Gender Differences in Prioritizing Rewarding Tasks*

Bruna Borges† Fernanda Estevan‡ Louis-Philippe Morin§

November 30, 2022

*Financial support from the Social Sciences and Humanities Research Council of Canada (430-2013-1033), Sao Paulo Research Foundation (FAPESP - grant # 2015/21640-3, 2017/50134-4 and 2019/25033-5), CALDO (2017/148485), British Academy and the Newton Fund (Newton Advanced Fellowship, AF140079), CNPq, FGV EESP, and University of Ottawa are gratefully acknowledged. Bruna received financial support from Sao Paulo Research Foundation (FAPESP - grants # 2017/23593-8 and # 2021/08880-6) and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Finance Code 001). We are very grateful to COMVEST, UNICAMP's admission office, for providing the data and assistance during the project. This research has benefited from discussions with Ana Maria Fonseca de Almeida, Sonia Bhalotra, Pierre Brochu, Abel Brodeur, Lena Edlund, Hülya Eraslan, Raissa Fabregas, Bruno Ferman, Jason Garred, Paola Giuliano, Daniel Haanwinckel, Rodolfo Jardim de Azevedo, Adam Lavecchia, Mario Macis, Rafael Maia, Daniele Paserman, Aureo de Paula, Christopher R. Taber, and Eric Verhoogen, and comments from seminar participants at PUC-Rio, University of Ottawa, FGV EESP, Universidad de la República, SBE Economia Aplicada, and the 48th ANPEC National Meeting. We also thank Derek Rice, Duangsuda Sopchokchai, Bogdan Urban, and Tiago Ferraz for their excellent research assistance. All remaining errors are ours.

†Sao Paulo School of Economics - FGV, Rua Itapeva, 474, Sao Paulo, SP, Brazil; email: brupugialli@gmail.com.

‡Sao Paulo School of Economics - FGV, Rua Itapeva, 474, Sao Paulo, SP, Brazil; email: fernanda.estevan@fgv.br.

§Department of Economics, University of Ottawa, 120 University, Ottawa ON K1N 6N5, Canada; email: lmorin@uottawa.ca.

Abstract

We investigate gender performance differences in reaction to task-rewards variation in a competitive environment. We use admission exam data from a selective Brazilian university, UNICAMP, to verify how the female-male performance gap changes in parts of the exam that count relatively more towards the final admission exam score, i.e., in which rewards are larger. In this environment, female and male applicants answer identical questions. However, the questions' rewards (weights) vary based on the applicant's major choice. The richness of our data allows us to flexibly control for applicants' major-choice self-selection issues through applicant fixed effects combined with multiple subject-specific ability measures. Our results suggest that females' performance decreases relative to males' when facing a larger-reward subject, even in a setting where all exams are high-stakes, there are no changes in the competition level, the pool of competitors, or the exam formats. Our question-level data allow us to use Item Response Theory (IRT) to investigate differences in exam strategy, something that the previous literature has been mostly unable to exploit. The evidence indicates that gender differences in confidence and exam strategy can explain our findings. Women tend to omit relatively more questions when the stakes increase, while men are more likely to answer them and earn a zero score. The gender gap in omissions emerges in male-dominated subjects, suggesting that gender stereotypes play a role. Also, women tend to spread their effort more equally across subjects, putting a higher effort on questions with lower rewards.

Keywords: post-secondary education, gender gap, high-stakes assessments.
JEL Codes: J16, I23, D80.

1 Introduction

Despite recent educational attainment progress, women are still underrepresented in high-profile jobs (Bertrand, 2018). Reaching top occupations typically involves performing well in competitive environments and prioritizing high- rather than low-rewarding endeavors. Not only is there evidence that women tend to underperform in competitive situations relative to men (Gneezy et al., 2003), but it also seems that women often devote more effort than men to comparatively less rewarding activities in the workplace (Babcock et al., 2017a,b).

In this paper, we investigate whether, as suggested by the previous literature, females and males react differently to variations in task rewards within a competitive setting where the potential payoffs are substantial. Our environment is one that many people will face at some point in their life, and for which the task is common: writing a university admission exam. Since we find gender differences in reaction to increased rewards, we next use the richness of our data to inquire into the potential channels through which this gender difference emerges.

We use admission exam data from a selective Brazilian university, UNICAMP, to verify how the female-male performance gap changes in parts of the exam that count relatively more towards the final admission exam score, i.e., in which rewards are higher. In our setting, we observe female and male applicants taking the same exams but with varying rewards within the exam. More specifically, UNICAMP applicants must write the admission exam in two phases, nearly two months apart. Both stages are composed of open-ended questions on typical high school subjects (e.g., biology, history, mathematics). In Phase 1, the final score is the unweighted average of all subjects. Therefore, there is no advantage in doing exceptionally well on a specific subject. In Phase 2, applicants must again answer questions on the same subjects. Still, their major choice determines the so-called 'priority subjects,' i.e., one or two subjects that receive a weight of two instead of one in the final score calculation. To increase their likelihood of being admitted, applicants should, all else equal, try to do better in these priority subjects.

Our setting allows us to overcome many pervasive obstacles in the literature. The timing, format, and content of both admission exam phases are identical for all applicants. Thus, we can analyze performance differences in priority subjects, allowing for subject-specific ability differences across genders. Notably, our data enable us to control both for applicant's overall ability (using individual fixed effects) and their subject-specific ability (using their performance in Phase 1, when exam items are equally weighted). Furthermore, we can rule out the possibility that females or males do not provide significant effort (or slack off) in Phases 1 or 2. Indeed, both phases are competitive since one must pass Phase 1 to advance to Phase 2, and relative (rather than absolute) performance determines admission.

These features allow us to avoid the potential confounding factors one faces when applicants have an objective, like attaining a specific final grade, and adjust their effort based on past performance (on midterms, for example).

We show that increased rewards affect the gender performance gap in a significant way. Moving from a non-priority to a priority subject reduces the female performance advantage by the equivalent of 9% of the within-applicant standard deviation. This effect is larger for higher-ability applicants, affecting admission in competitive majors. Indeed, we simulate admission without the gender gap in priority subjects and show that closing that gap would have a sizeable effect on majors with lower-than-average admission rates, such as medicine, engineering, and economics.

Not only do we observe applicants' performance in each subject, but we also have information on their performance on each of the twelve questions per subject. This level of detail in our data is particularly suitable for investigating potential channels explaining our findings. Results from our question-level analysis suggest that reaction to difficult questions and mental fatigue (Yoon et al., 2009) cannot explain our findings. Females actually perform relatively better on more difficult questions and questions at the end of the exam in priority subjects. Moreover, we can rule out gender-specific selection into priority subjects, gender differences in knowledge about UNICAMP rules, or reaction to negative shocks as potential mechanisms underlying our results.

We present evidence that gender differences in confidence and exam strategy are behind our results. First, we show that females and males adopt different strategies when faced with questions they are uncertain about the answer. Indeed, female applicants tend to omit relatively more priority-subject questions, while men attempt to answer them, but they are also more likely to get a score of zero. Time management issues do not seem to explain such patterns. Moreover, this omission pattern occurs precisely in male-typed subjects such as physics and mathematics, consistent with the literature documenting that confidence issues arise mainly in situations in which females suffer stereotype threats (e.g., Coffman, 2014; Bordalo et al., 2019; Exley and Kessler, 2022). Second, we show that women tend to put in a higher effort or preparation than men in questions associated with lower rewards. Conditional on ability, women have lower coefficients of variation across subjects in Phase 2, suggesting that they are more inclined than men to equalize their effort over distinct subjects. Using Item Response Theory, we calculate each applicant-question predicted score based on her estimated ability level and the question difficulty level. We show that female applicants' scores are closer to their predicted scores in priority subjects than male applicants, while the differences are much smaller in non-priority subjects. This evidence suggests that women also spread their effort more equally within priority-subject exams than men.

Our study closely relates to the literature stream that discusses gender differences in performance in high- versus low-stakes settings. While this literature suggests that women perform relatively worse than men when there are increased stakes, studies differ significantly in their capacity to account for other factors inherent to their settings.

A branch of this literature compares an individual's performance in different types of exams with varying stakes levels or across cohorts taking similar exams under different stakes levels. Ors et al. (2013) show that men outperform women at the highly selective HEC Paris admission exam (high-stakes) compared with the end-of-high-school and first-year core courses at HEC, which have lower stakes and are less competitive settings. Using data from a high-achieving private school in Barcelona, Azmat et al. (2016) show that men outperform women when stakes are higher by comparing their performance in different school exams with varying weights (from 5% to 27% of the final grade) and an end-of-high-school exam. More recently, Arenas and Calsamiglia (2022) examine the impact of a policy that increased the weight of the standardized high-stakes exam (from 40% to 57%) on college admission scores in Catalonia, reducing the relative weight of high school grades on admission scores. They find that the reform decreased female admission rates, mostly affecting high-achieving women. Better female performance in high school (compared to the high-stake exam) drives a substantial part of the impact. Still, it cannot explain all the differences, indicating that some of the policy's effects are due to behavioral responses.

Finally, other studies compare an individual's performance in a high-stakes exam with a mock or voluntary section of that same exam and have the advantage of comparing tests with the same format. Comparing mock results and the Chinese Gaokao exam from a province, Cai et al. (2019)'s results also show that men overperform women when stakes increase. However, they note that they cannot disentangle the exam 'stakes' effect and the degree of competition as the mock and actual examinations differ across both dimensions. One mechanism investigated by Cai et al. (2019) is whether men do not take the mock exam seriously since it does not affect the likelihood of university admission. They run a survey on Gaokao preparation (in particular, study time) and do not find evidence of gender differences in study effort (for the mock exam or actual Gaokao) or increased effort as the exam approaches. However, based on the high-stakes actual GRE and a low-stakes voluntary experimental GRE section, Schlosser et al. (2019) argue that the primary mechanism to explain differences in performance between mock and real exams is that men react more to incentives. At the same time, women tend to put in some effort even when stakes are 'nearly zero.'

Our work contributes to this literature by showing that the gender gap persists even in a context in which all exams are high-stakes, there are no changes in the competition level, the pool of competitors, or the exam formats, and applicants cannot adjust their effort based on previous performance, as they only learn about their ranking by the end of the admission process. Moreover, our detailed data at the exam-question level allows us to investigate differences in exam strategy at the question level that the previous literature has been mostly unable to exploit.

We organize the rest of the paper as follows. The next section details UNICAMP's admission exam process. We describe the administrative data used in our analysis in Section 3. Section 4 presents our empirical strategy. Section 5 shows our main results, including estimates for heterogeneity across subjects and applicant ability and robustness checks. In Section 6, we present the impact of the gender gap in terms of university admission. Finally, we investigate mechanisms that could potentially explain our results in Section 7 and provide a conclusion in Section 8.

2 UNICAMP Admission Exam

Each year, individuals applying to UNICAMP must write an admission exam (*vestibular*).[1] When registering for the admission exam (about two months before writing the exam), candidates apply to up to three majors (ranked first, second, and third options), so admission is program-specific. Given that UNICAMP uses an admission allocation based on the Boston mechanism and admission is competitive, the vast majority of applicants who are denied admission to their first option are not admitted to any of their three options (Estevan et al., 2019a).[2]

The admission exam is composed of two sequential phases (hereafter referred to as P_1 and P_2, respectively).[3] Applicants write P_1 in November, and P_2 in January (the academic year begins in late February). In both stages, all applicants write exactly the same examinations, regardless of their major choice.

In P_1, all applicants answer the same twelve written-answer questions on six high school subjects (i.e., two questions per subject), irrespective of the major they apply to. These subjects are biology, chemistry, geography, history, mathematics, and physics. Each subject carries the same weight in the calculation of the P_1 score — each question is worth 2.5 points. Importantly, applicants do not have any incentive to perform particularly well on a specific

[1]This section describes the rules governing the admission exam between 2001 and 2004, our analysis period. While rules changed over time, few changes occurred over our analysis period. We highlight these changes along with the discussion.

[2]Around 9% of UNICAMP applicants are admitted into the university. Among admitted applicants, almost 90% are admitted to their first-choice major.

[3]Additionally, some majors, like Performing Arts, require an aptitude test. We drop majors requiring an aptitude test from the analysis since their exam weighting scheme differs from the other majors.

high school subject in P_1 as they all carry the same weight — the rewards are identical across subjects. Applicants must also write an essay in P_1 worth 30 points. UNICAMP computes the applicants' P_1 score using two different formulas and selects the most favorable one for each applicant. The first formula corresponds to the sum of the essay and the general questions to determine the applicant's P_1 score. In the second formula, the sum of the essay and general questions receives a weight of 80%, and the applicant's ENEM score, an end-of-high-school exam score calibrated to a maximum score of 60, gets a weight of 20%.[4] Thus, under both formulas, the maximum possible P_1 score is 60 points.[5]

Applicants' P_1 score must be above a major-specific cutoff to qualify for P_2. The baseline cutoff score is 50% (or 30 points). However, UNICAMP adjusts the major-specific cutoff scores upward or downward to guarantee that the number of applicants per major in P_2 is between three and eight per available slot. For example, in 2003, the P_1 cutoff for Medicine was 84%, while for Statistics, it was 43%.

Applicants are allowed to write P_2 only if they pass P_1. The P_2 exam covers the same subjects as in P_1, plus Portuguese and a foreign language (English or French). There are twelve equally-weighted questions for each subject. P_2 is administered over four days (Day 1: Portuguese and biology; Day 2: chemistry and history; Day 3: physics and geography; Day 4: mathematics and English/French). Each day, applicants must complete the exams of both subjects and have a maximum of four hours to submit their answers. On any given day, both subject exams are provided to applicants simultaneously, and they are free to choose the order in which they answer each subject/question. However, we observe that the questions' average scores decrease with the questions' order, suggesting that applicants tend to answer the questions following the exam's layout.

As in P_1, P_2 exams are identical for all applicants and composed of written-answer questions (no multiple-choice questions). However, in contrast to P_1, applicants' scores on each subject are not equally weighted in the calculation of the P_2 score. Instead, depending on the applicant's major choice, one or two subjects are considered *priority subjects* and receive

[4]ENEM (*Exame Nacional do Ensino Médio*) is a national end-of-high-school exam used by some universities as the only admission criterion or as part of the admission process (like UNICAMP). Between 2001 and 2004, ENEM had 63 multiple-choice questions based on high-school subjects and an essay. UNICAMP considers only the ENEM score based on the multiple-choice questions. The applicant must have taken the ENEM in the two previous years and provided UNICAMP with her ENEM id. The formula that considers the applicant's ENEM score is advantageous for 85% of the applicants. In our analysis, we drop applicants without a valid ENEM score. Considering all UNICAMP candidates (excluding students writing the admission as a practice test), 88% submit valid ENEM scores. Among individuals who passed P_1 (the individuals we will focus on in our empirical analysis), the ENEM submission rate is 94%.

