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

Catalina Franco† Marcela Gomez-Ruiz‡

December 5, 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, Bertil Tungodden, 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, the FAIR CELE coffee meeting, and the 2025 AEA/ASSA annual 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 explanation for gender differences in exam performance is that pressure may negatively impact women more than men. For instance, Cai et al. (2019) found a significant performance drop for women—but not men—from a mock exam to the high-stakes *Gaokao* in China. Another factor is women's higher tendency to leave questions unanswered (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; Balart et al., 2022; Karle et al., 2022), linked to greater risk aversion (when incorrect answers are penalized) and lower self-confidence. 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 trial conducted by a public agency in Uruguay for admission to its Coding Program. Applicants were randomly assigned to an exam version covering verbal, mathematics, concentration, and logic. The versions were equivalent in difficulty, lacked penalties for wrong answers, and were classified into one of three groups: *Control* (two standard versions), *Nudge* (three versions encouraging attempting all questions and clarifying no penalties for incorrect responses), and *Nudge+Stress* (one version adding a stress-reframing intervention and a brief meditation).[1]

Both the *Nudge* and *Nudge+Stress* treatments significantly improve women's performance by reducing omitted questions and increasing correct answers. The *Nudge+Stress* intervention is particularly effective, narrowing the gender gap in Coding Program admissions by helping more women surpass the admission threshold. Our main contribution is to show that simple nudges and stress-reframing interventions significantly reduce omitted questions, which in turn have a substantial impact on performance. By highlighting omitted questions as a key, malleable driver of gender performance gaps, our study provides actionable insights into addressing these disparities

[1]The interventions, designed by the Behavioral Science Lab at Ceibal, drew on Harris et al. (2019) to increase female admission rates.

through targeted, low-cost strategies.

Our three main findings are as follows. First, compared to the *Control* group, women's overall scores increased by 0.08 standard deviations (SD) in *Nudge* and 0.18 SD in *Nudge+Stress*, with no effects for men. Women's gains came from improved performance in verbal (0.11 SD) and mathematics (0.09 SD) under *Nudge*, and in verbal (0.22 SD), mathematics (0.18 SD), and concentration (0.21 SD) under *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. Additionally, the 30-second meditation break in *Nudge+Stress* appears to boost performance in the subsequent concentration section for both genders.

Second, the treatments reduce omitted questions for women, with no effect on men. The effect is stronger in *Nudge+Stress*, where omissions decrease across all subjects, compared to *Nudge*, which impacts only verbal and logic. Importantly, accuracy (i.e., correct/attempted) increases slightly or remains unchanged, showing that attempting more questions does not compromise correctness. 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*.[2] 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 and graduation rates 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 seems to stem from increased effort, reflected in a higher rate of attempted questions and, to a lesser extent, improved accuracy. The greater impact on women likely arises because men were already attempting most questions, leaving less room for

[2]There is a small and non-significant reduction in admissions for men. Unlike our case, in settings where the number of slots is fixed, a higher admissions for women would necessarily imply lower admissions for men.

improvement.[3] One explanation for the *Nudge* intervention's success is that men and women have different thresholds for answering questions they are not sure about, with the intervention lowering this threshold for women, whose threshols is presumably higher. The *Nudge+Stress* intervention may also boost confidence, especially for students prone to second-guessing.[4] Psychological theory suggests stress impacts performance by diverting cognitive resources to worry, reducing problem-solving capacity (Beilock, 2011). By alleviating this cognitive load, the interventions likely enhance performance.

We cannot directly validate the proposed theories, so we assess alternative mechanisms. First, covariate differences between men and women are controlled for in our regressions, addressing baseline differences as potential confounders. Second, both genders engage similarly with the intervention, as shown by comparable word counts in the stress reappraisal exercise and exam duration. Third, women's improved performance may stem from being in a "friendly environment" rather than stress reframing, with the meditation prompt potentially serving more as a break. While we cannot definitively rule out this alternative, the positive effects of the *Nudge+Stress* intervention are valuable regardless of the mechanism. Our findings suggest that less threatening testing environments—whether by reducing stress or offering simple accommodations—can significantly boost women's performance.

We contribute to the literature on the gender gap in academic performance by highlighting nudges and stress management as low-cost tools to improve effort and exam outcomes. 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 to the literature exploring the link between a lower willingness to guess in exams and overall performance, especially in scenarios where penalties for incorrect answers are applied (Pekkarinen, 2015; Funk and Perrone, 2016; Iriberri and Rey-Biel, 2021; Akyol et al., 2022). While the standard policy recommendation has been to remove such penalties, gender gaps persist even in their absence. Our findings introduce a novel and policy-actionable approach to addressing this issue. Taken together, our results suggest that nudges and short stress reframing 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.

[3] Our null result for men aligns with De Paola and Gioia (2016) on time pressure and Cavatorta et al. (2021) on exam anxiety.

[4] OECD (2015) highlights that girls often report greater anxiety and fear of mistakes, particularly in mathematics, despite high intrinsic motivation.

2 Setting and Research Design

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

Uruguay was a global pioneer in implementing the *One Laptop Per Child* initiative through Ceibal, a public agency providing connectivity and educational content to all public school students. However, the country continues to face significant educational challenges, including some of the lowest secondary and tertiary graduation rates in the region (USAID, 2022).

Short-term programs offering marketable skills can attract students seeking to enhance their skill set, especially in the face of high dropout and low graduation rates. The Coding Program addresses this need by providing free training in a high-demand field for individuals aged 18 to 30 who have completed at least the third grade of high school.[5] The program is also appealing to youths still within the formal education system due to its strong ties with private sector employers. Demand for the program has grown steadily; since 2017, it has trained over 4,100 youths, and in 2023, more than 9,000 expressed interest.

2.2 The Coding Program Design and Entrance Exam

The program trains participants in a coding language, English, and soft skills for the labor market, such as CV and interview preparation. It consists of three phases over one year. In phase 1 (March–June), students complete basic courses asynchronously online. In phase 2 (July–December), those who pass phase 1 through teacher assessments and homework gain advanced coding knowledge (e.g., testing, web development) in small synchronous classes. Phase 3 provides work placement assistance in the tech sector.

Applicants must meet eligibility criteria, including completing at least the third grade of high school, being at least 18 years old, and scoring at least 50% on a 64-question entrance exam administered online in January.[6] The exam covers verbal (21 questions), math (20), concentration (9), and logic (14) in that order.[7] Correct answers are worth one point, with no penalties for incorrect responses. To minimize cheating, applicants are randomly assigned one of seven exam versions of similar difficulty.[8] The exam is computer-graded, eliminating score manipulation. While the exam

[5] In Uruguay, high school comprises six years, with third grade typically completed at age 15.

[6] The cutoff is not disclosed; applicants are encouraged to perform their best.

[7] Concentration tasks involve tedious, effort-based activities like counting occurrences of letters or numbers, requiring no prior knowledge.

[8] Some questions may repeat across versions or are of equivalent difficulty. No other measures to prevent cheating are known.

has a 180-minute limit, some applicants exceed this and are automatically disqualified.[9]

2.3 Data, Randomization and Analytical Sample

The data consist of administrative records from the 2023 admissions, including enrollment in phase 1 and continuation to phase 2. The dataset includes exam-related information: baseline characteristics from registration, overall and subject-specific performance, admission decisions, total exam time, and, for *Nudge+Stress* applicants, self-reported stress and stress reappraisal responses. It also contains response-level data for all 64 questions, indicating whether they were answered and correct. Post-exam data include teachers' phase 1 performance assessments, students' decisions to continue to phase 2, and graduation.

