(0.006)	(0.016)	(0.032)	(0.029)	(0.028)	(0.023)
Male	0.154***	−0.019***	0.103***	0.151***	0.106***	0.118***	0.094***
	(0.023)	(0.006)	(0.013)	(0.023)	(0.023)	(0.022)	(0.020)
Nudge × Male	−0.023	0.005	−0.025	−0.012	−0.017	−0.027	−0.026
	(0.029)	(0.008)	(0.016)	(0.030)	(0.029)	(0.028)	(0.024)
Nudge+Stress × Male	−0.132***	0.024***	−0.063***	−0.135***	−0.062	−0.068*	−0.039
	(0.040)	(0.007)	(0.020)	(0.041)	(0.040)	(0.039)	(0.034)
Constant	0.655***	0.024***	0.831***	0.629***	0.231***	0.208***	0.136***
	(0.018)	(0.006)	(0.011)	(0.018)	(0.016)	(0.015)	(0.013)
Pval Diff. Nudge-Stress	0.014	0.000	0.001	0.015	0.531	0.644	0.641
Mean control women	0.655	0.024	0.831	0.629	0.231	0.208	0.136
Controls	No	No	No	No	No	No	No
Obs.	4,658	4,658	4,658	4,658	4,658	4,658	4,658

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1. At the bottom of the table we report the point estimate and p-value of the treatment effect on women. Column 1 displays the estimates of the probability of program admission. Column 2 presents the estimates of the likelihood of answering zero questions. Column 3 reports the estimates of the fraction of the exam completed. Column 4 displays the estimates of taking longer than 180 minutes to complete the exam. Column 5 reports the estimates of the probability of enrollment in Phase 1 for those who are admitted. Column 6 reports the estimates of the likelihood of approving Phase 1. Column 7 reports the estimates of the probability of enrollment in Phase 2 for those who approved Phase 1. 792 students did not make the entrance exam cutoff. Robust standard errors in parentheses. $* \; p < 0.10$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table A5: Comparison of treated and control applicants' baseline covariates by gender

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	Control	Nudge and Stress.	Diff. (1)-(2)	Control	Nudge and Stress.	Diff. (4)-(5)
Sociodemographics and applicant education						
Age	24.009	24.151	−0.143	23.560	23.421	0.138
	(3.355)	(3.347)	(0.155)	(3.489)	(3.495)	(0.149)
Secondary or lower	0.518	0.513	0.005	0.599	0.622	−0.023
	(0.500)	(0.500)	(0.023)	(0.490)	(0.485)	(0.021)
Some college or higher	0.344	0.352	−0.009	0.278	0.261	0.017
	(0.475)	(0.478)	(0.022)	(0.448)	(0.439)	(0.019)
Other type of education	0.139	0.135	0.004	0.123	0.117	0.006
	(0.346)	(0.342)	(0.016)	(0.328)	(0.321)	(0.014)
Attended public education inst.	0.902	0.924	−0.022*	0.867	0.894	−0.027*
	(0.298)	(0.265)	(0.013)	(0.340)	(0.308)	(0.014)
STEM track	0.126	0.119	0.007	0.289	0.261	0.028
	(0.332)	(0.323)	(0.015)	(0.454)	(0.439)	(0.019)
Plan to study something else	0.781	0.784	−0.003	0.762	0.753	0.009
	(0.414)	(0.411)	(0.020)	(0.426)	(0.431)	(0.019)
Prior knowledge of coding	0.129	0.118	0.011	0.282	0.265	0.017
	(0.335)	(0.322)	(0.016)	(0.450)	(0.441)	(0.019)
High English level	0.478	0.485	−0.007	0.590	0.548	0.042**
	(0.500)	(0.500)	(0.023)	(0.492)	(0.498)	(0.021)
Household and Socioedemographic characteristics						
Low SES	0.471	0.464	0.007	0.376	0.381	−0.006
	(0.500)	(0.499)	(0.026)	(0.485)	(0.486)	(0.023)
Residing in capital city	0.496	0.540	−0.044*	0.548	0.535	0.013
	(0.500)	(0.499)	(0.023)	(0.498)	(0.499)	(0.021)
Household size	3.082	3.015	0.067	3.163	3.058	0.105
	(1.742)	(1.822)	(0.082)	(2.399)	(1.647)	(0.093)
Head of household	0.226	0.258	−0.031	0.287	0.291	−0.003
	(0.419)	(0.438)	(0.020)	(0.453)	(0.454)	(0.020)
Has children	0.216	0.199	0.016	0.071	0.084	−0.013
	(0.412)	(0.400)	(0.019)	(0.257)	(0.277)	(0.011)
Parent with tertiary education	0.295	0.295	−0.000	0.368	0.336	0.032
	(0.456)	(0.456)	(0.021)	(0.483)	(0.473)	(0.021)
More than 50 books at home	0.299	0.284	0.015	0.270	0.247	0.023
	(0.458)	(0.451)	(0.021)	(0.444)	(0.431)	(0.019)
Owns computer	0.874	0.862	0.012	0.942	0.942	−0.000
	(0.332)	(0.345)	(0.016)	(0.235)	(0.235)	(0.010)
Access to internet	0.817	0.837	−0.020	0.904	0.902	0.002
	(0.387)	(0.370)	(0.018)	(0.295)	(0.297)	(0.013)
Not working and looking for a job	0.487	0.458	0.029	0.436	0.419	0.017
	(0.500)	(0.498)	(0.023)	(0.496)	(0.494)	(0.021)
Has private health insurance	0.612	0.635	−0.023	0.655	0.656	−0.002
	(0.488)	(0.482)	(0.023)	(0.476)	(0.475)	(0.021)
Obs.	707	1,408	2,115	823	1,720	2,543