[5]In 2004, the maximum P_1 score was 120 points. Each question was worth 5 points, totalizing 60 points per subject. The essay was worth 60 points. Since the relative weights remained unchanged in 2004, this difference does not affect our analysis.

a weight of two (instead of one) in the final score calculation.

Appendix Table A.1 presents the list of majors considered in our empirical analysis along with their priority subjects, admission cutoff (the lowest-ranked P_2 score admitted in the major), the proportion of females among their applicants, and the percentage of candidates applying to this major (measured over the 2001-2004 period).[6] Note that we also include two majors offered at FAMERP (*Faculdade de Medicina de São José do Rio Preto*), medicine and nursing, as they use UNICAMP's exam for their admission. Most majors have two priority subjects, and priority subjects vary significantly across majors. There are some clusters of priorities (e.g., biology and chemistry for life-science programs, mathematics and physics for engineering programs), and these types of programs are more popular with a particular gender (e.g., life-science programs are usually more prevalent among females). However, and importantly, there remains significant and non-trivial variation in priority subjects and the proportion of females across majors. For example, chemical engineering, for which 45 percent of applicants are females, has chemistry and mathematics as priority subjects. Economics has history and mathematics as priorities, and around 40 percent of its applicants are females. Finally, 76% of food engineering applicants, who have mathematics and physics as priorities, are females.

An applicant's final admission score is the weighted average of her normalized:[7] 1) P_1 score, with a weight of two; 2) P_2 priority-discipline scores, each with a weight of two, and; 3) P_2 non-priority discipline scores, each with a weight of one. Thus, for a typical major with two priority subjects and no aptitude test, priority subjects count for one-third of applicants' final admission scores. Importantly, applicants who advanced to P_2 do not know their normalized P_1 score when they write P_2. Given that offers are based on applicants' ranking within a major choice, applicants do not know their admission likelihood even if they know their P_1 absolute performance. As such, we do not expect applicants to adjust their effort in P_2 based on their P_1 performance.

UNICAMP ranks applicants based on their final scores and major choices. The allocation mechanism is a version of the Boston mechanism, which initially considers applicants who chose the major as their first choice before those who put it as the second or third option.[8]

[6] Programs offered during daytime and the ones offered in the evening (labeled with '(Eve.)' in Appendix Table A.1) are treated as separate majors as they attract different types of potential students and are usually different in terms of admission competition (cutoffs).

[7] The P_1 score and P_2 subject scores are normalized to have a mean of 500 and a standard deviation of 100. Until 2003, the standardization of P_2 exams was done separately for applicants of majors within four defined areas. From 2004 on, the standardization considered the grades of P_2 exams of all candidates who participated in the exam. In all admission years, the standardization of the P_1 scores only considers the scores of candidates who passed P_1.

[8] See Estevan et al. (2019b) for more details on the admission procedure.

To summarize, the admission exam is designed such that two applicants writing identical exams could face different incentives regarding which questions/subjects should be prioritized given the exam rewards (weights) structure. Since P_1 is a qualifying exam, applicants cannot slack off in P_1, even if it only counts for a small portion of the final score. Due to this relatively small weight of P_1 on the final score and that applicants do not know their normalized P_1 performance before writing P_2, we would not expect applicants to adjust their performance/preparation for P_2 based on their P_1 raw score. In particular, given the competitive nature of the exam and that offers are based on relative performance, we are fairly confident that applicants with high P_1 scores will not slack off on P_2. Furthermore, applicants are not informed about their performance on individual P_2 subjects before receiving their overall final score. Hence, we do not expect applicants to adjust their preparation for, say, Day 4's subjects (i.e., mathematics and English/French) based on their performance on Day 1's (i.e., Portuguese and biology). Looking ahead, even though applicants with similar priority subjects may apply to different programs since one's major choice determines one's priority subjects, we will pay particular attention to major self-selection in our empirical strategy.

3 Data

We work with data provided by *Comissão Permanente para os Vestibulares* (COMVEST), UNICAMP's admission office. The dataset is available for 2000-2008 and contains individual-level information on all applicants who wrote UNICAMP's admission exam.

In this paper, we limit our analysis to 2001-2004, since UNICAMP implemented its affirmative action policy in 2005 (Estevan et al., 2019b). Focusing on the pre-affirmative action period allows us to leave any changes in the pool of applicants caused by the policy out of our analysis. In addition, we excluded the 2000 admission year since the number of admitted applicants to Phase 2 was between three and twenty per slot in 2000 instead of two to eight per slot in 2001-2004. Therefore, the number of applicants writing P_2 is much larger in 2000 than in 2001-2004, and their profiles are not necessarily comparable.

We concentrate on the six subjects, biology, chemistry, history, geography, mathematics, and physics, that are covered in both P_1 and P_2 (there are no questions specific to Portuguese or a foreign language in P_1). We do so to be able to control for applicants' subject-specific ability in our empirical approach.[9]

Importantly, for P_1 survivors, we see the applicants' grades in each question of P_1 and P_2

[9] As a robustness check, we use the essay as the P_1 score for Portuguese, and our results remain unchanged (see Table A.4).

in our dataset. More specifically, for each applicant, we observe twelve P_1 grades (two of each of the six subjects) and 72 P_2 scores (twelve of each of the six subjects).[10] For 2001-2002, we can also distinguish between an omitted question and an answered question that received a zero score. Since we know the applicants' major choices, we can easily identify their priority subjects (i.e., the P_2 subjects that carry a weight of two instead of one in the final-score calculation). The question-level performance information will allow us to investigate some potential channels through which a gender performance difference in reaction to increasing stakes could emerge.

Finally, the dataset contains ENEM scores for those applicants who provided their ENEM ids, corresponding to approximately 90% of the sample each year. We also know applicants' gender and other socioeconomic control variables such as age, parental education, and high school type.

Our initial sample contains 45,687 applicants who attended both phases of the exam for admission (not as a practice test)[11] and applied to a major not requiring an aptitude test. In addition, we drop applicants with missing gender information (0.6 percent), with ages below 16 or above 27 (2.5 percent), and with a missing ENEM score (4 percent). Our final sample contains 42,462 applicants for the 2001 to 2004 UNICAMP admission exams.

It is worth mentioning that we are looking at a selected pool of applicants, as we focus on applicants that attended both stages (Phases 1 and 2). UNICAMP is a highly competitive admission exam, and only 28% of the applicants pass Phase 1. Thus, we analyze gender gaps in performance for applicants that obtain better outcomes in a competitive environment, which is probably more similar to the real-world highly competitive labor markets.

Table 1 describes applicant-level information for our sample of interest. Female and male applicants differ in many meaningful dimensions, which will justify our empirical strategy presented in the next section. First, female applicants have, on average, significantly lower ENEM scores and apply to less competitive majors (based on the P_2 score cutoffs). Both differences are substantial. Females' average ENEM score is 0.25 s.d. below the overall average, while males' is 0.18 s.d. above. The difference in the average major cutoff (about 26 points) represents 0.33 s.d. of the major-cutoff distribution. The gender difference in P_1 performance is not as large as in ENEM performance, suggesting that females prepare better for the admission exam.

Given females and males apply to different majors (see Table A.1), it is not surprising that their priority subjects differ significantly. Forty percent of females have biology as a

[10] We actually observe 96 P_2 grades (twelve questions for each of the eight subjects) and the essay score (Phase 1). However, as mentioned above, our paper focuses on the six subjects covered in both stages.

[11] The admission exam is also open to students who wish to practice before finishing high school. Since these *trainees* are not eligible for admission, we exclude them from our analysis.

Table 1: Descriptive Statistics

	Full sample	Female	Male	Difference
Female	0.43			
	(0.49)			
Age	19.18	19.18	19.18	-0.00
	(1.62)	(1.56)	(1.67)	
Norm. ENEM scores	0.00	-0.25	0.18	-0.43***
	(1.00)	(1.04)	(0.93)	
# priority subjects	1.78	1.71	1.84	-0.14***
	(0.41)	(0.46)	(0.36)	
Major cutoff	518.21	503.38	529.35	-25.97***
	(78.19)	(82.81)	(72.56)	
Biology is a priority subject	0.27	0.40	0.18	0.22***
	(0.45)	(0.49)	(0.39)	
Chemistry is a priority subject	0.23	0.28	0.19	0.10***
	(0.42)	(0.45)	(0.39)	
Geography is a priority subject	0.02	0.01	0.02	-0.00***
	(0.13)	(0.12)	(0.13)	
History is a priority subject	0.18	0.22	0.14	0.08***
	(0.38)	(0.41)	(0.35)	
Mathematics is a priority subject	0.58	0.39	0.71	-0.32***
	(0.49)	(0.49)	(0.45)	
Physics is a priority subject	0.41	0.24	0.54	-0.30***
	(0.49)	(0.43)	(0.50)	
Portuguese is a priority subject	0.10	0.16	0.06	0.10***
	(0.30)	(0.37)	(0.24)	
Normalized P1 scores (average)	0.00	-0.09	0.07	-0.17***
	(0.63)	(0.66)	(0.59)	
Normalized P2 scores (average)	-0.00	-0.12	0.09	-0.21***
	(0.78)	(0.80)	(0.76)	
Normalized P2 scores (weighted average)	0.07	-0.06	0.16	-0.22***
	(0.79)	(0.81)	(0.77)	
Phase 2 score standard deviation	7.28	7.15	7.37	-0.22***
	(2.51)	(2.46)	(2.54)	
Phase 2 score coefficient of variation	0.32	0.34	0.31	0.03***
	(0.17)	(0.18)	(0.16)	
Within-applicant stand.dev - Norm. P2 scores	0.62	0.60	0.64	
Within-applicant stand.dev - Norm. P1 scores	0.78	0.78	0.78	
# Applicants	42,462	18,222	24,240	

Notes: We compute our descriptive statistics at the applicant level. The variable 'Age' is a student's age in June of the admission year. The ENEM scores are normalized to have a mean of zero and a standard deviation of one each year. The 'Major cutoff' corresponds to the final admission score of the student admitted with the lowest score in her first choice of major. We compute the normalized P_1 and P_2 scores by averaging our six subjects' normalized scores (biology, chemistry, geography, history, mathematics, and physics). The subject-specific scores are normalized to have a mean of zero and a standard deviation of one for each subject-year. We calculate the normalized Phase 1 and 2 'scores (average)' using equal weights for the six subjects. In contrast, 'Phase 2 scores (weighted average)' uses a weight of two for priority subjects and one for non-priority subjects. The within-applicant standard deviation captures the variation in performance across the six subjects within applicants. * significant at 10%; ** significant at 5%; *** significant at 1%.

priority subject, a proportion twice as large as for males. Biology is a priority subject for biological sciences, dentistry, medicine, nursing, pharmacy, physical education, and phonology. A number of these majors have only one priority subject (e.g., biological sciences, dentistry, and nursing), which explains why the number of priority subjects is lower for females. In contrast, males are significantly more likely to have mathematics (71% vs. 39%) or physics (54% vs. 24%) as priority subjects. Mathematics is a priority subject for mathematics, statistics, engineering programs, economics, information technology, construction technology, physics, computer science, media studies, environmental sanitation technology, and telecommunication technology.

Table 1 also shows gender differences in performance on P_2 for the weighted (using different weights for priority subjects relative to non-priority subjects) and unweighted averages. Males do better than females, but this is not surprising given the differences in ENEM and P_1 performances. The full-sample weighted average above 0 indicates that applicants perform better in their priority subjects and select topics they are relatively better at. Finally, we present the within-applicant standard deviation for both P_1 and P_2 performance. Since our variation of interest occurs at the applicant level (priority vs. non-priority subjects), the within-applicant variation is more informative than the overall variation (which combines both the within- and across-applicant variation in performance) to gauge the magnitude of our effect. As expected, the within-applicant standard deviation (0.62) is significantly lower than the overall standard deviation (normalized at 1.00).

In most cases, gender differences in major selection, and therefore priority-subject choice, would jeopardize the identification of our parameter of interest (i.e., the effect of increased rewards), especially if this selection is based on expected performance in the priority subjects. The richness of our data — being able to observe applicants' previous subject-specific performance and to include candidate fixed effects (or an overall academic ability measure) given that we observe six outcomes per applicant — along with our empirical strategy will allow us to control for major self-selection based on subject-specific ability (or gains in performance) in a flexible way.

4 Empirical Model

We now present an analytical framework to motivate our regression model and highlight its identification challenges. Imagine an applicant i writing an exam consisting of questions on different subjects, s. Some subjects, called priority subjects, are weighted more heavily than others to determine the exam's final score. Applicant i's performance on a specific subject

s, y_i^s, can be expressed as:

$$y_i^s = \rho^s + \pi_i + \gamma_i^s + \mathbb{P}_i^s \phi_i + \mathbb{P}_i^s \omega_i^s + \varepsilon_i^s, \tag{1}$$

where ρ^s represents the overall level of difficulty of subject s, π_i is the applicant's general academic ability, and γ_i^s is the applicant's ability specific for subject s. That is, some applicants might be better at some subjects than others. $\mathbb{P}_i^s$ is an indicator function equal to one if subject s is a priority subject for applicant i. ϕ_i is the applicant's overall (average) performance change when a subject is a priority while ω_i^s is the applicant's subject-specific additional performance change (over and above ϕ_i) when s is a priority. In this setup, we allow candidates' ability and their reaction to a priority subject to differ across subjects. Finally, ε_i^s is a purely random performance shock. Note that none of the terms on the right-hand side of the equation (1) are observed.