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 presents the balance across 21 baseline covariates between control and treated applicants (*Nudge* and *Nudge+Stress* together) in columns 1-3, and gender differences in columns 4-6. Covariate means across treatment and control are not statistically different.[12] The gender differences in the sample are substantial and interesting. Women constitute about 55% 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). Women are more likely than men to come from disadvantaged backgrounds (e.g., low SES households, parents without tertiary education) and less likely to own a computer, have prior STEM exposure, or coding knowledge.

[9]The disqualification rule is not clearly communicated, but less than 3% exceed the time limit. The program website states the exam takes 2.5 hours.

[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 A).

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

[12]Missing values are imputed for regressions to retain sample size, with indicators for missingness added. Using the full covariate set, only 1 of 21 variables is unbalanced (F-statistic of the joint balance test = 1.14, p-value = 0.267).

However, women may be positively selected on unobservable factors, as they are more likely to pursue or hold 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.

Since randomization occurred across exam versions rather than within, we assess whether treatment and control versions are equivalent. Differences in version difficulty could confound treatment effects if treatments were assigned to easier versions. Using identical exam versions from 2021 and 2022, we regress the likelihood of obtaining a correct answer by version (see Table A2). Point estimates are small and insignificant, except for version 5, which appears harder and is excluded from analysis. Descriptive evidence also shows no differential selection into versions by gender, education, or socioeconomic background (see Table A3). Overall, we find no evidence that control and treatment versions differ (see Appendix A).

3 Intervention Details and Empirical Strategy

The *Nudge* and *Nudge+Stress* treatments address constraints that may disproportionately affect women's performance. These interventions target factors unrelated to academic ability (e.g., Duquennois, 2022; Franco and Povea, 2023).[13] Low self-confidence may prevent applicants from answering questions they are unsure of, a barrier the *Nudge* treatment seeks to reduce by encouraging applicants to "complete as many questions as they can." Stress, which can deplete working memory through intrusive thoughts about failure (Beilock, 2011), may further impair performance. The *Nudge+Stress* intervention addresses this by reframing stress as performance-enhancing in addition to the nudge to attempt all questions.

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.

[13] These studies examine "external" constraints affecting performance, such as exam design features. The *Nudge* and *Nudge+Stress* treatments focus on "internal" constraints, such as how exam takers feel during the test.

3.2 *Nudge+Stress* Treatment

Applicants assigned to this treatment (version 4) follow these steps: First, they answer a question about their current anxiety level. Second, they read a paragraph about stress responses and how to interpret them. Third, they respond to a reflection question on the paragraph. Fourth, they read the three sentences from the *Nudge* treatment. Fifth, they begin the exam. Finally, after the verbal and math sections, they are reminded of the stress text and prompted to practice one of three anxiety-reducing techniques for 30 seconds.

We provide a summary of the main points from the stress treatment below, with full English translations available in Appendix B. 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. The emotional response we have depends largely on how we choose to interpret stress and arousal. If we interpret physiological arousal as negative, we experience emotions such as fear and threat. Instead, if we interpret it as positive, we experience emotions such as arousal and anticipation. People who respond well to stressful situations are those who interpret their body's arousal positively: they get excited because their body is ready for peak performance during a test or presentation.*"

The text then prompts exam takers 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 completing the verbal and math sections, applicants read: "*Remember that people experience a physiological stress response in many situations, such as taking a course exam. This response is necessary for increased alertness and responsiveness.*" They are then presented with three techniques to reduce anxiety: deep breathing, visualizing a calming place, or progressive muscle relaxation. The prompt instructs them to choose one technique and spend 30 seconds practicing it. Afterward, they proceed to the concentration and logic sections.[14]

[14]Returning to previous questions is technically possible but requires clicking back repeatedly, which may deter applicants.

3.3 Econometric Specification

Our econometric specification involves estimating the the treatment effects on outcomes such as exam performance, omitted questions and admissions for each gender separately:

$$y_i = \beta_0 + \beta_1 \text{Nudge}_i + \beta_2 \text{Nudge+Stress}_i + X_i\gamma + \varepsilon_i \tag{1}$$

The coefficient β_1 provides the effect of the nudge while β_2 provides the effect of the *Nudge+Stress* intervention. In the appendix we report estimates of the model including an indicator for male and its interactions with the treatments, that is, the difference-in-differences (DID) coefficients measuring the change in the gender gap generated by the treatments. 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 A13 and A10). 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.

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 that women's scores in the control group have a larger left tail and less density above the 50% cutoff compared to men. On average, women score 57%, while men score 20.8 percentage points higher (Table 2, Column 1).

4.2 Effects on Exam Performance

Figure 1, Panels (b) to (e) show raw performance by treatment for women and men. Panels (b) and (d) reveal fewer women with very low scores, particularly under *Nudge+Stress*, where the distribution shifts rightward, benefiting women across the lower tail of the score distribution, not just those near the cutoff. In contrast, Panels (c) and (e) show no effect on men's score distributions, as fewer men scored very low initially. However, *Nudge+Stress* appears to boost scores for men already performing at higher levels.

Table 2, Panel A shows effects on standardized scores for the overall exam and individual subjects. Women's overall scores increase by 0.08SD in *Nudge* and 0.18SD in *Nudge+Stress*, indicating the combined intervention's effect size is twice that of the nudge alone (p-value=0.037). Panel B shows near-zero point estimates for men, confirming that the observed gains are specific to women, as also reflected in the DID coefficient in Table A4.

By exam subject (Columns 2-5 of Table 2), the *Nudge* treatment significantly improves women's performance in verbal (0.11 SD) and math (0.09 SD) but not in concentration or logic. In *Nudge+Stress*, women show higher gains across verbal (0.22 SD), math (0.18 SD), and concentration (0.21 SD), while the effect on logic is smaller and insignificant. For men (Panel B), significant effects are seen only in verbal (*Nudge*) and concentration (*Nudge+Stress*). Table A4 highlights large gender gaps in the control group, with women scoring 0.2 SD lower than men in most subjects except verbal. The *Nudge+Stress* treatment substantially reduces these gaps, with DID coefficients significant and similar to treatment effects for women, except in concentration, where gains are comparable for both genders.

We offer several observations on our results. First, the small gender gap favoring men in verbal is reversed in both treatments, aligning with evidence that women often outperform men in verbal subjects (Coffman, 2014). Second, the interventions have the strongest effect on the first exam subject, immediately following the treatment. Third, effects diminish in the last subject, likely due to mental fatigue during the exam (Brown et al., 2022). Finally, effects reappear in the concentration section—the second-to-last subject—after the meditation prompt in *Nudge+Stress*, suggesting it either reinforces the stress reappraisal or acts as a beneficial break. Both interpretations offer valuable insights for exam design.

To contextualize our findings, we reference the broader education literature, as few interventions target students during exams. The effect size of the *Nudge+Stress* treatment on overall performance is comparable to interventions like incentivizing teachers to reduce absenteeism (Duflo et al., 2012) and substantially larger than costly programs like the One Laptop per Child initiative, which shows null effects across countries (Cristia et al., 2017; Falck et al., 2018; Yanguas, 2020).

