Gender Differences in Prioritizing Rewarding Tasks and Labor Market Outcomes

Bruna Borges* Fernanda Estevan† Louis-Philippe Morin‡

Abstract

We investigate gender performance differences in reaction to task-rewards (weights) variation based on the applicant's major choice in the admission exam from a selective Brazilian university. Our data allows us to control for applicants' major-choice self-selection issues flexibly and to follow applicants in the formal labor market several years after the admission exam. We show that females' performance decreases relative to males' when facing a larger-reward subject due to gender differences in exam strategy. Looking at future labor market outcomes, we find that performing well on priority subjects positively relates to wages. However, our findings cast doubt on whether gender differences in prioritizing rewarding tasks in an exam environment can explain the gender wage gap.

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

*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.

1 Introduction

Despite progress in educational attainment and labor market outcomes, females remain underrepresented in high-earning jobs (Bertrand, 2018). Reaching top positions within occupations typically involves performing well in competitive environments and prioritizing high- rather than low-rewarding endeavors. Recent literature investigates gender differences in non-cognitive skills as one potential explanation for the gender gaps in labor market outcomes.

One strand of this literature focuses on real-world settings, such as exams.[1] There is evidence that females, relative to males, tend to underperform in competitive situations (Jurajda and Münich, 2011; Ors et al., 2013; Morin, 2015; Iriberri and Rey-Biel, 2019) and are more risk-averse (Pekkarinen, 2015). Moreover, the literature suggests that females perform relatively worse than males when stakes are higher (Azmat et al., 2016; Cai et al., 2019; Schlosser et al., 2019).[2] The extent to which these gaps in exam performance can explain gender gaps in the labor market is still an open research question (Blau and Kahn, 2017; Bertrand, 2018).

In this paper, we show that, as found 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 face at some point in their life, and for which the task is common: writing a university admission exam. The richness of our data allows us to move this literature forward in two important ways. First, we investigate potential channels driving this gender gap. Second, and crucially, we link our university data to labor-market data (up to 14 years after students took the exam) to assess whether gender differences in prioritizing rewarding tasks can help explain gender wage gaps.

Specifically, 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. Specifically, UNICAMP applicants must write the admission exam in two phases, nearly two months apart. Both stages are composed of open-ended questions

[1] Alternatively, other studies use laboratory experiments to examine gender differences in several traits, such as competitiveness, overconfidence, and risky behavior (e.g., Gneezy et al., 2003; Niederle and Vesterlund, 2007; Buser and Peter, 2012; Coffman, 2014). Croson and Gneezy (2009) present a literature review of the experimental evidence regarding gender differences in preferences.

[2] Surprisingly, Bandiera et al. (2021) find no gender differences in response to incentives in experimental settings. Using a Bayesian hierarchical model, the authors aggregate evidence from multiple laboratory and field experiments. While the increase in performance pay increases overall performance, they find that women and men are equally responsive to incentives.

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. However, 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 a goal, like attaining a specific final grade, and adjust their effort based on past performance (on midterms, for example). Lastly, although the stakes and competition level may increase between Phase 1 and 2, we are comparing exams with different weights within Phase 2. Therefore, we hold environmental elements constant, such as the competition level and pool of competitors, and examine the differences related to an increase in the exam's reward.

We show that increased rewards affect the gender performance gap in a significant way. Moving from a non-priority to a priority subject reduces females' relative performance by the equivalent of 8% of the within-applicant standard deviation. This effect is larger for higher-ability applicants, potentially affecting admission in competitive majors. Indeed, we simulate admissions without the gender performance gap in priority subjects and find that closing the gap would have a modest effect on majors with lower-than-average admission rates, such as medicine, engineering, and a sizable effect in economics.

Our evidence suggests that gender differences in exam strategy may be behind these 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 males attempt to answer them, but they are also more likely to get a score of zero. Second, females spread their effort more equally across Phase 2 subjects and questions within priority-subject exams.

Interestingly, our results show that applicants who perform better on priority subjects earn more at the formal labor market. Nevertheless, female underperformance in priority subjects does not help explain the gender wage gap in the formal labor market. To show that, we extract residuals from Phase 2 performance regressions (that control for applicant fixed effects) that will capture applicants' 'unexpected' performance in each subject. We then compute, for each individual, the difference between the average residuals in priority and non-priority subjects to capture applicants' relative ability to perform well in priority subjects. Finally, we estimate wage regressions controlling for the ability to perform well in priority subjects to investigate whether it can decrease the estimated gender wage gap. While our 'priority-subject-ability' measure is positively associated with higher wages, its inclusion in our wage regressions does not materially affect the gender gap.