Given that we are interested in the group (i.e., gender) average performance changes when facing priority subjects, it is useful to express equation (1) in terms of deviations from group means:

$$y_i^s = \rho^s + \pi^g + \widetilde{\pi}_i^g + \gamma^{s,g} + \widetilde{\gamma}_i^{s,g} + \mathbb{P}_i^s[\phi^g + \widetilde{\phi}_i^g] + \mathbb{P}_i^s[\omega^{s,g} + \widetilde{\omega}_i^{s,g}] + \varepsilon_i^s, \tag{2}$$

where the parameters without i subscripts represent group averages (e.g., $\pi^g \equiv E(\pi_i|g)$ and $\omega^{s,g} \equiv E(\omega_i^{s,g}|s,g)$), and g stands for gender. Parameters with tildes are the applicant's deviations from these group averages (e.g., $\widetilde{\pi}_i^g \equiv \pi_i - \pi^g$). By construction, the expectations of tilde parameters are all equal to zero.[12]

Equation (2) suggests the following regression equation:

$$\begin{aligned} y_i^s &= \pi^m + \mathbb{F}_i\Delta\pi + \mathbb{P}_i^s(\phi^m + \omega^{s,m}) + \mathbb{P}_i^s\mathbb{F}_i(\Delta\phi + \Delta\omega^s) + \rho^s + \widetilde{\pi}_i^g + \gamma^{s,g} + \widetilde{\gamma}_i^{s,g} + \mathbb{P}_i^s[\widetilde{\phi}_i^g + \widetilde{\omega}_i^{s,g}] + \varepsilon_i^s \\ &= \beta_1 + \mathbb{F}_i\beta_2 + \mathbb{P}_i^s\beta_{3,s} + \mathbb{P}_i^s\mathbb{F}_i\beta_{4,s} + u_i^s, \end{aligned} \tag{3}$$

where $\mathbb{F}_i$ is a dummy variable equal to 1 if applicant i is a female (zero otherwise), m and Δ stand for male and gender difference, respectively. For example, π^m is male applicants' average general academic ability, and $\Delta\pi$ is the gender gap in general academic ability $(\pi^f - \pi^m)$. Finally,

$$u_i^s \equiv \rho^s + \widetilde{\pi}_i^g + \gamma^{s,g} + \widetilde{\gamma}_i^{s,g} + \mathbb{P}_i^s[\widetilde{\phi}_i^g + \widetilde{\omega}_i^{s,g}] + \varepsilon_i^s. \tag{4}$$

Note that if we were to assume that applicants' subject-specific reaction to facing a

[12]Note that, since all applicants write each and every subject, $E(\pi_i|s,g) = E(\pi_i|g), \forall s$.

priority subject is, on average, homogeneous across gender and subjects (i.e., $\omega^{s,g} = \alpha$, $\forall s, g$), then $\beta_{3,s}$ and $\beta_{4,s}$ in equation (3) would become subject invariant (e.g., $\beta_{4,s} = \beta_4, \forall s$).[13] In this case, our parameter of interest would be β_4, the effect of increasing rewards on the gender performance gap (or the gender difference in performance change when moving from a non-priority to a priority subject, all else equal). If, instead, we let applicants' reactions vary across subjects, then we will have to estimate a full set of $\beta_{3,s}$ and $\beta_{4,s}$. We will present results for both specifications.

The definition of the error term in equation (4) highlights the main challenges when trying to estimate $\beta_{4,s}$ using a standard difference-in-difference approach (whether we allow it to vary across subjects or not). The first two terms in the error term (ρ^s and $\widetilde{\pi}_i^g$) are usually controlled for in the previous literature (see, for e.g., Azmat et al., 2016; Cai et al., 2019) using 'test-type' and individual fixed effects. In our case, one should control for each subject's difficulty (ρ^s), especially if harder subjects are more likely to be a priority for a specific gender. We do this by including subject fixed effects in our regressions. Moreover, if the general-academic distributions differ across gender and females and males apply to different programs, then we would expect some correlation between $\mathbb{P}_i^s \mathbb{F}_i$ and the applicant's relative general academic ability ($\widetilde{\pi}_i^g$). In some specifications, we will use $\widetilde{ENEM}_i^g$, the applicant's relative performance on the ENEM exam (performance minus group g average, as defined above) to control for relative general academic ability. However, we will use applicant fixed effects in our preferred specifications, as in Azmat et al. (2016) and Cai et al. (2019).

An additional concern for us is that applicants are tested on different subjects, some of which may favor female or male applicants (captured by $\gamma^{s,g}$ in equation (4)).[14] If priority subjects are more likely to be subjects for which male applicants are, on average, better, then we could falsely attribute the gender difference in performance change when moving from a non-priority to a priority subject to the effect of increasing rewards. This potential issue motivates using gender-specific subject fixed effects in our regression.

The presence of $\widetilde{\gamma}_i^{s,g} + \mathbb{P}_i^s[\widetilde{\phi}_i^g + \widetilde{\omega}_i^{s,g}]$ highlights a common issue when trying to estimate a model with random coefficients. One would wish these terms to be uncorrelated with our regressor of interest ($\mathbb{P}_i^s \mathbb{F}_i$), but given that applicants choose their major, and therefore their priority subjects, such assumption is unlikely to hold. It is quite plausible that applicants choose their major based on their relative overall performance gain ($\widetilde{\phi}_i^g$) or their comparative advantages (e.g., $\widetilde{\gamma}_i^{s,g}$ or $\widetilde{\omega}_i^{s,g}$), and major self-selection could differ across gen-

[13]Furthermore, if we assume that applicants' subject-specific reaction to facing a priority subject is, on average, homogeneous across gender, for each subject (i.e., $\omega^{s,g} = \alpha^s, \forall s, g$), then $\beta_{4,s}$ will become subject invariant (e.g., $\beta_{4,s} = \beta_4, \forall s$), but $\beta_{3,s}$ will still vary across subjects.

[14]See, for example, Ellison and Swanson (2010), Niederle and Vesterlund (2010), and Ellison and Swanson (forthcoming) for discussions on gender differences in mathematics ability.

der. Such self-selection based on comparative advantage (or performance gain) is the main ingredient of correlated random coefficient (CRC) models (see, e.g., Heckman and Vytlacil, 1998; Wooldridge, 2005). Usually, estimating a treatment effect in the presence of correlated coefficients is challenging and requires using control functions, instrumental variables, or selection models (Dahl, 2002; Heckman et al., 2006).

Fortunately, the richness of our data allows us to control for such selection. Each applicant has to answer questions on the same subjects during Phase 1 and Phase 2 of the admission exam. Importantly, in Phase 1, each subject is equally weighted to determine who will move on to the second phase and overall performance in Phase 1 has a small impact on the likelihood of being admitted (it weights 2 in the calculation of the final grade, equivalent to one priority subject), conditional on making it to the second stage.[15] There is no incentive to do well in particular subjects. Hence, we can use applicants' performance on each Phase 1 subject to control for their relative subject-specific ability ($\widetilde{\gamma}_i^{s,g}$).

Finally, to control for the overall and subject-specific performance gains from priority subjects, we assume that the main predictors of these gains are flexible functions of the applicant's relative general and subject-specific abilities. More specifically, we use quartic functions of $\widetilde{P1}_i^{s,g}$ and $\widetilde{ENEM}_i^{g}$, each interacted with the priority dummy variable to model $\mathbb{P}_i^s[\widetilde{\phi}_i^g + \widetilde{\omega}_i^{s,g}]$.

Equations (3) and (4), along with the available applicant information, motivate our main regression equation:

$$y_i^s = \beta_{3,s}\mathbb{P}_i^s + \beta_{4,s}\mathbb{P}_i^s\mathbb{F}_i + G(\widetilde{P1}_i^{s,g}) + H(\mathbb{P}_i^s \times \widetilde{ENEM}_i^{g}) + J(\mathbb{P}_i^s \times \widetilde{P1}_i^{s,g}) + \rho^s + \pi_i + \gamma^{s,g} + v_i^s, \tag{5}$$

where $\widetilde{P1}_i^{s,g}$ is the applicant's relative performance in subject s on Phase 1 and $\widetilde{ENEM}_i^{g}$ is the applicant's relative performance on the ENEM exam. $G(\cdot)$, $H(\cdot)$ and $J(\cdot)$ are flexible functions meant to capture $\widetilde{\gamma}_i^{s,g}$, $\mathbb{P}_i^s\widetilde{\phi}_i^g$ and $\mathbb{P}_i^s\widetilde{\omega}_i^{s,g}$, respectively.[16] Subject and individual fixed effects will capture ρ^s and π_i, while gender-subjects interactions will capture $\gamma^{s,g}$. Note

[15]The fact that P_1 performance has only a weight of 2 in the calculation of the final admission grade (and that applicants do not know their P_1 relative performance or ranking) is another important advantage over some of the previous studies on the subject (e.g., Azmat et al., 2016). Suppose P_1 had a significant weight in the final grade (like a midterm counts significantly towards a course final grade), and applicants were aware of their P_1 relative performance (i.e., what matters for admission). In that case, they could react by providing more or less effort on P_2. For example, some applicants could slack off on P_2 (or a final exam), knowing that their P_1 (or midterm) relative performance allows them to do so. If the gender differential in reaction to P_1 achievement varied across priority and non-priority subjects, such differential would lead to a biased estimation of β_4.

[16]In our main specification, these functions have a quartic form. For example, $G(\widetilde{P1}_i^{s,g}) \equiv \sum_{j=1}^{4} \alpha_j (\widetilde{P1}_i^{s,g})^j$.

that the constant term (β_1), the gender gap in overall performance (β_2), and the applicant relative ability ($\widetilde{\pi}_i^g$) will be absorbed by the individual fixed effects (π_i). Our performance measure y_i^s is the applicant's Phase 2 score in subject s, normalized by subject and admission year.[17]

4.1 Validity

We now show that there is no evidence of gender-specific selection differences into priority subjects, minimizing concerns of biases in our estimated coefficients. A potential concern would arise if, for instance, men consider their comparative advantage when selecting their major and women do not, or consider it to a lesser extent. In that case, the gender gap could derive from gender differences in selection, not gender differences in reaction to rewards.

As we discussed above, our empirical strategy deals with gender differences in comparative advantage, and we do not expect this endogeneity issue to affect our estimates. In any case, we present evidence that supports our identification strategy's validity. Specifically, we investigate whether we observe gender performance differences in 'future' priority subjects during Phase 1 of the admission exam (remember that in P_1 all subjects are equally weighted). If gender differences in comparative advantage explained our findings, we would expect females to perform worse during P_1 in subjects that will become priorities in P_2. Besides, if females were to concentrate on P_1 disciplines that will be priorities in P_2 while males do not, then one could argue that females cannot improve much more (at least not as much as males) in priority subjects between P_1 and P_2. Such a situation would have consequences for interpreting our results from our preferred specification (5), where we control for P_1 subject-specific performance. Female specialization in P_1 and P_2 combined with a lack of reaction in P_1 and specialization in P_2 from males could lead to a negative estimate for our parameter of interest (i.e., a negative bias), despite females being the ones reacting more to the increased rewards.

Table 2 presents the results from estimating regressions where applicants' P_1 subject-specific scores are regressed on a dummy variable, 'Future Priority,' equal to 1 if the subject will be a priority in P_2 (and zero otherwise), and a 'Female $\times$ Future Priority' interaction term, along with the same regressors as in our main specification (except for the controls for P_1 scores). Two findings come out of Table 2. First, applicants do better in subjects that will be a priority in P_2, suggesting that they apply to majors in which they have a comparative advantage. This finding suggests that we should control for P_1 scores in our P_2 regressions to control for such selection. Second, and importantly, females do not seem

[17] As a robustness test, we use raw Phase 2 scores as our dependent variable, controlling for year fixed effects. Results remain similar. We report these estimates in Table A.3.

to concentrate more or underperform in subjects that will become priorities in P_2. The 'Female × Future Priority' parameter estimates are statistically non-significant and small compared to the within-applicant standard deviation for P_1 scores (0.78). These findings suggest that applicants' potential reaction to future priority subjects in P_1 is not a major cause for concern when interpreting our main results.

Table 2: Priority and P_1 Subject-Specific Performance

	(1)	(2)	(3)	(4)
Dependent variable: Phase 1 normalized subject-specific scores				
Female	-0.160***	-0.112***		
	(0.005)	(0.009)		
Future priority	0.301***	0.359***	0.325***	0.341***
	(0.005)	(0.006)	(0.006)	(0.007)
Female × Future priority	0.025***	0.006	-0.009	-0.007
	(0.008)	(0.009)	(0.009)	(0.009)
ENEM	0.385***	0.384***		
	(0.003)	(0.003)		
Number of observations	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes
ENEM × Priority	No	No	No	Yes

Notes: P_1 subject-specific scores are normalized to mean 0 and standard deviation 1 for each subject-year. 'Future priority' is a dummy variable equal to 1 if the subject will be a priority in P_2. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicant's normalized ENEM score minus her/his gender-year group average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. We use a quartic function to control for relative ENEM performance interacted with 'Future priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Furthermore, we analyzed whether P_1 performance in a subject correlates with the likelihood of having this subject as a priority in Phase 2 (see Table A.2). Our response variable is a priority subject binary variable that equals one if a subject was a priority for the applicant in Phase 2 and zero otherwise. We regress the priority subject dummy on normalized Phase 1 score, a female dummy, and the interaction of the gender dummy and P_1 scores. In column (2), we include subject fixed effects; in column (3), we add subject-gender fixed effects; and in column (4), we control for the normalized ENEM scores. Column (5) includes individual fixed effects. We find that a higher score in Phase 1 positively relates to the priority subjects' choice. Once we control for individual fixed effects (column (5)), we observe that a one-standard-deviation increase in P_1 score associates with an increase in the probability that the applicant selects the priority subject by eight percentage points for men and nine percentage points for women. If anything, females seem slightly more inclined to rely on their comparative advantages when choosing their major, not less. Overall, while one's comparative advantages, measured by higher subject-specific scores, correlate with the choice of priority subjects, the gender differences in the correlations are small and do not seem to threaten our empirical strategy.

5 Main Results

Table 3 presents the results from estimating different variations of equation (5), imposing a common β_4 across subjects. In all specifications, we cluster our standard errors at the applicant level. In column (1), we look at the overall gender difference in performance, controlling for overall ability. Our measure of ability, ENEM, looks like a good predictor of applicants' performance. A one-standard-deviation increase in relative ENEM performance is associated with a 0.54 s.d. increase in performance (with a t-statistic over 100). Note that when we control for group relative ENEM performance ($\widetilde{ENEM}_i^g$), the 'Female' coefficient estimate captures the gender gap in non-priority-subject performance emerging from two sources: the part that is due to the average ENEM gender performance gap, and the part that is unexplained by ENEM.

The specification in column (1) of Table 3 does not attempt to control for applicants' selection into majors other than through a linear control for overall relative ability. Since Table 1 suggests gender differences in major selection, it is unlikely that the β_4 estimate in column (1) represents causality unless our measure of ability completely captures this selection. Therefore, the following specifications sequentially attempt to capture more complex gender/individual differences in major selection.

Column (2) includes subject fixed effects and female-subject interaction terms, which allow females to perform, on average, better (or worse) in different subjects. In this case, the parameter for 'Female' represents the gender performance gap in biology. The estimate on the 'Female × Priority' is sizable as it represents about 8 percent of the within-applicant standard deviation (0.62) — the within-applicant standard deviation captures the variation in performance across all subjects within applicants. Looking at the subject fixed effects, applicants tend to perform on average best in geography and worst in mathematics. The female-subject interaction terms suggest that females perform, on average, worse than males (controlling for ENEM) in all six subjects, with the largest gender gaps in physics and chemistry.

We introduce a more flexible way to capture overall ability in column (3) by dropping ENEM and replacing it with individual fixed effects.[18] The main impact of using fixed effects

[18]Doing so, the fixed effects absorb the 'Female' dummy.