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. 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 presents the results for omitted questions using the main specification. On average, women in the control group omit 10.8 questions (Panel A), while men omit only 4.2 questions (Panel B). The *Nudge* treatment reduces the total fraction of omitted questions among women by 2.5 percentage points, while the *Nudge+Stress* treatment reduces it by 5.5 pp, nearly closing the gender gap in omitted questions equal to 6.7 pp in the control group (Table A5).

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, compared to 0.05% of men. Control women complete 83.1% of the exam on average, with the *Nudge+Stress* treatment raising this by 5.5 pp. 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 1.5 of 21 verbal questions, 3.7 of 20 math questions, 2.1 of 9 concentration questions, and 3.5 of 14 logic questions. The equivalent numbers for men oscillate between 0.5 and 1.5 omitted questions. The effects of the *Nudge+Stress* treatment on the fraction omitted by women are -3.3 pp in verbal, -6.4 pp in math, -5.9 pp in concentration, and -7.3 pp in logic. With no effects for men, these effects reduce the gender gap in omitted questions in the control group by at least half (Table A5).

Turning to accuracy rates, attempting more questions may not raise scores if the newly attempted questions are incorrect. However, even with unchanged accuracy, attempting more questions increases the fraction correct and overall score. Table A6 shows that control women have an accuracy rate of 66.6% compared to 71% for men. The treatments have limited effects on overall accuracy (Column 1), with notable gains only in specific subjects. Concentration accuracy increases significantly by 4.3 pp for women and 5.2 pp for men in the *Nudge+Stress* treatment, suggesting the meditation had a positive impact. Overall, small gains in accuracy are observed, but this is not the primary margin of improvement from the intervention.

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

[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.

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.

Our findings highlight a key insight: encouraging women to omit fewer questions can significantly reduce gender gaps in exam performance. Helping women attempt more questions boosts overall scores, even if accuracy rates remain unchanged, suggesting they may be "leaving money on the table" by omitting questions despite no penalties for incorrect answers. This is particularly promising given the well-documented gender differences in omitting questions and willingness to guess, with few studies identifying effective ways to address these gaps (Iriberri and Rey-Biel, 2021).

4.4 Effects on Admissions, Program Continuation and Graduation

Table 4 shows results for program admission, progression, and graduation. In the control group, 81% of men and 65.5% of women are admitted. The *Nudge* treatment has no effect on women's admissions, but *Nudge+Stress* raises their admission rate by 5.9 pp, reducing the gender gap by nearly 70%. Effects for men are negligible and not significant. Overall, the intervention disproportionately benefits women, leading to a 9% increase in female admissions to the Coding Program.

The effects on progression 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 treatment effects for wither gender, except for a small but marginally insignificant effect on progression of men in the *Nudge+Stress* treatment. 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 final performance outcome we examine is graduation (Table 4, Column 7). Among 4,658 applicants, 13.6% of women and 23% of men graduate. The treatments have no effect on graduation, suggesting that women admitted due to the interventions are of similar quality to those in the control group. Higher graduation rates among men may reflect greater intrinsic interest in the program's topics (Table 1) or uncertain employment prospects for female programmers in a male-dominated field.

4.5 Heterogeneity Analysis

We conduct two heterogeneity analyses using education level and socio-economic status (SES). Three key findings emerge. First, the interventions benefit less-educated women most, with the *Nudge+Stress* treatment increasing admission rates for women with less than some college by 10.5 pp (Table A8, Panel A), while there are no effects for low-SES women (Panel B). Second, the interventions do not affect men at all (Table A9). Third, there are no effects for men or women of different SES levels.

5 Underlying mechanisms

This section explores theories explaining why the interventions improved women's performance and primarily benefited women over men. We also consider alternative stress-unrelated mechanisms and provide evidence on how a *Nudge+Stress*-like intervention influences stress perceptions during an exam.

The *Nudge* intervention encourages exam takers to attempt all items. One model that could explain why women are more likely than men to omit questions is differing thresholds for answering. For instance, if both men and women are 60% sure their answer is correct, but women require 70% certainty while men need only 50%, men would answer while women would not, even though both are equally confident and likely to be correct. By encouraging test takers to attempt all questions, the *Nudge* treatment may lower women's threshold for answering, bringing it closer to that of men.

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* group were 13-19% less likely to report that stress reduced performance and 46-56% more likely to say it enhanced performance.[16] Women may benefit more from stress-reframing interventions due to higher reported anxiety levels (Remes et al., 2016; OECD, 2015) and poorer performance in stress-induced competitions (Cahlíková et al., 2020). Similarly, Cavatorta et al. (2021) show that anxious women benefit from interventions reducing focus on negative stimuli, leading them to attempt more questions in cognitive tasks, a pattern consistent with our findings.

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

5.1 Gender Differences in Covariates

Table 1 shows significant observable differences between men and women, suggesting potential differences in unobservables as well. Gender differences in covariates indicate that observables play an important role. However, the trial remains internally valid within gender, as no systematic baseline differences are observed (Table A14).

We present the results separately by gender and including controls to address the potential issue raised by gender differences in covariates. In addition, we conduct robustness exercises in Tables A15 and A16, 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.[17] 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

If the intervention reduces stress for all and lower stress improves performance, differential effects for women may arise if they are more likely to engage with the intervention. For instance, Shreekumar and Vautrey (2022) report strong gender selection in their study, with men comprising only 15% of those interested in a meditation app. This may suggest that men are less interested in mindfulness techniques, doubt their efficacy, or simply see no benefit.

[16] Respondents may have echoed the stress reappraisal messaging rather than genuinely feeling improved performance; see Appendix C.

[17] 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 A17.

Using several proxies for engagement, we conclude that men took the stress reappraisal exercise as seriously as women. First, applicants may have believed the exercise was part of the exam due to the lack of instructions indicating otherwise. Second, both genders wrote an average of 35 words, with similar word count distributions (Figure A3). Third, while we cannot confirm if applicants followed meditation instructions, treated participants of both genders spent more time on the exam than controls (Figure A4), suggesting engagement. Finally, large and significant treatment effects for both genders immediately after the meditation prompt indicate that differences in engagement do not explain why the intervention primarily benefits women.

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. 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—such as those about women in mathematics—can impair performance by creating cognitive burdens and distractions (Steele and Aronson, 1995; Spencer et al., 1999). The *Nudge+Stress* intervention may counteract these effects by framing stress as normal and beneficial, reducing fear of failure and pressure to disprove stereotypes. Alternatively, it may have fostered a testing environment where women felt less constrained by external expectations, improving performance.

Evidence from other contexts supports this idea. Studies show women perform better in same-sex competitions, such as in the lab (Gneezy et al., 2003) and all-women speedboat races, while men excel in mixed-sex settings (Booth and Yamamura, 2018). Similarly, in the 2024 Coding Program admissions, where only women were invited to apply, the *Nudge+Stress* intervention had no significant effects. Women had similar exam completion and omission rates as men in 2023, suggesting their performance was unaffected by the intervention when men were absent.[18] 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 text.

Based on prior research and the 2024 exam data, we conclude that women's perception of the testing environment may significantly influence performance. While the *Nudge+Stress* intervention may

[18] Gomez-Ruiz et al. (2024) found women performed better on this entrance exam in 2019 without men.

affect stress responses, it is equally plausible that its impact stems from creating a more supportive, less intimidating environment. Regardless of the mechanism, framing the test as non-threatening may be key to reducing gender gaps in performance

6 Discussion and Conclusion

Our study suggests that exams measure not only academic ability but also factors like the impact of stress or other behavioral barriers on performance. Using data from an entrance exam in Uruguay, we evaluate two interventions: one encouraging applicants to answer all questions and another combining this nudge with stress reframing. The combined intervention significantly benefits women, reducing gender gaps in performance and admissions, with no effect on men.