Notes: The table shows baseline covariate means for men and women by treatment assignment. Columns 3 and 6 compute the within-gender difference in means for men and women, respectively. Variable definitions are in Table A1. Standard deviations below the means and standard errors below the differences in parentheses.

Table A6: Effects on performance (reweighing using IPW)

	Performance				
	(1)	(2)	(3)	(4)	(5)
	Total score	Verbal	Math	Concentration	Logic
Nudge	0.065	0.092**	0.077*	0.043	0.019
	(0.042)	(0.044)	(0.043)	(0.044)	(0.044)
Male	0.191***	0.067	0.224***	0.159***	0.204***
	(0.042)	(0.044)	(0.042)	(0.046)	(0.046)
Nudge × Male	−0.015	0.022	−0.043	0.005	−0.015
	(0.054)	(0.057)	(0.055)	(0.059)	(0.057)
Nudge+Stress	0.183***	0.241***	0.162***	0.217***	0.080
	(0.055)	(0.054)	(0.056)	(0.063)	(0.061)
Nudge+Stress × Male	−0.151**	−0.163**	−0.148**	−0.098	−0.123
	(0.072)	(0.074)	(0.074)	(0.082)	(0.079)
Constant	0.110**	0.094**	0.095**	0.112**	0.104**
	(0.045)	(0.047)	(0.047)	(0.051)	(0.050)
Pval Diff. Nudge-Stress	0.023	0.003	0.110	0.004	0.276
Mean dep.var (women)	0.57	12.63	11.89	4.38	7.64
SD dep.var (raw)	0.266	4.177	6.488	2.945	5.165
Questions	64	21	20	9	14
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	4,658	4,658	4,658	4,658	4,658

Notes: This table presents the main estimates after reweighing with IPW. To get the weights, we first estimate a logit model to predict which characteristics are more predictive of "being a woman" and calculate the propensity score. Then we check whether there is enough overlap in the distribution of the propensity scores by gender. Finally, we weight the observations by $\frac{1}{pscore}$ for women and $\frac{1}{1-pscore}$ for men. Robust standard errors.$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A7: Effects on admission, exam completion and program continuation (reweighing using IPW)