Table 3: Priority Subjects and Gender Performance Gap

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.196***	-0.084***				
	(0.006)	(0.008)				
Priority	0.482***	0.582***	0.544***	0.481***	0.497***	0.515***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.001	-0.051***	-0.065***	-0.062***	-0.054***	-0.053***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
ENEM	0.542***	0.540***				
	(0.003)	(0.003)				
Chemistry		0.099***	0.099***	0.101***	0.101***	0.099***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Geography		0.168***	0.162***	0.149***	0.149***	0.147***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
History		0.073***	0.071***	0.066***	0.066***	0.064***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mathematics		-0.134***	-0.114***	-0.088***	-0.088***	-0.094***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Physics		-0.001	0.013**	0.026***	0.025***	0.021***
		(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Female × Chemistry		-0.174***	-0.180***	-0.189***	-0.190***	-0.188***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Geography		-0.061***	-0.074***	-0.087***	-0.087***	-0.087***
		(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Female × History		-0.044***	-0.052***	-0.061***	-0.061***	-0.061***
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Female × Mathematics		-0.096***	-0.116***	-0.148***	-0.147***	-0.145***
		(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Female × Physics		-0.193***	-0.214***	-0.244***	-0.243***	-0.241***
		(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Number of observations	254,772	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462	42,462
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and the standard deviation 1 for each subject-year. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

instead of ENEM (column (3) versus (2)) is to increase the magnitude of our parameter estimate of interest to 10 percent of a within-applicant standard deviation.

Column (4) introduces a flexible (quartic) function of the applicant's subject-specific relative performance in P_1. These additional covariates allow us to go a step further in controlling for potential selection based on comparative advantage. Under this specification, the applicant's overall ability will be captured by the individual fixed effects, while her (additional) subject-specific ability will be captured by the quartic function in P_1 scores. The main parameter estimate does not change significantly.

Finally, in columns (5) and (6), we add more covariates to control for major self-selection by interacting both the applicants' relative overall and subject-specific performances with the priority-subject indicator variable. The introduction of these covariates is meant to capture major selection based on individual absolute and comparative advantages ($\mathbb{P}_i^s[\widetilde{\phi}_i^g + \widetilde{\omega}_i^{s,g}]$ in equation (4)). Despite controlling for overall and subject-specific ability in fairly flexible ways, we still find that applicants' performance increases significantly when facing priority subjects, but females do not react as much as males when facing these increased rewards.

5.1 Heterogeneity

The imposition of a common effect of increased rewards across subjects in Table 3 may seem restrictive, especially since our results suggest that females and males perform differently from one subject to the other, even after controlling for ENEM and P_1 scores. To investigate the potential heterogeneity in the priority-subject effects across subjects, we estimate equation (5) allowing β_3 and β_4 to vary across subjects.

Table 4 presents the results when we estimate equation (5) without imposing common β_3 and β_4. We focus our discussion on specification (6), in which we have our full set of controls. We do not find gender differences in reaction when rewards are larger in biology or chemistry but observe large reactions (around -0.11, or 18 percent of a within-applicant standard deviation) when rewards increase in any other subjects. Looking back at our descriptive statistics (Table 1), we see that the two subjects for which we do not find gender differences in reaction to rewards (biology and chemistry) are also the two subjects where the gender gap in priority-subject proportions are the largest in favor of females.[19] We will

[19]Biology and chemistry are the priority subjects for medicine majors (UNICAMP and FAMERP), and medicine is the most popular and competitive major at UNICAMP. We test whether our results are robust to the exclusion of medicine applicants. The β_4 estimate is larger when we restrict our sample: female applicants' performance is 0.086 s.d lower (14% of the within-applicant standard deviation). Looking at heterogeneous impacts by subject, we still observe no gender differences in reaction to increased rewards in biology and a relatively better female performance in chemistry when the discipline is a priority (by 0.055 s.d. and statistically significant at 10%). We present these results in the Online Appendix Tables O.1 and

Table 4: Heterogeneity Across Subjects

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.196***	-0.125***				
	(0.006)	(0.009)				
Priority	0.649***	0.742***	0.372***	0.311***	0.311***	0.333***
	(0.011)	(0.012)	(0.010)	(0.010)	(0.010)	(0.011)
Female × Priority	0.035***	-0.036**	0.003	0.009	0.017	0.012
	(0.013)	(0.016)	(0.013)	(0.013)	(0.013)	(0.013)
Female × Priority × Chemistry	-0.078***	0.086***	0.034*	0.027	0.021	0.024
	(0.017)	(0.021)	(0.018)	(0.018)	(0.018)	(0.018)
Female × Priority × Geography	-0.102	-0.083	-0.114*	-0.107*	-0.131**	-0.126**
	(0.066)	(0.068)	(0.064)	(0.063)	(0.063)	(0.063)
Female × Priority × History	-0.101***	-0.112***	-0.110***	-0.119***	-0.115***	-0.112***
	(0.023)	(0.028)	(0.025)	(0.024)	(0.024)	(0.024)
Female × Priority × Mathematics	-0.125***	-0.026	-0.097***	-0.107***	-0.106***	-0.098***
	(0.017)	(0.026)	(0.022)	(0.021)	(0.021)	(0.021)
Female × Priority × Physics	-0.261***	-0.126***	-0.124***	-0.125***	-0.123***	-0.115***
	(0.018)	(0.025)	(0.021)	(0.020)	(0.020)	(0.020)
Priority × Chemistry	-0.012	-0.136***	-0.168***	-0.143***	-0.144***	-0.146***
	(0.013)	(0.015)	(0.013)	(0.013)	(0.013)	(0.013)
Priority × Geography	0.152***	-0.044	0.569***	0.544***	0.585***	0.576***
	(0.045)	(0.046)	(0.043)	(0.042)	(0.042)	(0.042)
Priority × History	0.017	-0.072***	0.564***	0.537***	0.550***	0.543***
	(0.017)	(0.020)	(0.018)	(0.017)	(0.017)	(0.017)
Priority × Mathematics	-0.309***	-0.314***	0.219***	0.230***	0.240***	0.227***
	(0.013)	(0.019)	(0.016)	(0.016)	(0.016)	(0.016)
Priority × Physics	-0.147***	-0.193***	0.220***	0.206***	0.213***	0.201***
	(0.013)	(0.018)	(0.015)	(0.014)	(0.014)	(0.014)
ENEM	0.536***	0.535***				
	(0.003)	(0.003)				
Number of observations	254,772	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized to mean 0 and standard deviation 1 for each subject-year. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

return to these findings when discussing the potential channels through which the change in the gender performance gap emerges as rewards increase.

Next, we consider whether the effect of increasing rewards on the gender performance gap varies with applicant ability. There are a few ways to approach this question, given our setup. We present the results where we interact 'Female × Priority' with the ENEM relative performance measure and use a linear function of 'ENEM × Priority' instead of a quartic function. We make this last change to see more easily if, on average, the effect of increased rewards on the gender performance gap increases or decreases with ENEM.[20]

Table 5 suggests that the impact of higher rewards on the gender performance gap increases in magnitude with applicants' relative ENEM performance. At the mean relative ENEM performance, the effect of increased rewards on the gender gap is close to what we found in Table 3 (-0.058 versus -0.053). The difference in performance between priority and non-priority subjects increases for males with higher ENEM scores. All else equal, a male applicant with an ENEM score one standard deviation above its group mean would have improved his performance on a priority subject by 0.051 s.d. more than a male applicant with an ENEM score at the mean. For females, this difference would be 0.020 (0.051-0.031). This difference is statistically significant at a 1% level. So, as we compare females and males with higher ENEM scores, the gender performance gap becomes more and more in favor of male applicants. For applicants one standard deviation above the group mean, increasing rewards will change the gender performance gap by 0.089 (14% of the within-applicant standard deviation) in favor of males.[21] Given that admission to UNICAMP is very competitive, our findings suggest that the gender difference in reaction to increased rewards could affect the gender representation of admitted students.[22]

O.2.

[20]Using a quartic function instead of a linear function has little impact on our 'Female × Priority' and 'Female × Priority × ENEM' parameter estimates. We display the results in the Online Appendix Table O.3.

[21]Doing a similar exercise, but looking at the link between P_1 subject-specific performance (instead of ENEM) and the effect of increased rewards on the gender gap yields similar results. See Online Appendix Table O.4.

[22]We have run the regressions separately for female and male applicants to analyze the pattern of high-achieving candidates in priority subjects by gender. Men and women perform better in priority subjects, and the reaction to rewards is higher for the 'high-ability' applicants in both genders. However, the coefficients for males are larger both for the average and high-performing applicants. We can interpret this result as further evidence that women react less than men to an increase in rewards. We display these results in Online Appendix Tables O.5 and O.6.

Table 5: Heterogeneity Across Academic Ability

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Female	-0.195***	-0.084***			
	(0.006)	(0.008)			
Priority	0.481***	0.580***	0.542***	0.478***	0.501***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.004	-0.052***	-0.062***	-0.060***	-0.058***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Priority × ENEM	-0.007	-0.012*	-0.031***	-0.034***	-0.031***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Female × ENEM	0.019***	0.021***			
	(0.006)	(0.006)			
Priority × ENEM	0.048***	0.041***	0.036***	0.039***	0.051***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
ENEM	0.521***	0.521***			
	(0.005)	(0.005)			
Number of observations	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized to mean 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use a quartic function to control for the relative P_1 performance and its interaction with 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

5.2 Robustness Checks

As a first robustness check, we ran the same regressions, including applicants' performance in Portuguese, to investigate whether its exclusion from our main analysis affects our findings. We use the P_1 essay score to control the applicant's Portuguese ability. Note that the regressions are estimated on the same set of applicants, but the number of observations increases since we add one P_2 performance outcome for each applicant. The results presented in Appendix Table A.4 are very similar once we include subject-gender and individual fixed effects.

Our analytical framework suggests using equation (5) as our main econometric model and controlling for relative performance in Phase 1 scores in our regressions. However, one could model performance changes between P_1 and P_2 instead to capture applicants' improvement, as done in Cai et al. (2019) and Schlosser et al. (2019). In our case, since we control for P_1 in our main specifications, estimating our model in difference without controlling for P_1 performance is equivalent to imposing $G(\widetilde{P1}_i^{s,g}) = \widetilde{P1}_i^{s,g}$ in equation (5).[23] In any case, estimating the model in difference allows us to shed more light on the robustness of our results. Therefore, we use the difference between Phase 2 and Phase 1 normalized scores as the dependent variable and estimate our model without controls for $G(\widetilde{P1}_i^{s,g})$. We present the results using this alternative response variable in Table A.5. Our coefficient estimates for 'Female × Priority' are very close to the ones presented in Table 3 once we control for subject-gender fixed effects. They vary between -0.048 and -0.057 in Table A.5 and between -0.051 and -0.065 in Table 3.[24] The only noteworthy difference is that the coefficient estimates for 'Priority' are significantly smaller when we do not control for P_1 performance. So, overall, our results suggest that our findings are not sensitive to whether we consider the P_2 performance in level while controlling for P_1 or consider our model in difference, as in Cai et al. (2019) and Schlosser et al. (2019).

6 Does the Gender Gap Matter for Admission?

Our results presented so far show that females' performance increases less than males' when facing larger rewards and that this difference increases for higher-ability applicants. Although these findings suggest that the gender performance gap could impact university admission, the actual effect hinges upon the share of female applicants and their performance distribution within majors.

To grasp the impact of the gender performance gap, we simulate university admission, considering only the applicants' first choice of major, using their actual and adjusted scores. More precisely, we adjust female applicants' scores in priority subject exams by considering the coefficient estimates obtained in a specification analogous to column (5) of Table 5, but where we replace the normalized subject-specific scores for the raw Phase 2 scores. First, we run the regressions to obtain the 'Female × Priority' and 'Female × Priority × ENEM' coefficient estimates. Next, we construct adjusted scores eliminating the gender gap in priority subjects. Then, we use the adjusted scores, taking the total number of available slots per major as given, to simulate how admission rates would behave in a setting where

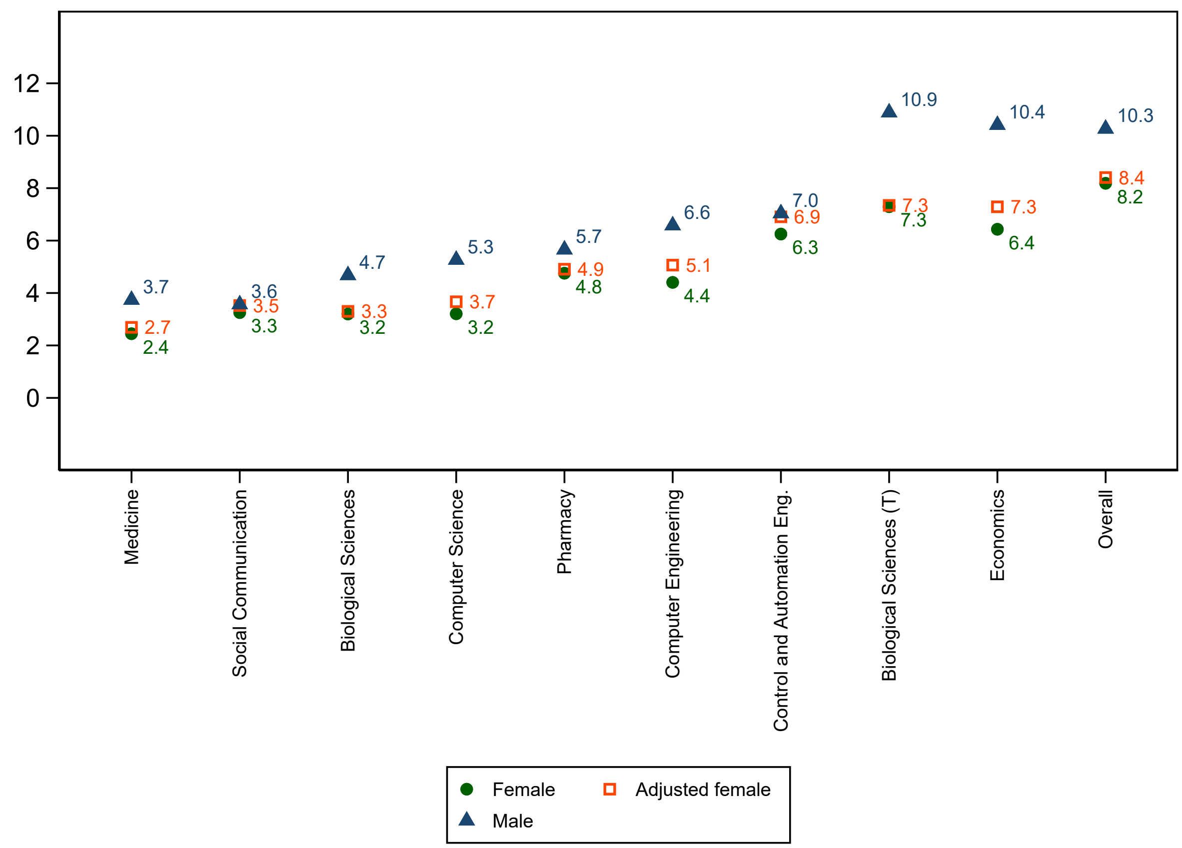
[23] Said differently, estimating the model in difference while allowing the improvement between P_1 and P_2 to depend on P_1 performance yields the same estimates for our coefficients of interest in Table 3 (in specifications where we control for P_1).