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.

The implications of our findings extend beyond this specific trial, highlighting the importance of designing testing environments that address non-academic barriers and optimize conditions for all test-takers. Accommodations such as stress-reframing prompts, brief meditation exercises, or other targeted interventions can reduce cognitive and emotional distractions during exams. These practical and scalable measures offer a pathway to more equitable assessment practices without requiring major changes to exam structures or administration. Ultimately, 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

Akyol, P., J. Key, and K. Krishna (2022). Hit or miss? test taking behavior in multiple choice exams. *Annals of Economics and Statistics* (147), 3–50.

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.

Funk, P. and H. Perrone (2016). Gender differences in academic performance: The role of negative marking in multiple-choice exams.

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

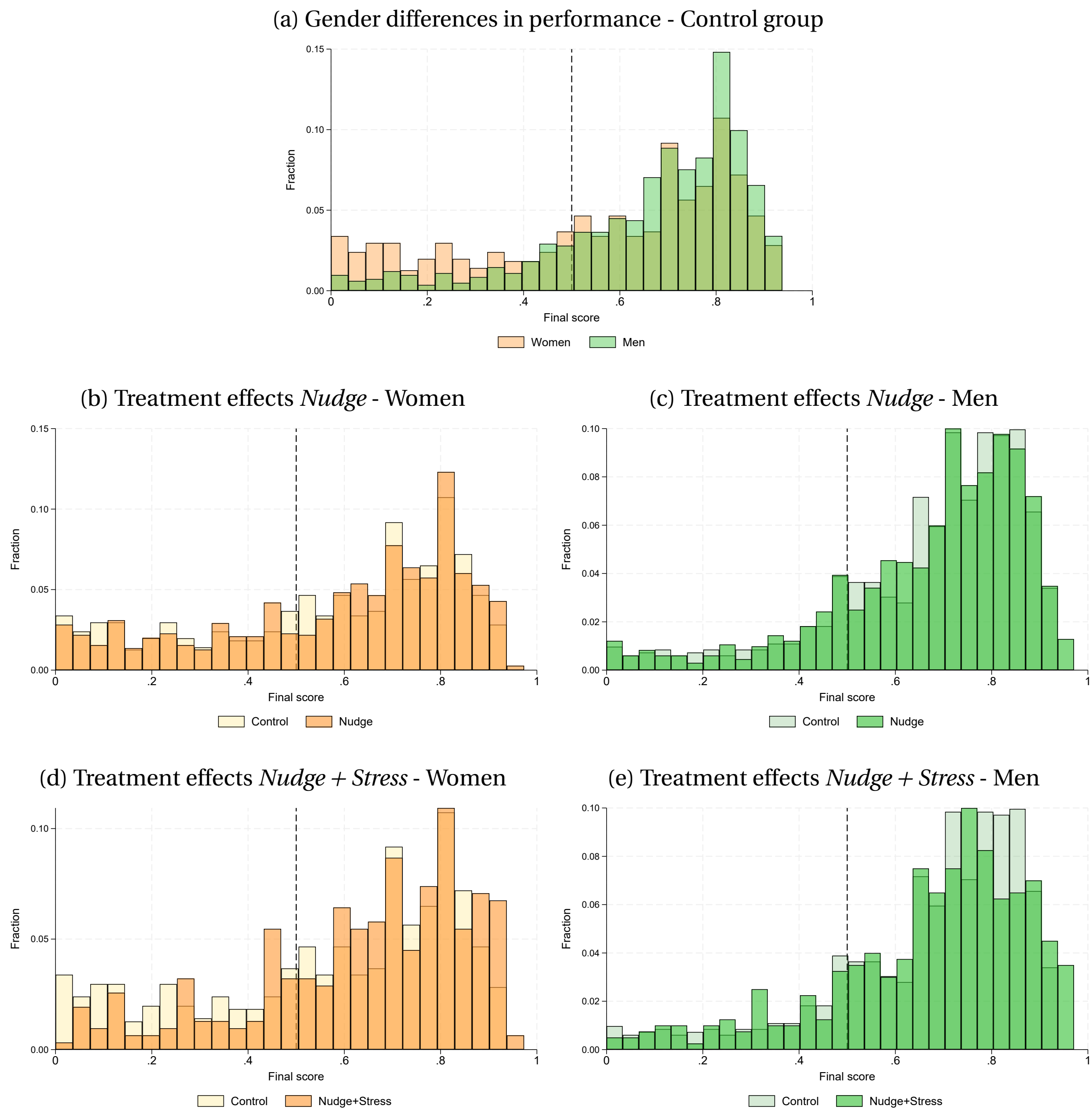


Figure 1: Gender differences in control group performance and treatment effects across the exam performance distribution by gender

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) Women	(5) Men	(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 (pooling the two treatments together), 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 Table A1. 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
Panel A: Women					
Nudge	0.078**	0.109***	0.091**	0.050	0.027
	(0.039)	(0.041)	(0.041)	(0.042)	(0.040)
Nudge+Stress	0.183***	0.223***	0.176***	0.208***	0.085
	(0.053)	(0.054)	(0.056)	(0.059)	(0.056)
Constant	0.162**	0.112*	0.156**	0.146**	0.163**
	(0.063)	(0.065)	(0.066)	(0.069)	(0.066)
Raw mean dep.var.	0.571	12.629	11.885	4.380	7.632
Pval Diff. Nudge-Stress	0.037	0.025	0.106	0.005	0.268
Obs.	2,115	2,115	2,115	2,115	2,115
Panel B: Men					
Nudge	0.030	0.106***	0.011	0.043	-0.024
	(0.031)	(0.032)	(0.031)	(0.037)	(0.034)
Nudge+Stress	0.018	0.063	-0.007	0.132***	-0.058
	(0.043)	(0.047)	(0.043)	(0.049)	(0.046)
Constant	0.295***	0.153***	0.308***	0.263***	0.313***
	(0.046)	(0.051)	(0.048)	(0.054)	(0.049)
Raw mean dep.var.	0.672	13.549	14.553	5.281	9.622
Pval Diff. Nudge-Stress	0.764	0.334	0.659	0.052	0.433
Questions	64	21	20	9	14
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	2,543	2,543	2,543	2,543	2,543

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 for the corresponding gender, along with the mean 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 women 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

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total omitted	Verbal	Math	Concent.	Logic
Panel A: Women					
Nudge	-0.025**	-0.019**	-0.025	-0.020	-0.037**
	(0.013)	(0.009)	(0.015)	(0.017)	(0.017)
Nudge+Stress	-0.055***	-0.033***	-0.064***	-0.059***	-0.073***
	(0.016)	(0.010)	(0.019)	(0.022)	(0.023)
Constant	0.141***	0.057***	0.142***	0.196***	0.232***
	(0.018)	(0.012)	(0.021)	(0.025)	(0.027)
Raw mean dep.var.	10.820	1.523	3.656	2.107	3.533
Pval Diff. Nudge-Stress	0.034	0.094	0.025	0.054	0.093
Obs.	2,115	2,115	2,115	2,115	2,115
Panel B: Men					
Nudge	-0.002	-0.001	0.001	0.008	-0.013
	(0.008)	(0.005)	(0.010)	(0.011)	(0.012)
Nudge+Stress	-0.011	-0.004	-0.011	0.002	-0.028*
	(0.011)	(0.007)	(0.013)	(0.015)	(0.015)
Constant	0.083***	0.041***	0.088***	0.099***	0.129***
	(0.012)	(0.009)	(0.014)	(0.016)	(0.016)
Raw mean dep.var	4.222	0.536	1.355	0.801	1.531
Pval Diff. Nudge-Stress	0.376	0.587	0.323	0.642	0.313
Questions	64	21	20	9	14
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	2,543	2,543	2,543	2,543	2,543