	Admitted	Exam completed			Continuation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Above cutoff	None	Fraction	Overtime	Enroll 1	Passed	Enroll 2
Nudge	0.025	0.001	0.017	−0.004	0.004	−0.015	0.005
	(0.022)	(0.007)	(0.012)	(0.010)	(0.023)	(0.024)	(0.023)
Male	0.092***	−0.009	0.056***	−0.002	0.086***	0.049**	0.068***
	(0.022)	(0.006)	(0.012)	(0.010)	(0.023)	(0.025)	(0.025)
Nudge × Male	−0.024	−0.004	−0.006	0.006	−0.004	−0.005	−0.023
	(0.029)	(0.008)	(0.016)	(0.013)	(0.030)	(0.031)	(0.031)
Nudge+Stress	0.065**	−0.016***	0.043***	0.003	0.042	−0.013	0.002
	(0.030)	(0.005)	(0.014)	(0.014)	(0.032)	(0.034)	(0.033)
Nudge+Stress × Male	−0.094**	0.012*	−0.023	0.022	−0.089**	−0.032	−0.048
	(0.040)	(0.007)	(0.019)	(0.019)	(0.043)	(0.044)	(0.043)
Constant	0.684***	0.022***	0.857***	0.033**	0.666***	0.226***	0.211***
	(0.025)	(0.007)	(0.012)	(0.014)	(0.026)	(0.029)	(0.029)
Pval Diff. Nudge-Stress	0.159	0.000	0.047	0.593	0.207	0.956	0.932
Mean control women	0.655	0.024	0.831	0.028	0.629	0.231	0.208
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	4,658	4,658	4,658	4,658	4,658	4,658	4,658

Notes: This table presents the main estimates after reweighing with IPW. To get the weights, we first estimate a logit model to predict which characteristics are more predictive of “being a woman” and calculate the propensity score. Then we check whether there is enough overlap in the distribution of the propensity scores by gender. Finally, we weight the observations by $\frac{1}{pscore}$ for women and $\frac{1}{1-pscore}$ for men. Robust standard errors.$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A8: Gender differences in covariates using IPW weights

	Men	Women	Diff. (2)-(1)
	(1)	(2)	(3)
Sociodemographics and applicant education			
Age	23.761	23.727	-0.035
	(3.372)	(3.472)	
Secondary or lower	0.560	0.555	-0.005
	(0.497)	(0.497)	
Some college or higher	0.314	0.321	0.007
	(0.464)	(0.467)	
Other type of education	0.127	0.124	-0.003
	(0.333)	(0.329)	
Attended public education inst.	0.902	0.900	-0.002
	(0.297)	(0.300)	
STEM track	0.200	0.206	0.006
	(0.400)	(0.404)	
Plan to study something else	0.770	0.779	0.009
	(0.421)	(0.415)	
Prior knowledge of coding	0.205	0.205	0.000
	(0.404)	(0.404)	
High English level	0.531	0.535	0.004
	(0.499)	(0.499)	
Household and Sociodemographic characteristics			
Low SES	0.422	0.415	-0.006
	(0.494)	(0.493)	
Residing in capital city	0.546	0.546	-0.001
	(0.498)	(0.498)	
Household size	3.092	3.062	-0.029
	(2.245)	(1.742)	
Head of household	0.280	0.278	-0.003
	(0.449)	(0.448)	
Has children	0.134	0.126	-0.008
	(0.340)	(0.332)	
Parent with tertiary education	0.337	0.335	-0.001
	(0.473)	(0.472)	
More than 50 books at home	0.277	0.277	0.000
	(0.448)	(0.448)	
Owns computer	0.906	0.917	0.011
	(0.292)	(0.276)	
Not working and looking for a job	0.434	0.437	0.003
	(0.496)	(0.496)	
Has private health insurance	0.650	0.653	0.003
	(0.477)	(0.476)	

Notes: This table replicates columns 4 to 6 of Table 1 after applying the weights using IPW. There are no significant gender differences in baseline covariates after applying these weights. Standard deviations below the means and standard errors below the differences in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table A9: Difficulty of exam including stress management exercises (version 4) relative to other exam versions

	Correct answers (2021)		Correct answers (2022)	
	Full sample	Women	Full sample	Women
Test 1	−0.004	−0.019	0.012	−0.014
	(0.013)	(0.017)	(0.013)	(0.016)
Test 2	−0.013	−0.009	−0.013	−0.033**
	(0.013)	(0.017)	(0.014)	(0.017)
Test 3	−0.019	−0.042**	−0.003	−0.012
	(0.013)	(0.018)	(0.013)	(0.016)
Test 5	−0.026**	−0.034**	−0.027**	−0.062***
	(0.013)	(0.017)	(0.013)	(0.015)
Test 6	−0.017	−0.019	−0.006	−0.029*
	(0.013)	(0.017)	(0.014)	(0.016)
Test 7	−0.009	−0.014	−0.005	−0.023
	(0.013)	(0.017)	(0.013)	(0.016)
Women		−0.109***		−0.134***
		(0.018)		(0.019)
Test 1=1 × Women		0.032		0.060**
		(0.026)		(0.026)
Test 2=1 × Women		0.002		0.046*
		(0.026)		(0.027)
Test 3=1 × Women		0.049*		0.014
		(0.026)		(0.027)
Test 5=1 × Women		0.022		0.076***
		(0.025)		(0.026)
Test 6=1 × Women		−0.006		0.048*
		(0.026)		(0.027)
Test 7=1 × Women		0.013		0.040
		(0.026)		(0.027)
Obs.	345,920	345,920	333,696	333,696