[24] Note that the standard deviation of the difference between the normalized P_2 and P_1 is slightly below 1 (0.959). Therefore, in terms of standard deviations, our estimates are even closer than what Table A.5 suggests.

there are no gender differences in reaction to higher stakes.

Figure 1 presents the actual admission rates for male and female applicants, considering all applicants registered in the admission exam. Overall, we find a modest impact on admission, corresponding to 2.4% of the actual admission rates. However, the effect is large for competitive majors, with lower-than-average admission rates. In all of these majors, men have higher (unconditional) admission rates than females, as shown in Figure 1. Eliminating the gender performance gap would have a sizable impact on female applicants' admission rates, as shown by the adjusted rates. Indeed, the adjusted admission rates are around 10–15% larger than the actual admission rates in majors such as medicine, economics, and engineering majors.

Figure 1: Admission Rates with Actual and Adjusted scores



Notes: The figure presents actual admission rates for men and women, considering all applicants registered for the admission exam (and not only Phase 2 applicants). We show the overall admission rate and by major, focusing on majors characterized by lower-than-average admission rates. The adjusted female admission rate imputes the coefficients for 'Female × Priority' and 'Female × Priority × ENEM' for female applicants' priority subjects. The specification is analogous to column (5) in Table 5, but where the response variable is the raw score. We group medicine UNICAMP and FAMERP, as these majors are considered jointly in the admission process.

Thus, our estimated gender performance gap is sufficiently large to affect admission to competitive majors in a meaningful way.

7 Potential Channels

This section inquires into potential channels that could explain our main findings. The exam structure and its competitive nature rule out a few potential channels like males not caring about non-priority subjects (admission is highly competitive), applicants' reaction to their performance in other subjects (applicants do not know their relative performance or rank before the end of the admission process), or a change in the competition level (applicants compete against the same candidates throughout Phase 2). However, other channels could drive the results presented in Table 3.

To investigate some of the potential channels through which our findings emerge, we leverage the richness of our data and analyze gender performance differences across exam subjects and at the exam-question level. Recall that, for each applicant-subject observation, we observe their performance in each of the 12 subject-specific questions. Such detailed information allows us to dig deeper into the potential sources of the gender difference in performance across priority and non-priority subjects.

First, we present suggestive evidence that gender differences in confidence can partially explain our main results. Second, we examine the plausibility of gender disparities in exam strategies as an explanation. Lastly, we discuss mechanisms we can rule out, such as gender gaps in performance in difficult questions, time management, and information.

7.1 Confidence

For the 2001 and 2002 admission years, we can separate omitted questions from attempted questions that received a score of zero.[25] Table 6 (column (1)) suggests that females leave significantly more questions blank when stakes increase (especially when we compare the estimate to the average number of omitted questions). The results are reversed, in column (2), as increased stakes lower the number of zeros of females relative to males. Treating omitted questions as zeros results in a negligible effect of increased stakes on the gender gap in zeros (column (3)).

These results suggest that women and men use different strategies when facing questions they are uncertain about the correct answer. Apparently, women are more likely to shy away from answering questions they are uncertain about the answer. The previous literature

[25] Since there are 12 questions per subject, the number of omissions and zeros varies between 0 and 12.

Table 6: Priority Subjects and Number of Omitted and Zero Score Questions (2001-2002)

	Omissions	Zero scores	Zero + Omissions
	(1)	(2)	(3)
Priority	-0.516***	-0.565***	-1.080***
	(0.017)	(0.020)	(0.024)
Female × Priority	0.087***	-0.078***	0.008
	(0.019)	(0.022)	(0.026)
Mean dependent variable	0.77	2.09	2.85
Std.dev dependent variable	1.55	2.10	2.74
Number of observations	120,690	120,690	120,690
Number of applicants	20,115	20,115	20,115
Subject FE	Yes	Yes	Yes
Subject-gender FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Phase 1 scores	Yes	Yes	Yes
ENEM × Priority	Yes	Yes	Yes
Phase 1 scores × Priority	Yes	Yes	Yes

Notes: The dependent variables are the number of questions omitted by an applicant in a given P_2 subject (column (1)), the number of zeros obtained by an applicant in a given P_2 subject in attempted questions (exclude omissions) (column (2)) and the overall number of zeros obtained by an applicant in a given P_2 subject (attempted questions with a zero score plus omitted questions) (column (3)). The sample is restricted to 2001-2002. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

documents that men tend to be more overconfident than women (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Bertrand, 2011), especially in domains perceived as male-typed (Beyer, 1990; Beyer and Bowden, 1997). Thus, this omission pattern could relate to gender differences in confidence in their ability to answer a question correctly (Bucher-Koenen et al., 2021; Saygin and Atwater, 2021).

Gender stereotypes can lead to differences in self-assessments, leading to self-fulfilling prophecies, as inaccurate beliefs can influence behavior and performance. Bordalo et al. (2019) examines how gender stereotypes shape perceptions regarding abilities in several domains. Although men and women tend to overestimate their ability level, this pattern differs by category. Their evidence suggests gender differences in confidence in female and male-typed domains but negligible gaps in gender-neutral categories. The stereotypes distort women's beliefs in fields with larger previous gender differences in performance. Moreover, Exley and Kessler (2022) verifies gender discrepancies in self-evaluations in male-typed tasks even when they privately assess their performance (instead of sharing with potential employers) — where there are no incentives for self-promotion. They observe that the gender gap in self-assessment closes in a female-typed task. Coffman (2014) finds that women are less willing to answer questions in fields perceived as male-typed due to lower confidence levels. Conditional on ability, she finds that a higher 'maleness' of a domain leads to a lower female willingness to contribute her ideas to a group.

To understand if gender stereotypes help explain the response pattern we observe in Table 6, we present the heterogeneous impacts by subject of higher rewards on the number of omitted questions in Table 7. In our preferred specification (column (6)), we verify that women omit relatively more questions when the stakes are higher in geography, mathematics, and physics. At the same time, there are no gender differences in omissions in biology, chemistry, and history. Note that these results are in line with our findings presented in Table 4 (except for history), where we see that the impact of increased stakes on the gender performance gap is stronger for geography, history, mathematics and physics. Table 1 shows that a higher proportion of women select biology, chemistry, and history as priority subjects, while a larger share of men chooses mathematics, physics, and geography. Overall, our findings suggest that the omission pattern in priority subjects could relate to confidence, especially in the presence of stereotyped beliefs, with larger gender gaps in male-dominated subjects.[26]

If females had tried (e.g., guessed an answer) some of the questions they omitted and some of their answers translated in non-zero scores (without affecting their performance on other questions), they could have improved their performance relative to males. Given that the minimum possible score is 0, guessing more on questions can only improve their performance unless time constraints are binding, which does not seem to be the case, as we document in subsection 7.3.

To investigate how much of the gender gap in priority subjects can be explained by the decision to skip questions, we employ Item Response Theory (graded response model) to estimate each applicant's predicted score in an omitted question based on their ability in a given subject and on the question's difficulty level.[27] We present the results of this exercise in Table A.8 using the 2001-2002 sample. As we can observe in Table A.7 (analogous to

[26]As an alternative measure of the degree of the maleness of a priority subject, we verify the share of men among applicants (2001-2002) that selected each priority subject. The overall proportion of male applicants is 51%. Among the analyzed subjects, biology, history, and chemistry present the lowest male shares: 35%, 37% and 40% of the candidates that choose these priority subjects are men. On the other hand, the subjects that display the highest degree of maleness are physics (73%), mathematics (72%), and geography (61%).

[27]One could argue that using the predicted score based on ability will overestimate the score on would have obtained on a question since they decided to omit that question. However, the grade response model might also underestimate the predicted score as it does not consider which subjects are priorities when estimating the applicant's ability. More specifically, the model computes a subject-specific ability using only the performance information in that specific subject. Since we find that women perform worse than expected on priority subjects, the estimated ability of female applicants will be underestimated for these subjects.

Table 7: Number of Omitted Questions (2001-2002): Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: # omitted questions in Phase 2 per subject						
Female	0.117***	0.043**				
	(0.016)	(0.018)				
Priority	-0.499***	-0.178***	0.243***	0.253***	0.215***	0.205***
	(0.014)	(0.015)	(0.020)	(0.020)	(0.022)	(0.023)
Female × Priority	-0.116***	-0.042*	-0.020	-0.006	-0.017	-0.010
	(0.019)	(0.023)	(0.029)	(0.029)	(0.030)	(0.029)
Female × Priority × Chemistry	0.038	0.105**	0.034	0.007	0.012	0.003
	(0.028)	(0.041)	(0.040)	(0.040)	(0.040)	(0.040)
Female × Priority × Geography	0.031	0.076	0.351**	0.340*	0.380**	0.371**
	(0.046)	(0.051)	(0.179)	(0.177)	(0.179)	(0.178)
Female × Priority × History	-0.011	-0.018	-0.049	-0.055	-0.085	-0.089
	(0.026)	(0.037)	(0.064)	(0.063)	(0.064)	(0.064)
Female × Priority × Mathematics	-0.100***	0.249***	0.186***	0.175***	0.160***	0.151**
	(0.037)	(0.066)	(0.061)	(0.061)	(0.061)	(0.061)
Female × Priority × Physics	0.131***	0.273***	0.180***	0.141**	0.130**	0.123**
	(0.041)	(0.062)	(0.057)	(0.056)	(0.056)	(0.056)
Priority × Chemistry	0.192***	-0.393***	-0.401***	-0.417***	-0.409***	-0.402***
	(0.022)	(0.029)	(0.028)	(0.028)	(0.028)	(0.028)
Priority × Geography	-0.327***	-0.182***	-1.533***	-1.459***	-1.534***	-1.523***
	(0.036)	(0.038)	(0.117)	(0.114)	(0.116)	(0.116)
Priority × History	-0.119***	-0.103***	-1.019***	-0.964***	-0.976***	-0.967***
	(0.019)	(0.023)	(0.045)	(0.044)	(0.045)	(0.045)
Priority × Mathematics	0.986***	-0.572***	-1.038***	-1.008***	-1.021***	-1.004***
	(0.022)	(0.048)	(0.045)	(0.045)	(0.045)	(0.045)
Priority × Physics	0.590***	-0.677***	-1.011***	-0.937***	-0.941***	-0.931***
	(0.021)	(0.040)	(0.039)	(0.038)	(0.038)	(0.038)
ENEM	-0.317***	-0.318***				
	(0.009)	(0.009)				
Mean dependent variable	0.77					
Std.dev dependent variable	1.55					
Number of observations	120,690	120,690	120,690	120,690	120,690	120,690
Number of applicants	20,115	20,115	20,115	20,115	20,115	20,115
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variables are the number of questions omitted by an applicant in a given P_2 subject. The sample is restricted to 2001-2002. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3 for the 2001-2002 sample), women's relative performance decreases by 4.8% of a standard deviation as stakes increase. If the applicants answered all questions and achieved their predicted score instead of skipping items, the change in the gender gap in priority subjects would be 4.3% (a 10% reduction). Thus, confidence seems to play a role, especially in male-stereotyped subjects. However, it cannot account for a considerable proportion of the observed gender gap in reaction to higher rewards.

7.2 Exam Strategy

Another potential explanation for our findings is that females adopt different exam strategies relative to males in priority versus non-priority subjects. For instance, females may spread their effort more equally across and within subjects in Phase 2, while males focus on priority subjects. If that were the case, we expect the variation in performance across subjects to be smaller for females once we consider their average performance (and ability). To investigate this possibility, we first compute the coefficient of variation for each applicant over all the eight P_2 subject scores. We then regress the applicant's coefficient of variation on her ENEM score, a female dummy, and its interaction with the ENEM score.

Conditional on the ENEM performance, female applicants seem to equalize their effort across subjects relatively more than male applicants, incurring a lower coefficient of variation, as shown in Table 8. Moreover, the correlation is stronger (more negative) and highly statistically significant for high-performing women, the group with the largest gender performance gap in priority subjects, as shown in Table 5.

Interestingly, women also exhibit lower performance variation within subjects than men in priority subjects. To account for each question's difficulty level and the applicant's subject-specific ability, we again use the Item Response Theory (graded response model) to calculate each question's predicted score. Then we compute the difference between actual and predicted scores for each question, the 'IRT residuals.' Figure 2 presents the standard deviation distribution of questions' IRT residuals. While the distribution is similar in non-priority subjects for both genders, male applicants show more variation than female applicants in priority subjects relative to their respective predicted scores. Thus, we observe a similar pattern across and within subjects, with female applicants having less performance variation than males, especially in priority subjects. Note that we cannot distinguish whether these different strategies occur during the exam or its preparation, as both would have similar consequences in terms of performance variation.