Notes: The table presents estimates for each outcome variable in the panel and column headers following Equation 1. The outcome is defined as fraction omitted based on the total number of questions in each exam subject (specified in the Questions row). At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments. 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
Panel A: Women							
Nudge	0.015	-0.004	0.025**	-0.001	-0.019	-0.003	-0.008
	(0.020)	(0.007)	(0.013)	(0.021)	(0.019)	(0.019)	(0.016)
Nudge+Stress	0.059**	-0.021***	0.055***	0.049	-0.017	-0.006	-0.011
	(0.029)	(0.006)	(0.016)	(0.030)	(0.028)	(0.027)	(0.023)
Constant	0.702***	0.022**	0.859***	0.684***	0.206***	0.199***	0.142***
	(0.032)	(0.010)	(0.018)	(0.034)	(0.034)	(0.034)	(0.029)
Raw mean dep.var.	0.655	0.024	0.831	0.629	0.231	0.208	0.136
Pval Diff. Nudge-Stress	0.108	0.000	0.034	0.077	0.963	0.905	0.914
Obs.	2,115	2,115	2,115	2,115	2,115	2,115	2,115
Panel B: Men							
Nudge	0.004	0.001	0.002	-0.000	-0.022	-0.018	-0.027
	(0.017)	(0.003)	(0.008)	(0.018)	(0.020)	(0.020)	(0.018)
Nudge+Stress	-0.034	0.000	0.011	-0.049*	-0.053*	-0.050*	-0.034
	(0.024)	(0.004)	(0.011)	(0.026)	(0.027)	(0.027)	(0.024)
Constant	0.779***	0.014*	0.917***	0.751***	0.326***	0.322***	0.209***
	(0.024)	(0.007)	(0.012)	(0.026)	(0.031)	(0.031)	(0.027)
Raw mean dep.var.	0.809	0.005	0.934	0.780	0.337	0.326	0.230
Pval Diff. Nudge-Stress	0.092	0.768	0.376	0.045	0.221	0.199	0.772
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2,543	2,543	2,543	2,543	2,543	2,543	2,543

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1. At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments. 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 reports the estimates of the probability of enrollment in Phase 1. Column 5 reports the estimates of the likelihood of approving Phase 1. Column 6 reports the estimates of the probability of enrollment in Phase 2. Column 7 reports the likelihood of graduating. All columns show the fraction of students over the whole sample of applicants, regardless of whether they scored above the cutoff. Robust standard errors in parentheses. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

A 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 A2 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 A3. 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.

B Stress Reframing 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.

B.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"*

B.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.

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 A18.

Table A19 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.

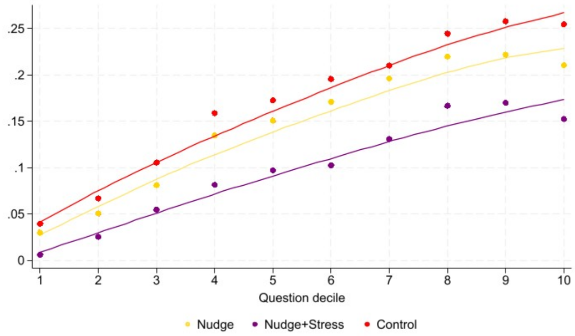
Second, the competition shifted to a single-sex environment, where applicants competed only against other women. Evidence from 2019, another year when only women participated, suggests that women's performance improved under such conditions (Gomez-Ruiz et al., 2024).[19] More broadly, research suggests 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.

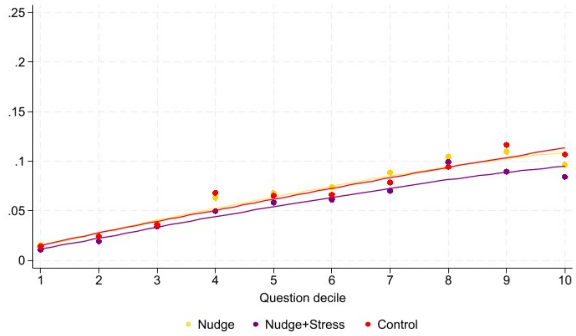
[19]The improved performance in 2019 has been attributed to the deactivation of gender stereotypes in women-only environments.

D Appendix Figures

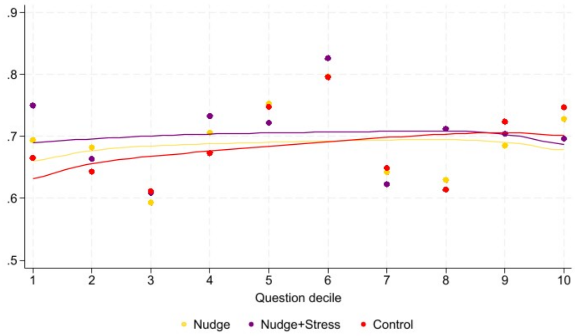
(a) Fraction omitted by question decile - Women



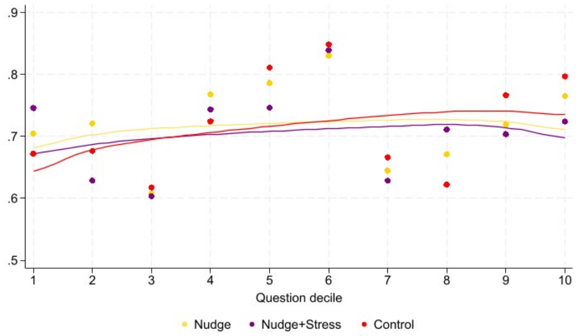
(b) Fraction omitted by question decile - Men



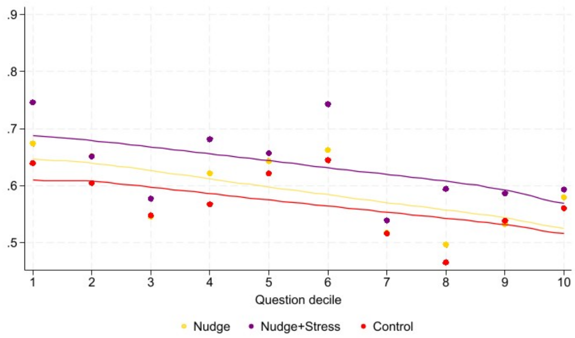
(c) Accuracy rate by question decile - Women



(d) Accuracy rate by question decile - Men



(e) Fraction correct by question decile - Women



(f) Fraction correct by question decile - Men

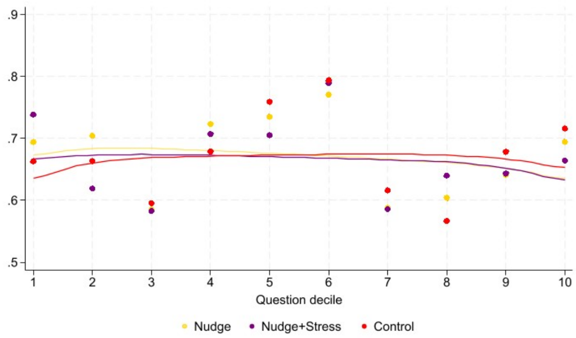


Figure A2: Fraction omitted, correct and accuracy rates by question decile and gender

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 A). All plots show, for treatment and control applicants separately, the mean fraction of omitted questions, accuracy rates and correct questions 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.