Our work contributes to the literature in important ways. First, we show that the gender performance 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. Secondly, we are examining a real-life scenario characterized by moderate changes in stakes, wherein tasks with negligible or near-zero stakes are non-existent. We believe this setting aligns more consistently with the actual situations in the labor market. Moreover, our detailed data at the exam-question level allow us to investigate differences in exam strategy at the question level, an aspect the previous literature has been mostly unable to exploit. Finally, our data allows us to test if exam performance in rewarding tasks can partially account for differences in future wages up to 14 years after the admission exam. We document that applicants who overperform on priority subjects earn higher annual wages. Still, the gender gaps in exam performance cannot account for the gender wage gaps we observe in the labor market, casting doubt on whether we can always extrapolate gender differences in exam performance beyond the academic environment. This last finding supports the recent literature showing small to moderate impacts of gender gaps in competitive behavior, as measured in laboratory experiments or through online surveys, on labor market outcomes (e.g., Reuben et al., 2015; Buser et al., 2021).

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. Next, we investigate mechanisms that could explain our results in Section 6. In Section 7, we present the impact of the gender gap in terms of university admission and future wages. Finally, Section 8 provides a conclusion.

2 UNICAMP Admission Exam

Each year, individuals applying to UNICAMP write an admission exam. Admission is competitive as only around 10% of applicants are admitted. When registering for the exam (about two months before writing it), candidates rank up to three majors, so admission is major-specific. However, since UNICAMP uses an admission allocation based on the Boston mechanism, most of the successful applicants (90%) are admitted to their first-choice major.[3]

The admission exam is composed of two sequential phases (hereafter referred to as P_1 and P_2, respectively).[4] Applicants write P_1 in November, P_2 in January, and the academic year begins in late February. In both phases, the examinations are identical for all applicants, regardless of their major choice.

In P_1, all applicants answer the same twelve written-answer questions in biology, chemistry, geography, history, mathematics, and physics (i.e., two questions per subject). Importantly, applicants do not have any incentive to perform particularly well on a specific subject as they carry the same weight (5 points each) in the calculation of the P_1 score — the rewards are identical across subjects. Applicants must also write an essay worth 30 points.

To qualify for P_2, applicants' P_1 score must be above a major-specific cutoff. The baseline cutoff score is 50%. However, after P_1 exams are graded, UNICAMP adjusts cutoff scores 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 cutoff score was 84% for medicine and 43% for statistics.

The P_2 exam covers the same subjects as in P_1, plus Portuguese and a foreign language. There are twelve equally-weighted questions for each subject. P_2 is administered over four days, testing two subjects per day. Each day, applicants have four hours to submit their answers for both subjects. 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 question average scores decrease with their order, suggesting that applicants tend to answer questions following the exam's layout.

As in P_1, P_2 exams are composed of written-answer questions and are identical for all applicants. However, in contrast to P_1, applicants' subject scores are not equally weighted in their P_2 score calculation. Instead, depending on the applicant's major choice, one or two subjects are considered *priority subjects* and receive a weight of two (instead of one) in the final score calculation.

Table 1 presents the list of majors in our sample along with their priority subjects,

[3]See the Online Appendix O.1 and Estevan et al. (2019) for more details on the admission procedure.

[4]Some majors, like Performing Arts, also require an aptitude test. We drop these majors from the analysis since their exam weighting schemes differ from the other majors.

admission cutoffs, the proportion of female applicants, and the percentage of candidates applying to this major in our sample of analysis (measured over the 2001-2004 period).[5] 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 majors; mathematics and physics for engineering majors), which are more popular with a particular gender (e.g., life-science programs are usually more prevalent among females). However, there remains significant and non-trivial variation in priority subjects and female proportions across majors. For example, chemical engineering, for which 45% of the applicants are females, has chemistry and mathematics as priority subjects. Economics has history and mathematics as priorities, and around 39% of its applicants are females. Finally, 76% of food engineering applicants, who have mathematics and physics as priorities, are females.

An applicant's final P_2 score is the weighted average of her normalized: 1) P_1 score, with a weight of two; 2) P_2 priority-subject scores, each with a weight of two, and; 3) P_2 non-priority subject scores, each with a weight of one. Thus, for a typical major with two priority subjects, priority subjects count for one-third of applicants' final admission scores. Once final scores are computed, UNICAMP ranks candidates within each major following the Boston mechanism and makes offers based on the number of available slots.