Table 8: Coefficient of Variation across Subjects

	(1)	(2)	(3)
Dependent variable: Phase 2 score coefficient of variation			
Female	0.030***	-0.006***	-0.006***
	(0.002)	(0.001)	(0.001)
Norm. ENEM scores		-0.084***	-0.077***
		(0.001)	(0.001)
Female × Norm. ENEM scores			-0.014***
			(0.002)
Mean of dependent variable	0.33		
Std.dev dependent variable	0.16		
Number of applicants	42,462	42,462	42,462
Year dummies	Yes	Yes	Yes

Notes: The dependent variable is the coefficient of variation across subject scores. We include foreign language and Portuguese scores to calculate the Phase 2 coefficient of variation. ENEM scores are normalized to mean 0 and standard deviation 1 for each year. Heteroskedasticity-robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

7.3 Alternative Explanations

Performance on Difficult Questions

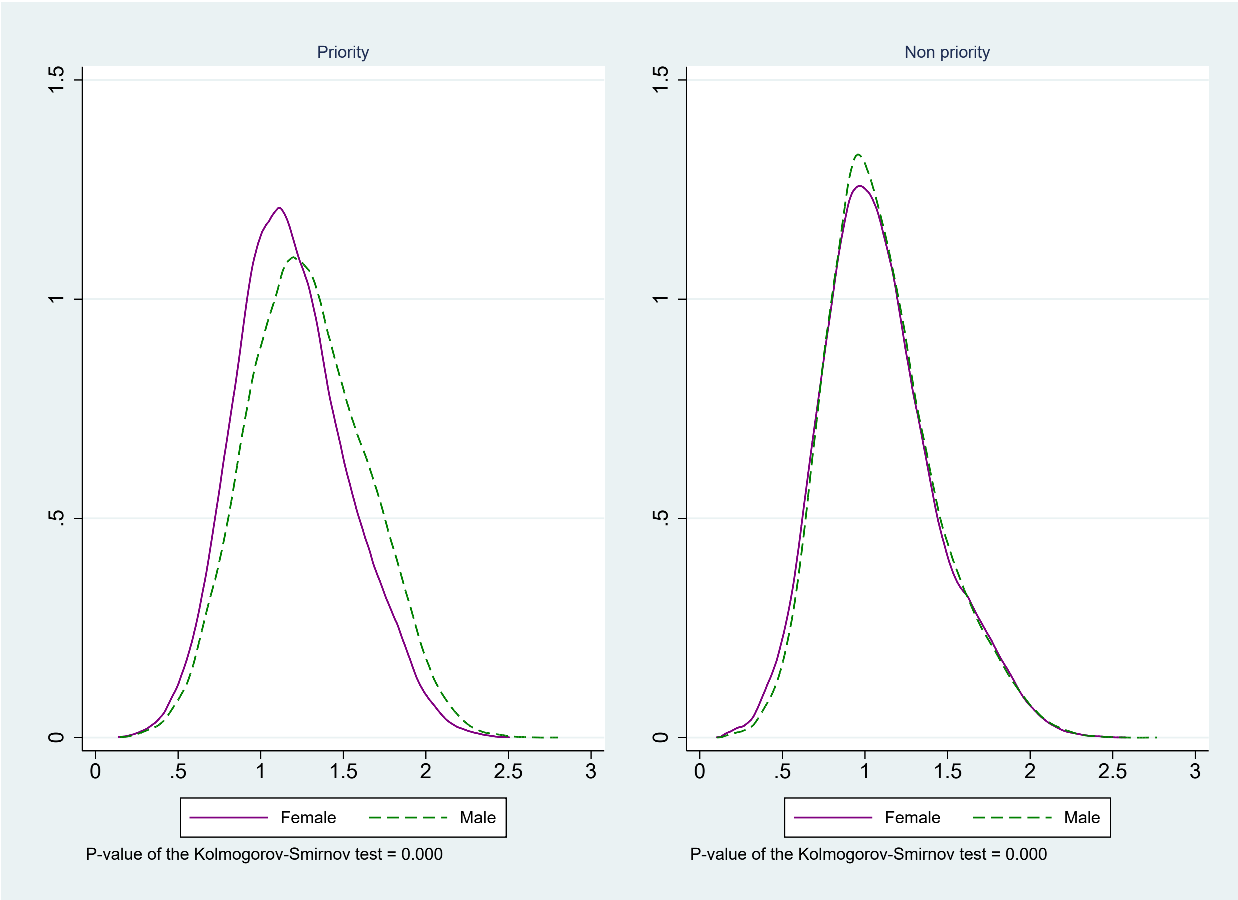
Given that we observe that the gender performance gap in priority subjects increases with ability, we could suspect that women do relatively worse on more difficult questions (assuming that most applicants answer the easier questions correctly). In order to explain our findings, such a pattern would have to be more pronounced in priority subjects — something that is possible if male applicants focus their P_2 exam on priority subjects.

To investigate this possibility, we follow Iriberri and Rey-Biel (2019) and classify each question as being *difficult* or not. We begin by computing the average performance on each question.[28] Then, we classify a question in a specific subject as *difficult* if the average performance on this question is below the median of question average scores for that subject.[29] We interact this dummy variable with 'Female × Priority' to investigate whether the questions' difficulty could exacerbate females' reaction to priority subjects. In this exercise, our dependent variable is the question's raw score, so we have twelve observations per subject

[28] We have run an alternative specification where we define the difficulty level by gender. Results are similar and available in the Online Appendix Table O.7.

[29] As robustness checks, we have also looked at different definitions for the difficulty level. In particular, we classify a question as *very difficult* if the average score on this question is below the bottom quartile. Finally, we have also defined a question as *most difficult* if it is the question with the lowest average performance (within the subject). Our results remain qualitatively similar and are available in Online Appendix Tables O.8 and O.9.

Figure 2: Standard Deviation of Question's IRT Residuals (Score - Predicited Score)



and 72 per applicant.

Table 9 suggests the opposite: While there is a large change in the gender performance gap in reaction to increasing rewards (e.g., -0.156 points or around 9% of a standard deviation in specification (6)) on easier questions, it is not the case when priority questions are difficult questions. If anything, the gender gap changes slightly in favor of women. For example, in specification (6), the estimated change in the gender gap on difficult priority-subject questions is +0.026 (0.182-0.156). While this change in the gender gap is small, it is statistically significant at the 10% level in all specifications.

Therefore, our main findings (in Table 3) are not driven by females' relative performance on difficult priority-subject questions. Women's relative gains on difficult priority-subject questions also suggest ruling out gender differences in applicants' ability levels to explain our main results. Indeed, the estimated impact of increasing stakes on the gender performance gap comes mainly from their reaction to easy priority-subject questions, i.e., those that most applicants answer correctly.

Table 9: Priority Subjects, Difficult Questions and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.159***	-0.049***				
	(0.006)	(0.007)				
Priority	0.590***	0.694***	0.656***	0.604***	0.613***	0.622***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Female × Priority	-0.131***	-0.161***	-0.169***	-0.166***	-0.157***	-0.156***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Female × Priority × Difficult question	0.182***	0.182***	0.182***	0.182***	0.182***	0.182***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Priority × Difficult question	-0.358***	-0.358***	-0.358***	-0.358***	-0.358***	-0.358***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Difficult question	-1.149***	-1.149***	-1.149***	-1.149***	-1.149***	-1.149***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Female × Difficult question	-0.038***	-0.038***	-0.038***	-0.038***	-0.038***	-0.038***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ENEM	0.444***	0.442***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev dependent variable	1.65					
# observations	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264
# applicants	42,462	42,462	42,462	42,462	42,462	42,462
Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the question's raw score (ranging from 0 to 5 points). 'Difficult question' is a dummy variable equal to 1 if *difficult* if the average performance on this question is below the median of question average scores for that subject. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Mental Fatigue and Time Management

Two related factors that could drive our results are mental fatigue and time management. While applicants are not required to answer questions in order of appearance (and we cannot identify the exact order in which applicants answer them), Figure B.1 suggests that they answer questions in the order presented to them. The proportions of zeros and omissions increase according to the question order. Mental fatigue and a lack of proper time management could be behind this pattern. We investigate whether these factors could explain our findings by first splitting each subject exam into three parts: *early* questions (questions 1 to 4), *mid-exam* questions (questions 5 to 8), and *late* questions (questions 9 to 12). We construct a dependent variable that is, for each subject, the difference between an applicant's average performance on early questions minus their average performance on late questions. This dependent variable allows us to control for applicants' subject-specific performance and will inform us whether they perform relatively worse on late questions.

Table 10: Priority Subjects and Within-Exam Performance (Early vs. Late Questions)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Average score first 4 questions - Average score last 4 questions						
Female	0.133***	-0.060***				
	(0.006)	(0.011)				
Priority	0.812***	0.014*	0.019***	0.011	0.045***	0.045***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.010)
Female × Priority	-0.663***	-0.037***	-0.046***	-0.043***	-0.035***	-0.032***
	(0.014)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
ENEM	0.043***	0.054***				
	(0.002)	(0.002)				
Mean dependent variable	0.52					
Std.dev dependent variable	1.30					
Number of observations	254,772	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the average score in early questions (1 to 4) minus the average score in the late questions (9 to 12) for each subject in Phase 2. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and the standard deviation is 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 10 presents results from estimating the same regressions presented in Table 3, but where the outcome variable is our early/late question performance difference. The average of our outcome variable is 0.52, indicating that applicants indeed perform worse on later questions. However, Table 10 suggests that while males' performance worsens more when facing priority subjects (as we move to questions toward the end of the exam), females' relative performance remains unchanged (focusing on our preferred specification in column (6)) whether they face a priority subject or not. The sum of the 'Priority' and 'Female × Priority' parameter estimates is close to zero and statistically insignificant. These findings suggest that mental fatigue and time management are not driving our results.

Negative Shocks

Another potential explanation for the gender gap is females' reaction to 'negative shocks.' Anaya et al. (2022) show that the random assignment of difficult questions at the beginning of an exam increases the likelihood of abandoning the test and the number of omitted questions, reducing the number of correct answers.

However, in our setting, we find mixed evidence that females' performance in priority subjects decreases more than males when facing negative negative shocks. We use the question's difficulty combined with its order in the exam to evaluate if female and male applicants react differently to negative shocks. We define (for each subject-year) an event as negative shock if a *difficult* (i.e., with average performance below the median of question average scores), *very difficult* (i.e., average score below the bottom quartile) or *most difficult* (i.e., with the lowest average performance) question was among the first four questions. We present the results from this exercise in Table A.6. Columns (1), (2), and (3) present results when we define a negative shock as the early arrival of a difficult, very difficult, and most difficult question, respectively. We can see that the impact of these shocks on the gender performance gap is sensitive to their definition. Focusing on the triple-interaction terms, we would conclude that females priority-subject performance is *more* affected by early *difficult* questions, is *equally* affected by early *very difficult* questions, and *less* affected by early *most difficult* questions than males. The larger the shock on priority subjects, the better females cope with these shocks compared to males. Thus, we conclude that gender differences in reaction to negative shocks cannot be behind our results.

Information

Gender differences in information about the UNICAMP admission exam's rules could drive our impacts. Suppose male applicants have better access to information than females.

Women could, in that case, not focus on priority subjects simply because they are unaware of the unequal weighting scheme of P_2, not because they react differently to increased rewards. To investigate this hypothesis, we estimate the main regression for different subsamples that potentially have distinct information levels about the admission exam. First, we split the sample into applicants taking the UNICAMP exam for the first time and those who have taken the exam in previous years. Candidates that have applied to UNICAMP before probably know how the university calculates the final admission score. Next, we examine if the results depend on whether the applicant did a preparatory course. Applicants can enroll in preparatory classes during or after high school to better prepare for university admission exams. Also, UNICAMP is located in Campinas city, and nearby schools may specialize in the UNICAMP admission exam. Thus, we compare the effects for students who attended schools in the Campinas metropolitan region with students from other cities. We also investigate heterogeneous impacts by school type (private or public) and parental educational level (university-educated parents vs. lower degrees). In all subsamples, we observe the same pattern of reduced women's performance in priority subjects, and the coefficients' sizes are also similar.[30]

8 Conclusion

This paper provides evidence that females and males react differently to increased rewards in a natural environment. We observe men and women taking identical exams but with varying stakes within exams. Our setting, combined with the richness of our data, allows us to flexibly control for applicants' major-choice self-selection issues through applicant fixed effects combined with multiple subject-specific ability measures – something that data rarely permits.

Our findings indicate that increasing rewards decreases females' performance relative to males, and this decrease is more pronounced for higher-ability applicants. This greater gender gap in reaction to stakes for the high-achieving candidates results in lower admission rates, particularly in competitive majors such as medicine, engineering, and economics.

After investigating applicants' performance in more detail (at the exam question level), we rule out gender differences in reaction to difficult questions, mental fatigue, and time-

[30]Results are available in the Online Appendix. Table O.10 presents the results comparing applicants taking the exam for the first time with other applicants. Table O.11 displays the results for applicants who enrolled in preparatory courses (compared with candidates who did not enroll). Table O.12 shows the regressions for applicants that attended school in the Campinas Metropolitan region versus other cities. Table O.13 splits our sample by school type: public and private high schools. Lastly, Table O.14 compares students whose parents have a university degree with students with lower parental schooling levels.

management issues as potential mechanisms. If anything, as rewards increase, female performance increases more relative to males on difficult questions and questions that appear late in the exam. Besides, gender differences in selection into priority subjects, reaction to negative shocks, or knowledge about UNICAMP rules cannot explain our result.

Our evidence suggests that women and men adopt distinct approaches when uncertain about a question's correct answer. In higher-stakes exams, females are more likely to omit questions, while males are more likely to try to answer them and obtain zero scores. Interestingly, the omission pattern emerges in male-typed subjects, suggesting the influence of gender-stereotyped beliefs. Also, we observe that women tend to spread their effort more equally across subjects and within priority subject exams. Taken together, the pieces of evidence suggest that gender differences in confidence and exam strategy can explain our findings.

References

Anaya, L., Iriberri, N., Rey-Biel, P., and Zamarro, G. (2022). Understanding performance in test taking: The role of question difficulty order. *Economics of Education Review*, 90:102293.

Arenas, A. and Calsamiglia, C. (2022). Gender differences in high-stakes performance and college admission policies. IZA Discussion Papers 15550, Institute of Labor Economics (IZA).

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

Babcock, L., Recalde, M. P., and Vesterlund, L. (2017a). Gender differences in the allocation of low-promotability tasks: The role of backlash. *American Economic Review*, 107(5):131–135.

Babcock, L., Recalde, M. P., Vesterlund, L., and Weingart, L. (2017b). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, 107(3):714–747.

Bertrand, M. (2011). New perspectives on gender. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4, chapter 17, pages 1543–1590. Elsevier.

Bertrand, M. (2018). Coase lecture – The glass ceiling. *Economica*, 85(338):205–231.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5):960–970.

Beyer, S. and Bowden, E. M. (1997). Gender differences in seff-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23(2):157–172.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2019). Beliefs about gender. *American Economic Review*, 109(3):739–773.

Bucher-Koenen, T., Alessie, R. J., Lusardi, A., and Van Rooij, M. (2021). Fearless woman: Financial literacy and stock market participation. NBER Working Paper 28723, National Bureau of Economic Research.

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

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

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2):448–74.

Dahl, G. B. (2002). Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica*, 70(6):2367–2420.

Ellison, G. and Swanson, A. (2010). The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, 24(2):109–128.

Ellison, G. and Swanson, A. (forthcoming). Dynamics of the gender gap in high math achievement. *Journal of Human Resources*.

Estevan, F., Gall, T., and Morin, L.-P. (2019a). On the road to social mobility? Affirmative action and major choice. Unpublished Manuscript.

Estevan, F., Gall, T., and Morin, L.-P. (2019b). Redistribution without distortion: Evidence from an affirmative action program at a large Brazilian university. *Economic Journal*, 129(619):1182–1220.

Exley, C. L. and Kessler, J. B. (2022). The gender gap in self-promotion. *Quarterly Journal of Economics*, 137(3):1345–1381.

Gneezy, U., Niederle, M., and Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics*, 118(3):1049–1074.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics*, 88(3):389–432.

Heckman, J. J. and Vytlacil, E. (1998). Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling. *Journal of Human Resources*, 33(4):974–987.

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

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.