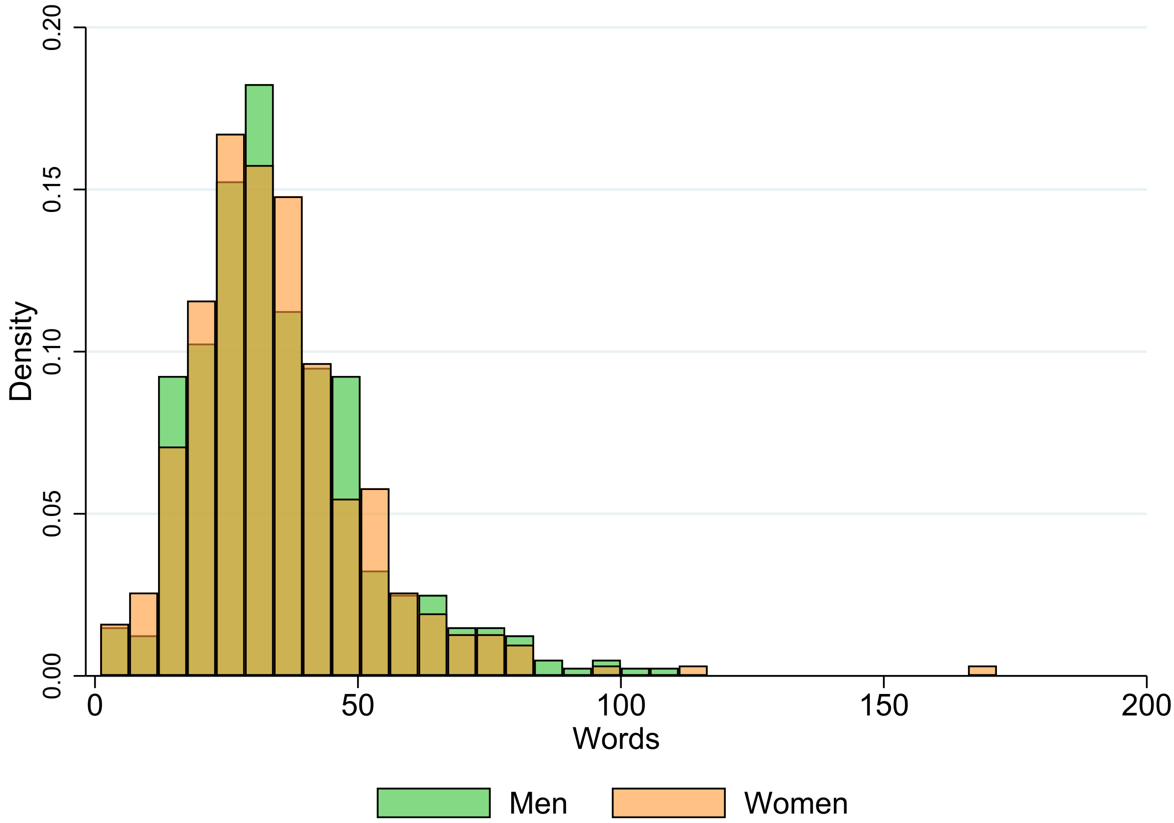


Figure A3: Gender differences in number of words written (treatment group only)

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.

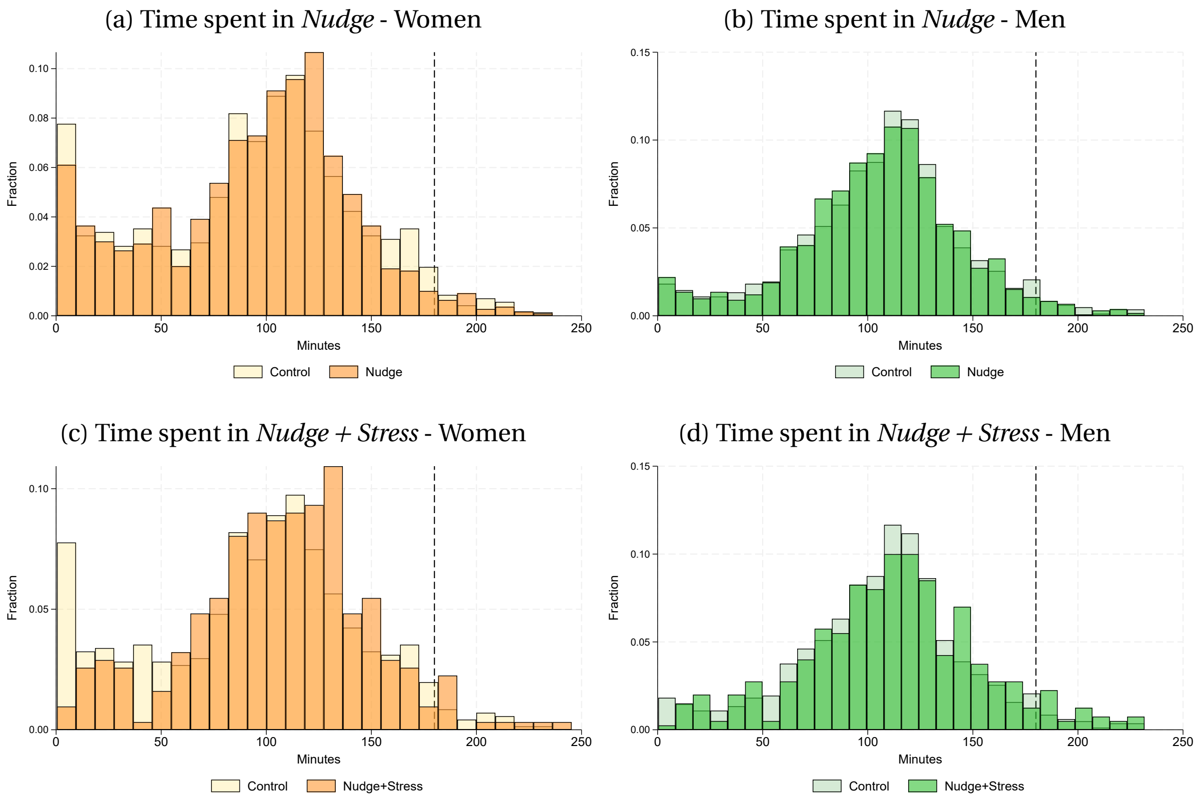


Figure A4: Gender differences in time spent answering the exam by treatment and gender

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: Difficulty of exam including stress 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 applicant-question level. The purpose is to compare the difficulty of exam version 4, which contained the stress reframing exercises, with the other six exam versions. $* \; p < 0.10$, $** \; p < 0.05$, $*** \; p < 0.01$.