To summarize, the admission exam is designed such that two applicants writing identical Phase 2 exams could face different incentives regarding which questions/subjects should be prioritized given the exam rewards (weights) structure. Since less than 30% of candidates makes it to P_2, applicants cannot slack off in P_1, even if it only counts for a small portion of the final score. Importantly, applicants who advanced to P_2 do not know their normalized P_1 score when they write P_2. Since 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. Finally, applicants are not informed about their performance on individual P_2 subjects before receiving their final score, precluding effort adjustments across P_2 exam days.

3 Data

We work with data from UNICAMP's admission office (*Comissão Permanente para os Vestibulares*, COMVEST) matched with employer-employee data from *Relação Anual de Informações Sociais* (RAIS).

[5]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.

Table 1: Priority Subjects by Major

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	0.045
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	0.031
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	0.011
Food Engineering (Eve.)						✓	✓	514	0.717	0.012
History	✓			✓				514	0.524	0.018
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	0.010
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	0.038
Geology and Geography					✓			463	0.405	0.009
Linguistics	✓			✓				451	0.717	0.005
Philosophy	✓							450	0.436	0.001
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	0.010
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: Majors with (Eve.) are offered in the evening. (T) and (TB) are teaching (*licenciaturas*) and teaching/bachelor majors, respectively. 'Cutoff' is the final score of the student admitted with the lowest score in her first-choice major. The proportion of females (Prop. Female), and the proportion of applicants who chose the major as their first choice (Prop. App.) are measured over the 2001-2004 period, after our sample restrictions. We exclude Architecture and Urban Planning, Arts, Dance, Music Composition, Music Composition and Conducting, Music Conducting, Music Instruments, Popular Music and Scenic Arts as they require aptitude tests.

The COMVEST dataset contains individual-level information on all 2001-2004 UNICAMP applicants.[6] We focus on the six subjects covered in both P_1 and P_2, i.e., biology, chemistry, history, geography, mathematics, and physics, to be able to control for appli-

[6]We focus on the pre-affirmative action period (i.e., pre-2005) to avoid potential changes in the pool of applicants caused by the policy (Estevan et al., 2019). We exclude years before 2001 since the number of applicants admitted to P_2 was much larger than in 2001-2004, and their profiles are not necessarily comparable.

cants' subject-specific ability in our empirical approach.[7] Since we know the applicants' major choices, we can easily identify their priority subjects.

We observe applicants' grades on each of the 12 P_1 and 72 P_2 questions. For 2001-2002, we can also distinguish between an omitted question and an answered question that received a score of zero. The question-level performance information will allow us to investigate channels through which a gender performance difference in reaction to increasing rewards could emerge.

Finally, the dataset contains ENEM scores for those applicants who provided their ENEM ids (96% of our sample), applicants' gender, and socioeconomic variables such as age, parental education, and high school type. ENEM (*Exame Nacional do Ensino Médio*) is a national end-of-high-school exam, externally graded, used by some universities as their only admission criterion or as part of their admission process. Therefore, the ENEM score provides an ability measure independent of the applicant's performance on the UNICAMP admission exam.

Our initial sample contains 45,687 applicants who attended both P_1 and P_2 for admission (not as a practice test) and applied to a major not requiring an aptitude test. We drop applicants with missing gender information (0.6%), with ages below 16 or above 27 (2.5%), a missing ENEM score (4%), and no priority subjects among the six covered subjects (0.3%).[8] Our final sample contains 42,275 applicants for the 2001-2004 admission years.

It is worth mentioning that we are looking at a selected pool of applicants (i.e., those attending both P_1 and P_2). The admission exam is very competitive as only 29% of applicants pass P_1. Thus, we analyze gender performance gaps for applicants who perform well in a competitive environment, which is probably more similar to the real-world competitive labor markets.

Table 2 presents applicant-level information for our sample of interest. Female and male applicants differ in many meaningful dimensions, justifying our empirical strategy presented in the next section. First, females have, on average, significantly lower ENEM scores and apply to less competitive majors (based on P_2 score cutoffs). Females' average ENEM score is 0.25 s.d. below the overall average, while males' is 0.19 s.d. above. The difference in the average major cutoff (about 26 pts) 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.