Niederle, M. and Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2):129–144.

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

Saygin, P. O. and Atwater, A. (2021). Gender differences in leaving questions blank on high-stakes standardized tests. *Economics of Education Review*, 84:102162.

Schlosser, A., Neeman, Z., and Attali, Y. (2019). Differential performance in high vs. low stakes tests: Evidence from the GRE test. *Economic Journal*, 129(623):2916–2948.

Wooldridge, J. M. (2005). Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 87(2):385–390.

Yoon, T., Keller, M. L., De-Lap, B. S., Harkins, A., Lepers, R., and Hunter, S. K. (2009). Sex differences in response to cognitive stress during a fatiguing contraction. *Journal of Applied Physiology*, 107(5):1486–1496.

Appendix A – Additional Tables

Table A.1: Priority Subjects by Program

Major	Port.	Bio.	Chem.	Hist.	Geo.	Math.	Phys.	Cutoff	Prop.Female	Prop. Applied
Medicine Unicamp		✓	✓					617	0.539	0.079
Medicine Famerp		✓	✓					603	0.551	0.044
Computer Engineering						✓	✓	584	0.100	0.062
Electrical Engineering						✓	✓	584	0.121	0.049
Control and Automation Engineering (Eve.)						✓	✓	577	0.083	0.032
Computer Science (Eve.)						✓	✓	575	0.123	0.030
Economics				✓		✓		574	0.388	0.047
Electrical Engineering (Eve.)						✓	✓	571	0.075	0.014
Social Communication (Media Studies)				✓		✓		565	0.560	0.005
Pharmacy		✓	✓					560	0.769	0.007
Economics (Eve.)				✓		✓		557	0.334	0.021
Chemical Engineering			✓			✓		557	0.445	0.037
Biological Sciences		✓						554	0.614	0.030
Food Engineering						✓	✓	551	0.759	0.042
Mechanical Engineering						✓	✓	544	0.075	0.054
Chemical Engineering (Eve.)			✓			✓		538	0.364	0.010
Food Engineering (Eve.)						✓	✓	514	0.717	0.012
History	✓			✓				514	0.524	0.017
Social Sciences	✓			✓				509	0.558	0.021
Language Studies (TB)	✓			✓				499	0.770	0.010
Civil Engineering						✓	✓	494	0.262	0.025
Chemistry Technology (Eve.)			✓					491	0.474	0.009
Biological Sciences (T) (Eve.)		✓						490	0.645	0.024
Chemistry			✓					488	0.585	0.024
Social Sciences (Eve.)	✓			✓				487	0.488	0.012
Physics (Eve.)						✓	✓	468	0.152	0.007
Physics, Math, Applied Math and Comp						✓	✓	466	0.298	0.037
Geology and Geography					✓			463	0.405	0.009
Linguistics	✓			✓				451	0.717	0.004
Philosophy	✓							450	0.393	0.006
Geography (Eve.)					✓			449	0.338	0.007
Language Studies (T) (Eve.)	✓			✓				449	0.743	0.007
Chemistry and Physics (T) (Eve.)			✓				✓	438	0.337	0.004
Dentistry		✓						438	0.695	0.034
Pedagogy (T)	✓			✓				423	0.958	0.010
Phonology	✓	✓						418	0.959	0.006
Information Technology (Eve.)						✓		415	0.228	0.010
Pedagogy (T) (Eve.)	✓			✓				415	0.929	0.009
Agricultural Engineering						✓	✓	407	0.313	0.016
Physical Education		✓		✓				407	0.497	0.013
Telecommunications Technology						✓		403	0.210	0.003
Nursing Unicamp		✓						402	0.946	0.012
Physical Education (Eve.)		✓		✓				398	0.385	0.012
Mathematics (T) (Eve.)						✓	✓	391	0.439	0.009
Statistics						✓	✓	377	0.505	0.018
Nursing Famerp		✓						370	0.938	0.015
Information Technology						✓		360	0.277	0.008
Environmental Sanitation Technology (Eve.)						✓		329	0.595	0.017
Construction Technology (Eve.)						✓		325	0.435	0.008
Environmental Sanitation Technology						✓		310	0.660	0.002
Overall								**533**	**0.429**	**1.000**

Notes: The majors with (Eve.) are programs offered in the evening. (T) and (TB) refer to teaching majors (*licenciaturas*) and teaching/bachelor majors, respectively. 'Cutoff' is the final admission score of the student admitted with the lowest score in her first choice of major. 'Prop. Female' is the proportion of applicants to the major that are females, and 'Prop. App.' is the proportion of applicants who chose the major as their first choice (both measured over the 2001-2004 period, after our sample restrictions – main analysis sample). We drop nine majors requiring an aptitude test from the analysis: Architecture and Urban Planning, Arts, Dance, Music Composition, Music Composition and Conducting, Music Conducting, Music Instruments, Popular Music and Scenic Arts.

Table A.2: Probability of Choosing Priority Subjects

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Priority subject dummy					
P1 normalized subject-specific scores	0.073***	0.067***	0.064***	0.065***	0.077***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female × P1 normalized scores	-0.003**	0.010***	0.008***	0.008***	0.010***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Female	-0.028***	-0.027***	0.220***	0.219***	
	(0.001)	(0.001)	(0.004)	(0.004)	
Norm. ENEM scores				-0.003***	
				(0.000)	
Number of observations	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes	Yes
Subject-gender FE	No	No	Yes	Yes	Yes
Individual FE	No	No	No	No	Yes

Notes: The dependent variable is a dummy that equals one if the subject is a priority in Phase 2. Phase 1 subject-specific scores are normalized such that the mean is 0 and standard deviation 1, for each subject-year. ENEM scores are normalized such that the mean is 0 and standard deviation 1, for each year. Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.3: Priority Subjects and Gender Performance Gap (Raw Phase 2 scores)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 subject-specific raw scores						
Female	-2.063***	-0.753***				
	(0.055)	(0.082)				
Priority	4.955***	6.214***	5.720***	4.959***	5.076***	5.316***
	(0.051)	(0.059)	(0.055)	(0.052)	(0.058)	(0.068)
Female × Priority	-0.445***	-0.818***	-0.936***	-0.851***	-0.752***	-0.699***
	(0.076)	(0.081)	(0.076)	(0.072)	(0.074)	(0.074)
ENEM	0.431***	0.428***				
	(0.002)	(0.002)				
Mean dependent variable	25.42					
Stand dev dependent variable	10.61					
Number of observations	254,772	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462	42,462
Year FE	Yes	Yes	No	No	No	No
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: We use raw ENEM, Phase 1 and Phase 2 scores in the regressions. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' ENEM score minus her/his gender-year group's average ENEM. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' P_1 subject score minus her/his gender-year group's average. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.4: Priority Subjects and Gender Performance Gap (Including Portuguese)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.107***	-0.074***				
	(0.006)	(0.008)				
Priority	0.498***	0.570***	0.549***	0.484***	0.495***	0.510***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.065***	-0.071***	-0.067***	-0.065***	-0.058***	-0.057***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
ENEM	0.535***	0.534***				
	(0.003)	(0.003)				
Number of observations	297,234	297,234	297,234	297,234	297,234	297,234
Number of applicants	42,462	42,462	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Alternative Dependent Variable: Phase 2 - Phase 1 Scores

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Normalized Phase 2 scores - Phase 1 scores					
Female	-0.036***	0.028***			
	(0.005)	(0.009)			
Priority	0.181***	0.222***	0.220***	0.224***	0.419***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Female × Priority	-0.026***	-0.057***	-0.055***	-0.048***	-0.050***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
ENEM	0.157***	0.157***			
	(0.002)	(0.002)			
Number of observations	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes
Phase 1 scores	No	No	No	No	No
ENEM × Priority	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	Yes

Notes: Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.6: 'Negative Shocks': Gender Differences in Phase 2 Normalized Scores

	(1)	(2)	(3)
Dependent variable: Phase 2 normalized subject-specific scores			
Priority	0.443***	0.464***	0.524***
	(0.010)	(0.008)	(0.007)
Female × Priority	-0.032**	-0.046***	-0.061***
	(0.013)	(0.010)	(0.008)
Difficult question among the first 4	-0.040***		
	(0.006)		
Female × difficult in first 4	0.001		
	(0.006)		
Priority × Difficult in first 4	0.084***		
	(0.009)		
Female × Priority × Difficult in first 4	-0.024*		
	(0.014)		
Very difficult question among the first 4		-0.028***	
		(0.006)	
Female × very difficult in first 4		0.013	
		(0.009)	
Priority × Very difficult in first 4		0.088***	
		(0.008)	
Female × Priority × Very difficult in first 4		-0.017	
		(0.013)	
Most difficult question is among the first 4			-0.011**
			(0.005)
Female × Most difficult in first 4			0.015**
			(0.007)
Priority × Most difficult in first 4			-0.033***
			(0.012)
Female × Priority × Most difficult in first 4			0.032**
			(0.016)
# observations	254,772	254,772	254,772
# applicants	42,462	42,462	42,462
Subject FE	Yes	Yes	Yes
Subject-gender FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Phase 1 scores	Yes	Yes	Yes
ENEM × Priority	Yes	Yes	Yes
Phase 1 scores × Priority	Yes	Yes	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The most difficult question is the item, per subject and year, that presented the lowest average score. We classify a question in a specific subject as *difficult* if the average performance on this question is below the median of question average scores for that subject, and a *very difficult* question if the average score is among the lowest 25% in the exam (subject-year). 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.7: Normalized Phase 2 Scores, Omissions = Zero Score (Main Results, 2001-2002 Sample)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.226***	-0.112***				
	(0.008)	(0.011)				
Priority	0.466***	0.559***	0.526***	0.469***	0.491***	0.500***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
Female × Priority	-0.015	-0.058***	-0.064***	-0.062***	-0.049***	-0.048***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
ENEM	0.573***	0.571***				
	(0.004)	(0.004)				
Number of observations	120,690	120,690	120,690	120,690	120,690	120,690
Number of applicants	20,115	20,115	20,115	20,115	20,115	20,115
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The sample is restricted to 2001-2002. P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

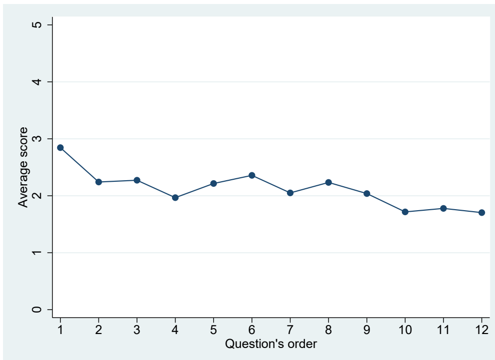
Table A.8: Normalized Phase 2 Scores, Omissions = Predicted IRT Score

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.244***	-0.116***				
	(0.008)	(0.011)				
Priority	0.451***	0.530***	0.502***	0.446***	0.468***	0.477***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
Female × Priority	-0.022**	-0.050***	-0.059***	-0.056***	-0.044***	-0.043***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
ENEM	0.581***	0.580***				
	(0.004)	(0.004)				
Number of observations	120,690	120,690	120,690	120,690	120,690	120,690
Number of applicants	20,115	20,115	20,115	20,115	20,115	20,115
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

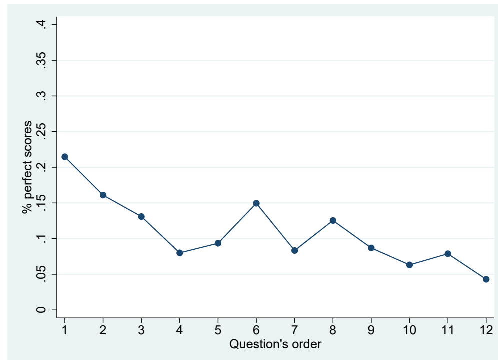
Notes: The sample is restricted to 2001-2002. Each omitted question score is replaced by the applicant's predicted IRT score. P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix B – Additional Figures

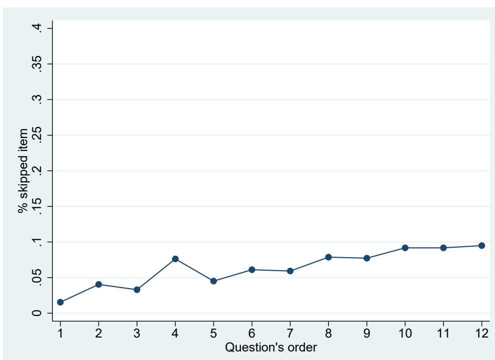
Figure B.1: Performance by Question Order



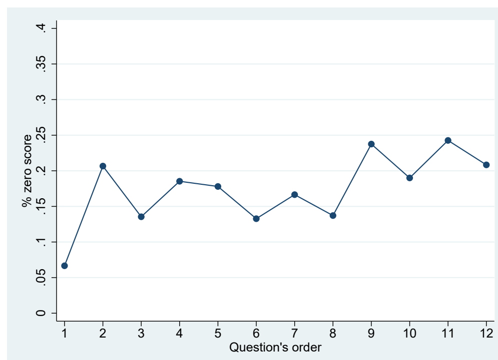
(a) Average score



(b) Share of 'perfect' scores



(c) Share of omissions



(d) Share of zero scores

Notes: The figures present the average applicant performance by the order the question is displayed in the exam (subject-year). Subfigures: (a) presents the question's average raw score; subfigure (b) the percentage of 'perfect' (maximum) scores; (c) the percentage of omitted questions; (d) the percentage of attempted questions that received a zero score. We present the performance focusing on our main estimation sample: 2001–2004 applicants, and six subjects covered in both stages (biology, chemistry, geography, history, mathematics, and physics.) In subfigures (c) and (d), we restrict the sample to 2001–2002, the years in which we can distinguish omissions and attempted questions with zero scores.