Notes: The outcome in this table is the likelihood of answering a question correctly and the data is at the student-question level. The purpose is to compare the difficulty of exam version 4, which contained the stress management exercises, with the remaining exam versions, which did not contain any stress-related exercises. The versions that we use in the final sample are 1, 4, 6 and 7 (see Appendix B for details). $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table A10: Statistics by exam version

	Test 1	**Test 2**	**Test 3**	**Test 4**	Test 5	**Test 6**	**Test 7**
Female	0.54	0.52	0.55	0.56	0.53	0.55	0.55
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Tertiary education	0.44	0.43	0.45	0.43	0.43	0.41	0.42
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.49)	(0.49)
Low SES	0.43	0.41	0.43	0.42	0.43	0.42	0.40
	(0.50)	(0.49)	(0.50)	(0.49)	(0.50)	(0.49)	(0.49)
Obs.	830	783	748	711	800	763	824

Notes: This table shows a set of descriptive statistics by exam version. Each column indicates the test version, starting from Test 1 to Test 7. In each row, we present the mean and standard deviations in parenthesis for 3 variables. Row 1, shows the distribution of women by version. Row 2, shows the distribution by education. Row 3, shows the distribution by household income.

Table A11: Covariate balance by treatment- 2024

	(1) Control	(2) T1	(3) Diff. (C)-(T1)	(4) T2	(5) Diff. (C)-(T2)
Sociodemographics and applicant education					
Age	24.324	24.098	0.226	24.335	−0.011
	(3.300)	(3.398)	(0.165)	(3.397)	(0.163)
Secondary or lower	0.449	0.436	0.013	0.433	0.016
	(0.498)	(0.496)	(0.024)	(0.496)	(0.024)
Some college or higher	0.409	0.396	0.013	0.409	0.000
	(0.492)	(0.489)	(0.024)	(0.492)	(0.024)
Other type of education	0.142	0.169	−0.026	0.158	−0.016
	(0.349)	(0.375)	(0.018)	(0.365)	(0.017)
Attended public education inst.	0.906	0.900	0.006	0.913	−0.007
	(0.292)	(0.301)	(0.014)	(0.282)	(0.014)
STEM track	0.156	0.164	−0.008	0.182	−0.027
	(0.363)	(0.370)	(0.018)	(0.386)	(0.018)
Plan to study something else	0.801	0.809	−0.008	0.810	−0.009
	(0.400)	(0.393)	(0.021)	(0.393)	(0.021)
Prior knowledge of coding	0.187	0.187	0.000	0.176	0.011
	(0.390)	(0.390)	(0.019)	(0.381)	(0.019)
High English level	0.497	0.525	−0.028	0.494	0.002
	(0.500)	(0.500)	(0.025)	(0.500)	(0.024)
Household and Socioedemographic characteristics					
Low SES	0.434	0.392	0.042	0.419	0.015
	(0.496)	(0.489)	(0.027)	(0.494)	(0.027)
Residing in capital city	0.557	0.597	−0.040*	0.545	0.012
	(0.497)	(0.491)	(0.024)	(0.498)	(0.024)
Household size	2.975	3.230	−0.255	3.084	−0.109
	(1.777)	(4.052)	(0.155)	(2.290)	(0.100)
Head of household	0.242	0.244	−0.002	0.231	0.011
	(0.429)	(0.430)	(0.021)	(0.422)	(0.021)
Has children	0.122	0.108	0.015	0.124	−0.001
	(0.328)	(0.310)	(0.016)	(0.330)	(0.016)
Parent with tertiary education	0.318	0.330	−0.012	0.343	−0.025
	(0.466)	(0.471)	(0.023)	(0.475)	(0.023)
More than 50 books at home	0.269	0.275	−0.006	0.265	0.004
	(0.444)	(0.447)	(0.022)	(0.442)	(0.022)
Owns computer	0.914	0.910	0.004	0.909	0.005
	(0.281)	(0.287)	(0.014)	(0.288)	(0.014)
Access to internet	0.827	0.870	−0.043**	0.850	−0.023
	(0.378)	(0.337)	(0.017)	(0.357)	(0.018)
Not working and looking for a job	0.502	0.453	0.049**	0.469	0.033
	(0.500)	(0.498)	(0.025)	(0.499)	(0.024)
Has private health insurance	0.647	0.690	−0.043*	0.653	−0.006
	(0.478)	(0.463)	(0.023)	(0.476)	(0.023)
Obs.	887	801	1,688	822	1,709