Table A3: 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 exam 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 A4: Effects on performance (male interaction)

		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. 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 women 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 A5: Omitted questions and accuracy rate (male interaction)

	(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. 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 A6: Accuracy

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total accuracy	Verbal	Math	Concent.	Logic
Panel A: Women					
Nudge	0.007	0.011*	0.019*	0.007	-0.017
	(0.007)	(0.006)	(0.010)	(0.012)	(0.013)
Nudge+Stress	0.018*	0.026***	0.021	0.043**	-0.019
	(0.010)	(0.009)	(0.014)	(0.017)	(0.017)
Constant	0.692***	0.656***	0.725***	0.648***	0.766***
	(0.012)	(0.010)	(0.016)	(0.019)	(0.018)
Raw mean dep.var.	0.666	0.643	0.720	0.635	0.725
Pval Diff. Nudge-Stress	0.271	0.097	0.865	0.027	0.856
Obs.	2,075	2,075	1,861	1,775	1,753
Panel B: Men					
Nudge	0.010	0.022***	0.005	0.023**	-0.018*
	(0.006)	(0.006)	(0.008)	(0.011)	(0.010)
Nudge+Stress	0.002	0.012	-0.010	0.052***	-0.044***
	(0.009)	(0.008)	(0.011)	(0.014)	(0.014)
Constant	0.695***	0.654***	0.753***	0.627***	0.747***
	(0.009)	(0.008)	(0.012)	(0.016)	(0.015)
Raw mean dep.var	0.710	0.659	0.778	0.642	0.767
Pval Diff. Nudge-Stress	0.322	0.216	0.178	0.024	0.053
Questions	64	21	20	9	14
Controls	Yes	Yes	Yes	Yes	Yes
Obs.	2,528	2,528	2,430	2,392	2,377

Notes: The table presents estimates for each outcome variable in the panel and column headers following Equation 1. At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments. 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 A7: Effects on admission, exam completion and program continuation (male interaction)

	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 reports the estimates of the probability of enrollment in Phase 1. Column 5 reports the estimates of the likelihood of approving Phase 1. Column 6 reports the estimates of the probability of enrollment in Phase 2. Column 7 reports the likelihood of graduating. All columns show the fraction of students over the whole sample of applicants, regardless of whether they scored above the cutoff. Robust standard errors in parentheses. $*\ p < 0.10$, $**\ p < 0.05$, $***\ p < 0.01$.

Table A8: Heterogeneity by education and socio-economic status for women

	Exam performance					Continuation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total	Verbal	Math	Concentration	Logic	Admitted	Graduated
Panel A: Educational level							
Nudge	0.134**	0.142**	0.128**	0.142**	0.084	0.036	-0.006
	(0.058)	(0.065)	(0.060)	(0.057)	(0.057)	(0.030)	(0.019)
Some college or higher	0.416***	0.366***	0.372***	0.422***	0.365***	0.177***	0.062**
	(0.064)	(0.065)	(0.066)	(0.067)	(0.068)	(0.034)	(0.027)
Nudge × Some college or higher	-0.117	-0.071	-0.078	-0.194**	-0.119	-0.043	-0.006
	(0.078)	(0.081)	(0.081)	(0.083)	(0.081)	(0.040)	(0.032)
Nudge+Stress	0.298***	0.324***	0.331***	0.278***	0.144*	0.105**	0.026
	(0.082)	(0.083)	(0.086)	(0.084)	(0.083)	(0.045)	(0.030)
Nudge+Stress × Some college or higher	-0.224**	-0.198*	-0.301***	-0.144	-0.119	-0.090	-0.070
	(0.107)	(0.109)	(0.111)	(0.117)	(0.113)	(0.058)	(0.046)
Obs.	2,115	2,115	2,115	2,115	2,115	2,115	2,115
Panel B: Low SES							
Nudge	0.034	0.085	0.039	-0.026	0.010	0.033	-0.047*
	(0.059)	(0.057)	(0.061)	(0.065)	(0.063)	(0.029)	(0.027)
Low SES	-0.177**	-0.080	-0.198***	-0.175**	-0.170**	-0.071*	-0.029
	(0.072)	(0.075)	(0.074)	(0.076)	(0.077)	(0.038)	(0.030)
Nudge × Low SES	0.059	-0.004	0.077	0.110	0.041	-0.038	0.051
	(0.088)	(0.092)	(0.092)	(0.094)	(0.091)	(0.046)	(0.036)
Nudge+Stress	0.082	0.156*	0.048	0.155*	-0.005	0.008	-0.047
	(0.081)	(0.080)	(0.081)	(0.091)	(0.085)	(0.042)	(0.038)
Nudge+Stress × Low SES	0.084	0.057	0.124	0.013	0.069	0.049	0.022
	(0.123)	(0.126)	(0.129)	(0.132)	(0.127)	(0.066)	(0.050)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1,688	1,688	1,688	1,688	1,688	1,688	1,688

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. 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 A9: Heterogeneity by education and socio-economic status for men

	Exam performance					Continuation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Total	Verbal	Math	Concentration	Logic	Admitted	Graduated
Panel A: Educational level							
Nudge	0.037	0.116**	0.031	0.053	-0.041	0.017	-0.037
	(0.043)	(0.045)	(0.043)	(0.049)	(0.045)	(0.023)	(0.023)
Some college or higher	0.169***	0.118**	0.191***	0.185***	0.115*	0.080**	-0.031
	(0.058)	(0.059)	(0.059)	(0.070)	(0.065)	(0.033)	(0.036)
Nudge × Some college or higher	-0.017	-0.027	-0.051	-0.023	0.043	-0.034	0.024
	(0.061)	(0.064)	(0.061)	(0.074)	(0.067)	(0.033)	(0.037)
Nudge+Stress	0.025	0.030	-0.008	0.173***	-0.031	-0.054	-0.066**
	(0.057)	(0.064)	(0.058)	(0.063)	(0.059)	(0.033)	(0.029)
Nudge+Stress × Some college or higher	-0.018	0.091	0.005	-0.112	-0.075	0.057	0.087*
	(0.087)	(0.091)	(0.087)	(0.102)	(0.096)	(0.046)	(0.051)
Obs.	2,543	2,543	2,543	2,543	2,543	2,543	2,543
Panel B: Low SES							
Nudge	0.030	0.098**	0.021	0.066	-0.045	-0.006	-0.028
	(0.043)	(0.044)	(0.042)	(0.052)	(0.046)	(0.022)	(0.026)
Low SES	-0.067	-0.023	-0.063	-0.034	-0.103	-0.044	-0.021
	(0.060)	(0.061)	(0.060)	(0.069)	(0.065)	(0.032)	(0.034)
Nudge × Low SES	-0.012	0.016	-0.008	-0.106	0.019	-0.000	-0.010
	(0.073)	(0.077)	(0.075)	(0.085)	(0.078)	(0.040)	(0.040)
Nudge+Stress	-0.007	0.079	-0.011	0.110	-0.136**	-0.075**	-0.045
	(0.060)	(0.064)	(0.059)	(0.070)	(0.066)	(0.033)	(0.035)
Nudge+Stress × Low SES	0.051	-0.013	0.002	-0.017	0.187*	0.065	0.011
	(0.103)	(0.108)	(0.105)	(0.114)	(0.107)	(0.056)	(0.053)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2,049	2,049	2,049	2,049	2,049	2,049	2,049

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. 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 A10: Effects on performance (without controls)