Online Appendix (Not for Publication)

Gender Differences in Prioritizing Rewarding Tasks

Bruna Borges, Fernanda Estevan and Louis-Philippe Morin

Table O.1: Priority Subjects and Gender Performance Gap (Excluding Medicine)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.225***	-0.121***				
	(0.006)	(0.008)				
Priority	0.498***	0.668***	0.629***	0.562***	0.569***	0.574***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Female × Priority	-0.022***	-0.107***	-0.092***	-0.091***	-0.087***	-0.086***
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ENEM	0.475***	0.473***				
	(0.003)	(0.003)				
Number of observations	223,458	223,458	223,458	223,458	223,458	223,458
Number of applicants	37,243	37,243	37,243	37,243	37,243	37,243
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: We exclude medicine applicants (UNICAMP and FAMERP). P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.2: Heterogeneity Across Subjects (Excluding Medicine)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.225***	-0.134***				
	(0.006)	(0.009)				
Priority	0.419***	0.453***	0.406***	0.348***	0.366***	0.373***
	(0.017)	(0.018)	(0.015)	(0.014)	(0.014)	(0.015)
Female × Priority	0.172***	0.081***	-0.020	-0.015	-0.019	-0.019
	(0.020)	(0.022)	(0.018)	(0.017)	(0.017)	(0.017)
Female × Priority × Chemistry	-0.141***	0.024	0.073***	0.060**	0.055**	0.055**
	(0.029)	(0.032)	(0.026)	(0.026)	(0.026)	(0.026)
Female × Priority × Geography	-0.192***	-0.160**	-0.103	-0.092	-0.109*	-0.110*
	(0.067)	(0.069)	(0.065)	(0.064)	(0.064)	(0.064)
Female × Priority × History	-0.217***	-0.184***	-0.073***	-0.080***	-0.064**	-0.064**
	(0.027)	(0.031)	(0.027)	(0.026)	(0.026)	(0.026)
Female × Priority × Mathematics	-0.240***	-0.249***	-0.090***	-0.105***	-0.095***	-0.093***
	(0.023)	(0.030)	(0.025)	(0.025)	(0.025)	(0.025)
Female × Priority × Physics	-0.377***	-0.288***	-0.108***	-0.111***	-0.098***	-0.096***
	(0.024)	(0.030)	(0.024)	(0.023)	(0.023)	(0.023)
Priority × Chemistry	0.077***	-0.046*	-0.117***	-0.098***	-0.102***	-0.102***
	(0.023)	(0.025)	(0.020)	(0.020)	(0.020)	(0.020)
Priority × Geography	0.411***	0.265***	0.502***	0.474***	0.499***	0.497***
	(0.046)	(0.047)	(0.044)	(0.043)	(0.043)	(0.043)
Priority × History	0.305***	0.288***	0.520***	0.485***	0.475***	0.475***
	(0.021)	(0.023)	(0.021)	(0.020)	(0.020)	(0.020)
Priority × Mathematics	-0.019	0.264***	0.232***	0.251***	0.239***	0.237***
	(0.018)	(0.024)	(0.020)	(0.019)	(0.019)	(0.019)
Priority × Physics	0.149***	0.311***	0.237***	0.213***	0.195***	0.194***
	(0.019)	(0.022)	(0.018)	(0.017)	(0.017)	(0.017)
ENEM	0.476***	0.471***				
	(0.003)	(0.003)				
Number of observations	223,458	223,458	223,458	223,458	223,458	223,458
Number of applicants	37,243	37,243	37,243	37,243	37,243	37,243
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: We exclude medicine applicants (UNICAMP and FAMERP). Biology is the baseline subject. P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.3: Heterogeneity Across Academic Ability (ENEM × Priority - Quartic Function)

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Female	-0.195***	-0.083***			
	(0.006)	(0.008)			
Priority	0.443***	0.543***	0.564***	0.495***	0.514***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.014**	-0.060***	-0.054***	-0.053***	-0.051***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Priority × ENEM	-0.015**	-0.018**	-0.020***	-0.024***	-0.021***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Female × ENEM	0.019***	0.021***			
	(0.006)	(0.006)			
ENEM	0.521***	0.521***			
	(0.005)	(0.005)			
Priority × ENEM	0.099***	0.099***	0.049***	0.059***	0.070***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Number of observations	254,772	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462	42,462
ENEM × Priority (Quartic)	Yes	Yes	Yes	Yes	Yes
Subject FE	No	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.4: Heterogeneity Across Academic Ability (Phase 1 scores)

	(1)	(2)	(3)	(4)
Dependent variable: Phase 2 normalized subject-specific scores				
Female	-0.193***	-0.056***		
	(0.005)	(0.007)		
Priority	0.392***	0.458***	0.481***	0.499***
	(0.004)	(0.005)	(0.005)	(0.005)
Female × Priority	-0.024***	-0.049***	-0.053***	-0.045***
	(0.006)	(0.007)	(0.007)	(0.007)
Female × Priority × Phase 1 score	0.004	-0.001	-0.016***	-0.015**
	(0.007)	(0.007)	(0.006)	(0.006)
Female × Phase 1 score	-0.011***	-0.009**	-0.017***	-0.018***
	(0.004)	(0.004)	(0.004)	(0.004)
Priority × Phase 1 score	-0.017***	-0.023***	-0.017***	-0.032***
	(0.005)	(0.005)	(0.004)	(0.004)
Phase 1 score	0.363***	0.360***	0.209***	0.213***
	(0.003)	(0.003)	(0.003)	(0.003)
Number of observations	254,772	254,772	254,772	254,772
Number of applicants	42,462	42,462	42,462	42,462
Subject FE	No	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes
ENEM × Priority	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use a quartic function to control for the relative ENEM performance and its interaction with 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.5: Heterogeneity Across Academic Ability (Females)

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Priority	0.477***	0.528***	0.479***	0.421***	0.444***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)
Priority × ENEM	0.040***	0.029***	0.005	0.005	0.020***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
ENEM	0.540***	0.542***			
	(0.004)	(0.004)			
Number of observations	109,332	109,332	109,332	109,332	109,332
Number of applicants	18,222	18,222	18,222	18,222	18,222
Subject FE	No	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.6: Heterogeneity Across Academic Ability (Males)

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Priority	0.481***	0.580***	0.542***	0.475***	0.499***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Priority × ENEM	0.048***	0.041***	0.036***	0.040***	0.052***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
ENEM	0.521***	0.521***			
	(0.005)	(0.005)			
Number of observations	145,440	145,440	145,440	145,440	145,440
Number of applicants	24,240	24,240	24,240	24,240	24,240
Subject FE	No	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. Biology is the baseline subject. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.7: Priority Subjects, Question's Difficulty (by Gender) and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.164***	-0.054***				
	(0.006)	(0.007)				
Priority	0.593***	0.697***	0.658***	0.606***	0.615***	0.625***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Female × Priority	-0.131***	-0.162***	-0.170***	-0.167***	-0.158***	-0.157***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Female × Priority × Difficult question	0.183***	0.183***	0.183***	0.183***	0.183***	0.183***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Priority × Difficult question	-0.363***	-0.363***	-0.363***	-0.363***	-0.363***	-0.363***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Difficult question	-1.151***	-1.151***	-1.151***	-1.151***	-1.151***	-1.151***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Female × Difficult question	-0.028***	-0.028***	-0.028***	-0.028***	-0.028***	-0.028***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ENEM	0.444***	0.442***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev. dependent variable	1.65					
# observations	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264
# applicants	42,462	42,462	42,462	42,462	42,462	42,462
Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the question's raw score (ranging from 0 to 5 points). 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.8: Priority Subjects, Very Difficult Questions and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.177***	-0.067***				
	(0.005)	(0.007)				
Priority	0.506***	0.610***	0.572***	0.520***	0.529***	0.538***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.084***	-0.114***	-0.122***	-0.119***	-0.110***	-0.110***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Priority × Very difficult question	0.176***	0.176***	0.176***	0.176***	0.176***	0.176***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Priority × Very difficult question	-0.380***	-0.380***	-0.380***	-0.380***	-0.380***	-0.380***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Very difficult question	-1.117***	-1.117***	-1.117***	-1.117***	-1.117***	-1.117***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Female × Very difficult question	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
ENEM	0.444***	0.442***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev. dependent variable	1.65					
# observations	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264
# applicants	42,462	42,462	42,462	42,462	42,462	42,462
Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the question's raw score (ranging from 0 to 5 points). 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.9: Priority Subjects, Most Difficult Questions and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.182***	-0.072***				
	(0.005)	(0.007)				
Priority	0.436***	0.540***	0.501***	0.449***	0.458***	0.468***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.047***	-0.078***	-0.086***	-0.083***	-0.074***	-0.073***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Female × Priority × Most difficult question	0.092***	0.092***	0.092***	0.092***	0.092***	0.092***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Priority × Most difficult question	-0.293***	-0.293***	-0.293***	-0.293***	-0.293***	-0.293***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Most difficult question	-1.203***	-1.203***	-1.203***	-1.203***	-1.203***	-1.203***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Female × Most difficult question	0.052***	0.052***	0.052***	0.052***	0.052***	0.052***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
ENEM	0.444***	0.442***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev. dependent variable	1.65					
# observations	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264	3,057,264
# applicants	42,462	42,462	42,462	42,462	42,462	42,462
Order FE	Yes	Yes	Yes	Yes	Yes	Yes
Subject FE	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the question's raw score (ranging from 0 to 5 points). 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.10: Priority Subjects and Gender Performance Gap: First time UNICAMP

	First time UNICAMP						Not first time					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: Phase 2 normalized subject-specific scores												
Female	-0.208***	-0.107***					-0.190***	-0.074***				
	(0.010)	(0.014)					(0.007)	(0.010)				
Priority	0.465***	0.573***	0.534***	0.481***	0.492***	0.499***	0.491***	0.586***	0.549***	0.483***	0.502***	0.524***
	(0.008)	(0.010)	(0.009)	(0.009)	(0.010)	(0.011)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)
Female × Priority	-0.024**	-0.066***	-0.062***	-0.060***	-0.052***	-0.050***	0.006	-0.048***	-0.071***	-0.067***	-0.058***	-0.057***
	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ENEM	0.556***	0.554***					0.535***	0.533***				
	(0.005)	(0.005)					(0.004)	(0.004)				
# observations	78,180	78,180	78,180	78,180	78,180	78,180	174,786	174,786	174,786	174,786	174,786	174,786
# applicants	13,030	13,030	13,030	13,030	13,030	13,030	29,131	29,131	29,131	29,131	29,131	29,131
Subject FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.11: Priority Subjects and Gender Performance Gap: Preparatory course

	Preparatory course						No prep course					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: Phase 2 normalized subject-specific scores												
Female	-0.196***	-0.075***					-0.198***	-0.109***				
	(0.007)	(0.009)					(0.011)	(0.015)				
Priority	0.491***	0.586***	0.550***	0.486***	0.504***	0.524***	0.463***	0.573***	0.531***	0.472***	0.484***	0.497***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)	(0.008)	(0.010)	(0.010)	(0.009)	(0.010)	(0.011)
Female × Priority	0.007	-0.045***	-0.066***	-0.065***	-0.056***	-0.054***	-0.025**	-0.075***	-0.072***	-0.064***	-0.058***	-0.057***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
ENEM	0.538***	0.536***					0.548***	0.546***				
	(0.004)	(0.004)					(0.006)	(0.006)				
# observations	178,530	178,530	178,530	178,530	178,530	178,530	74,580	74,580	74,580	74,580	74,580	74,580
# applicants	29,755	29,755	29,755	29,755	29,755	29,755	12,430	12,430	12,430	12,430	12,430	12,430
Subject FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.12: Priority Subjects and Gender Performance Gap: Campinas Metropolitan Region

	Campinas metropolitan region						Other cities					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: Phase 2 normalized subject-specific scores												
Female	-0.184***	-0.060***					-0.189***	-0.075***				
	(0.012)	(0.017)					(0.007)	(0.009)				
Priority	0.404***	0.590***	0.553***	0.490***	0.506***	0.526***	0.500***	0.570***	0.536***	0.474***	0.489***	0.506***
	(0.010)	(0.012)	(0.011)	(0.011)	(0.012)	(0.014)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)
Female × Priority	0.056***	-0.055***	-0.058***	-0.057***	-0.054***	-0.052***	-0.017**	-0.050***	-0.066***	-0.064***	-0.055***	-0.053***
	(0.014)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ENEM	0.486***	0.483***					0.542***	0.541***				
	(0.006)	(0.006)					(0.004)	(0.004)				
# observations	51,528	51,528	51,528	51,528	51,528	51,528	186,288	186,288	186,288	186,288	186,288	186,288
# applicants	8,588	8,588	8,588	8,588	8,588	8,588	31,048	31,048	31,048	31,048	31,048	31,048
Subject FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.13: Priority Subjects and Gender Performance Gap: Public × Private Schools

	Public school						Private school					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: Phase 2 normalized subject-specific scores												
Female	-0.253***	-0.129***					-0.177***	-0.068***				
	(0.011)	(0.015)					(0.006)	(0.009)				
Priority	0.468***	0.667***	0.623***	0.539***	0.558***	0.572***	0.485***	0.551***	0.519***	0.463***	0.477***	0.496***
	(0.009)	(0.011)	(0.011)	(0.010)	(0.011)	(0.013)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.007)
Female × Priority	0.002	-0.086***	-0.074***	-0.068***	-0.059***	-0.059***	-0.009	-0.047***	-0.063***	-0.062***	-0.054***	-0.052***
	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
ENEM	0.474***	0.470***					0.548***	0.547***				
	(0.005)	(0.005)					(0.004)	(0.004)				
# observations	64,500	64,500	64,500	64,500	64,500	64,500	189,768	189,768	189,768	189,768	189,768	189,768
# applicants	10,750	10,750	10,750	10,750	10,750	10,750	31,628	31,628	31,628	31,628	31,628	31,628
Subject FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.14: Priority Subjects and Gender Performance Gap: At Least One Parent Has a Higher Education Degree

	At least one parent has higher education degree						No parent has higher education degree					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: Phase 2 normalized subject-specific scores												
Female	-0.170***	-0.058***					-0.231***	-0.116***				
	(0.007)	(0.010)					(0.009)	(0.013)				
Priority	0.494***	0.554***	0.521***	0.465***	0.475***	0.492***	0.453***	0.624***	0.583***	0.508***	0.532***	0.550***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)	(0.009)	(0.009)	(0.008)	(0.009)	(0.010)
Female × Priority	-0.019**	-0.047***	-0.061***	-0.060***	-0.053***	-0.050***	0.025**	-0.069***	-0.074***	-0.069***	-0.061***	-0.060***
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
ENEM	0.549***	0.548***					0.503***	0.501***				
	(0.004)	(0.004)					(0.005)	(0.005)				
# observations	163,098	163,098	163,098	163,098	163,098	163,098	89,142	89,142	89,142	89,142	89,142	89,142
# applicants	27,183	27,183	27,183	27,183	27,183	27,183	14,857	14,857	14,857	14,857	14,857	14,857
Subject FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Subject-gender FE	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Individual FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Phase 1 scores	No	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes
ENEM × Priority	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Phase 1 scores × Priority	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Notes: P_2 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. 'ENEM' stands for the applicant's ENEM relative performance, i.e., the applicants' normalized ENEM score minus her/his gender-year group's average normalized ENEM. Individual ENEM scores are first normalized such that the mean is 0 and standard deviation 1 for each year. 'Phase 1 scores' stands for the applicant's subject-specific relative P_1 performance, i.e., the applicants' normalized P_1 subject score minus her/his gender-year group's average. Subject-specific P_1 scores are first normalized such that the mean is zero and the standard deviation is one for each subject-year. We use quartic functions to control for the relative P_1 performance and its interaction with 'Priority,' as well as for the interaction between the relative ENEM performance and 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.