Notes: Columns 1 and 2 show baseline covariate means by control and *T1*, respectively. Column 3 computes the difference between columns 1 and 2 and shows whether the difference is statistically significant. Columns 4 shows the baseline covariate means for *T2*. Column 5 tests whether the differences are significant. Standard deviations below the means and standard errors below the differences in parentheses. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$. We run two separate regressions: the first regressing the full set of covariates on *T1*, and the second on *T2*. In the first regression, we fail to reject the null hypothesis that all covariates are jointly equal to zero (F-statistic = 1.25, p-value =0.154). In the second regression, we also fail to reject the null hypothesis (F-statistic = 0.88, p-value = 0.664).

Table A12: Effects on performance, exam completion and admission - 2024

	Exam completed		Admitted	Performance				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Fraction	None	Above cutoff	Total score	Verbal	Math	Concent.	Logic
Nudge+Stress(verbal)	0.019*	0.001	−0.003	0.070	0.029	0.060	0.084*	0.079
	(0.011)	(0.001)	(0.022)	(0.048)	(0.047)	(0.048)	(0.048)	(0.048)
Nudge+Stress(math)	−0.011	0.064***	0.015	−0.008	−0.365***	0.182***	0.037	0.012
	(0.013)	(0.009)	(0.022)	(0.051)	(0.061)	(0.046)	(0.048)	(0.048)
Constant	0.886***	0.006	0.726***	−0.023	0.037	−0.040	−0.059	−0.018
	(0.012)	(0.005)	(0.021)	(0.048)	(0.052)	(0.047)	(0.046)	(0.047)
Mean (control)	0.88	0.00	0.72	40.42	14.56	13.15	4.68	8.03
Diff.	0.03	-0.06	-0.02	0.08	0.39	-0.12	0.05	0.07
p-value	0.02	0.00	0.41	0.12	0.00	0.01	0.33	0.17
Exam version FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Questions				64	21	20	9	14
Obs.	2,510	2,510	2,510	2,510	2,510	2,510	2,510	2,510

Notes: The *Nudge + Stress* intervention is the same one describe in the main text. The difference between the two rows in the table is what exam subject appears first after the stress reappraisal exercise. In the first row, verbal appears first, which is exactly the same as in the intervention in the main text. In the second row math appears first and verbal second. The control group answers verbal questions first. Robust standard errors.$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A13: How stress was perceived to affect performance - 2024

	(1) Reduced	(2) Enhanced	(3) Did not affect	(4) No stress
Nudge + Stress (verbal)	-0.055^{**}	0.100^{***}	-0.032	-0.013
	(0.026)	(0.022)	(0.024)	(0.014)
Nudge + Stress (math)	-0.079^{***}	0.083^{***}	-0.013	0.009
	(0.026)	(0.021)	(0.024)	(0.015)
Constant	0.410^{***}	0.181^{***}	0.333^{***}	0.076^{***}
	(0.025)	(0.021)	(0.024)	(0.014)
Diff. math-verbal	0.02	0.02	-0.02	-0.02
p-value	0.36	0.47	0.44	0.13
Controls	No	No	No	No
Obs.	2,146	2,146	2,146	2,146

Notes: The outcomes in each column are constructed from the survey question: *If at any point you felt stressed before or during the test, do you think:* The stress reduced your performance, The stress enhanced your performance, The stress did not affect your performance, I did not feel stressed before or during the test. Robust standard errors.$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.