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total score	Verbal	Math	Concentration	Logic
Panel A: Women					
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)
Constant	-0.000	-0.000	0.000	-0.000	-0.000
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)
Raw mean dep.var.	0.571	12.629	11.885	4.380	7.632
Pval Diff. Nudge-Stress	0.002	0.001	0.010	0.000	0.051
Obs.	2,115	2,115	2,115	2,115	2,115
Panel B: Men					
Nudge	0.005	0.106***	0.011	0.043	-0.024
	(0.034)	(0.032)	(0.031)	(0.037)	(0.034)
Nudge+Stress	-0.012	0.063	-0.007	0.132***	-0.058
	(0.047)	(0.047)	(0.043)	(0.049)	(0.046)
Constant	0.380***	0.153***	0.308***	0.263***	0.313***
	(0.026)	(0.051)	(0.048)	(0.054)	(0.049)
Raw mean dep.var.	0.699	0.334	0.659	0.052	0.433
Pval Diff. Nudge-Stress	0.672	13.549	14.553	5.281	9.622
Questions	64	21	20	9	14
Controls	No	No	No	No	No
Obs.	2,543	2,543	2,543	2,543	2,543

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1, excluding baseline covariates. At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments, along with the total number of exam questions considered in each outcome. All standardized outcomes are standardized based on the mean and SD of women 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 A11: Effects on omitted questions (without controls)

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total omitted	Verbal	Math	Concent.	Logic
Panel A: Women					
Nudge	-0.024*	-0.017*	-0.023	-0.020	-0.037**
	(0.014)	(0.010)	(0.017)	(0.019)	(0.017)
Nudge+Stress	-0.071***	-0.043***	-0.082***	-0.079***	-0.073***
	(0.016)	(0.010)	(0.020)	(0.023)	(0.023)
Constant	0.169***	0.073***	0.183***	0.234***	0.232***
	(0.011)	(0.008)	(0.013)	(0.015)	(0.027)
Raw mean dep.var.	10.820	1.523	3.656	2.107	3.533
Pval Diff. Nudge-Stress	0.001	0.003	0.001	0.004	0.093
Obs.	2,115	2,115	2,115	2,115	2,115
Panel B: Men					
Nudge	0.001	0.001	0.004	0.012	-0.008
	(0.008)	(0.005)	(0.010)	(0.011)	(0.012)
Nudge+Stress	-0.009	-0.003	-0.009	0.005	-0.025
	(0.011)	(0.007)	(0.013)	(0.015)	(0.016)
Constant	0.066***	0.026***	0.068***	0.089***	0.109***
	(0.006)	(0.004)	(0.008)	(0.009)	(0.010)
Raw mean dep.var.	4.222	0.536	1.355	0.801	1.531
Pval Diff. Nudge-Stress	0.307	0.504	0.258	0.581	0.254
Questions	64	21	20	9	14
Controls	No	No	No	No	No
Obs.	2,543	2,543	2,543	2,543	2,543

Notes: The table presents estimates for each outcome variable in the panel and column headers following Equation 1, excluding baseline covariates. The outcome is defined as fraction omitted based on the total number of questions in each exam subject (specified in the Questions row). At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments. Robust standard errors in parentheses. $* \, p < 0.10$, $** \, p < 0.05$, $*** \, p < 0.01$.

Table A12: Effects on accuracy (without controls)

		Performance by exam subject			
	(1)	(2)	(3)	(4)	(5)
	Total accuracy	Verbal	Math	Concent.	Logic
Panel A: Women					
Nudge	0.008	0.012*	0.017	0.005	-0.021
	(0.009)	(0.007)	(0.011)	(0.013)	(0.014)
Nudge+Stress	0.028**	0.035***	0.022	0.045**	-0.023
	(0.012)	(0.010)	(0.015)	(0.018)	(0.018)
Constant	0.666***	0.643***	0.720***	0.635***	0.725***
	(0.007)	(0.006)	(0.009)	(0.010)	(0.011)
Raw mean dep.var.	0.666	0.643	0.720	0.635	0.725
Pval Diff. Nudge-Stress	0.077	0.019	0.729	0.016	0.922
Obs.	2,075	2,075	1,861	1,775	1,753
Panel B: Men					
Nudge	0.005	0.018***	-0.000	0.016	-0.024**
	(0.007)	(0.006)	(0.008)	(0.011)	(0.010)
Nudge+Stress	-0.006	0.006	-0.019	0.042***	-0.053***
	(0.010)	(0.009)	(0.012)	(0.015)	(0.015)
Constant	0.710***	0.659***	0.778***	0.642***	0.767***
	(0.005)	(0.005)	(0.006)	(0.009)	(0.008)
Raw mean dep.var.	0.710	0.659	0.778	0.642	0.767
Pval Diff. Nudge-Stress	0.252	0.180	0.117	0.064	0.037
Questions	64	21	20	9	14
Controls	No	No	No	No	No
Obs.	2,528	2,528	2,430	2,392	2,377

Notes: The table presents estimates for each outcome variable in the panel and column headers following Equation 1, excluding baseline covariates. The outcome is defined as correct over attempted. At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments. Robust standard errors in parentheses. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table A13: 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
Panel A: Women							
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)
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)
Raw mean dep.var.	0.655	0.024	0.831	0.629	0.231	0.208	0.136
Pval Diff. Nudge-Stress	0.014	0.000	0.001	0.015	0.531	0.644	0.641
Obs.	2,115	2,115	2,115	2,115	2,115	2,115	2,115
Panel B: Men							
Nudge	-0.005	0.002	-0.001	-0.010	-0.036*	-0.032	-0.037**
	(0.018)	(0.003)	(0.008)	(0.019)	(0.021)	(0.021)	(0.018)
Nudge+Stress	-0.044*	0.000	0.009	-0.060**	-0.064**	-0.061**	-0.040
	(0.025)	(0.004)	(0.011)	(0.027)	(0.028)	(0.027)	(0.025)
Constant	0.809***	0.005**	0.934***	0.780***	0.337***	0.326***	0.230***
	(0.014)	(0.002)	(0.006)	(0.014)	(0.016)	(0.016)	(0.015)
Raw mean dep.var.	0.809	0.005	0.934	0.780	0.337	0.326	0.230
Pval Diff. Nudge-Stress	0.104	0.665	0.307	0.046	0.270	0.255	0.914
Controls	No	No	No	No	No	No	No
Obs.	2,543	2,543	2,543	2,543	2,543	2,543	2,543

Notes: The table presents estimates for each outcome variable in the column headers following Equation 1, excluding baseline covariates. At the bottom of each panel we report the mean of the raw dependent variable and the p-value of the differences across the two treatments. 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 reports the estimates of the probability of enrollment in Phase 1. Column 5 reports the estimates of the likelihood of approving Phase 1. Column 6 reports the estimates of the probability of enrollment in Phase 2. Column 7 reports the likelihood of graduating. All columns show the fraction of students over the whole sample of applicants, regardless of whether they scored above the cutoff. Robust standard errors in parentheses. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table A14: 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 A15: Effects on performance (reweighing using IPW)

	Performance				
	(1) Total score	(2) Verbal	(3) Math	(4) Concentration	(5) 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 A16: 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 A17: Gender differences in covariates using IPW weights

	Men (1)	Women (2)	Diff. (2)-(1) (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 A18: 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 A19: Effects on performance, exam completion and admission - 2024

	Exam completed		Admitted	Performance				
	(1) Fraction	(2) None	(3) Above cutoff	(4) Total score	(5) Verbal	(6) Math	(7) Concent.	(8) 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 described 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 A20: 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 in the 2024 admissions (women only): *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.$*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$.