Since females and males apply to different majors (Table 1), their priority subjects differ significantly. Forty percent of females have biology as a priority, a proportion twice as large

[7] There are no questions specific to Portuguese or foreign languages in P_1. Including Portuguese and using applicants' P_1 essay score as a Portuguese-specific ability measure yields similar results to our main estimates (see the Robustness Checks Subsection).

[8] In practice, we drop applicants that selected Philosophy after 2001. Starting in 2002, Mathematics was no longer a priority, and Portuguese became the only priority subject.

Table 2: 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	0.19	-0.43***
	(1.00)	(1.04)	(0.93)	
# of priority subjects	1.79	1.71	1.85	-0.14***
	(0.41)	(0.45)	(0.36)	
Major cutoff	518.51	503.59	529.73	-26.15***
	(78.23)	(82.90)	(72.53)	
Biology is a priority subject	0.28	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	0.40	0.72	-0.32***
	(0.49)	(0.49)	(0.45)	
Physics is a priority subject	0.41	0.24	0.54	-0.31***
	(0.49)	(0.43)	(0.50)	
Portuguese is a priority subject	0.10	0.16	0.05	0.10***
	(0.30)	(0.36)	(0.23)	
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	-0.22***
	(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)	
P2 score standard deviation	7.27	7.14	7.36	-0.22***
	(2.50)	(2.46)	(2.53)	
P2 score coefficient of variation	0.32	0.34	0.31	0.03***
	(0.17)	(0.18)	(0.16)	
Avg annual wages - 10 to 14 years after exam	42,437	35,936	47,148	-11,212***
	(29,019)	(24,078)	(31,296)	
Max annual wages - 10 to 14 years after exam	53,448	45,484	59,220	-13,736***
	(37,708)	(31,387)	(40,733)	
Match rate - RAIS 10 to 14 years after exam	0.78	0.77	0.80	-0.03***
	(0.41)	(0.42)	(0.40)	
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	42,275	18,151	24,124	

Notes: We compute our descriptive statistics at the applicant level. 'Age' is a student's age in June of the exam year. ENEM scores are normalized to have a mean of 0 and a standard deviation of 1 each year. 'Major cutoff' corresponds to the final score of the student admitted with the lowest score in her first-choice major. We compute the normalized P_1 and P_2 scores by averaging our six subjects' normalized scores. Subject-specific scores are normalized to have a mean of 0 and a standard deviation of 1 for each subject-year. We calculate normalized P_1 and P_2 'scores (average)' using equal weights for the six subjects. 'Phase 2 scores (weighted average)' uses a weight of 2 for priority subjects and 1 for non-priority subjects. 'Match rate - RAIS 10 to 14 years after' is the match rate of RAIS and UNICAMP administrative datasets between 10 and 14 years after they took the admission exam. For the matched sample, we compute the average ('Avg') and maximum ('Max') real annual wages (in 2002 Brazilian *reais*). 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%.

as males. Some of these majors have only one priority subject (e.g., biological sciences and nursing), which explains the smaller average number of priority subjects for females. In contrast, males are likelier to have mathematics (72% vs. 40%) or physics (54% vs. 24%) as priority subjects.

Table 2 also shows gender differences in P_2 performance 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 weighted average is above 0, indicating that applicants perform better in their priority subjects and select majors based on their relative advantages.

In theory, gender differences in major selection, and therefore priority-subject choice, would jeopardize the identification of the effect of increased rewards, especially if this selection is based on expected performance in priority subjects. The richness of our data — being able to observe applicants' previous subject-specific performance and to include applicant fixed effects (or an overall academic ability measure) given that we observe six outcomes per applicant — and 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.

We present the within-applicant standard deviation for both P_1 and P_2 performance. Since, for each applicant, our variation of interest occurs at the subject 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).

Lastly, we combine UNICAMP administrative data with those from the Brazilian formal labor market (RAIS) to obtain information on applicants' wages years after they wrote the admission exam. The RAIS is a matched employer-employee database covering the universe of formal labor market employees. Table 2 shows that the match rate of RAIS and UNICAMP administrative datasets from 10 to 14 years after the UNICAMP exam is 78%, with slightly lower match rates for females (three percentage points lower). Our wage measures compute the average and maximum wages observed between 10 and 14 years after candidates applied to UNICAMP. We report real wages in 2002 Brazilian *reais*.[9] Females earn significantly less than males, their annual wages being between 11,000 and 14,000 Brazilian *reais* (or 23-24 %) lower.

[9] We have RAIS information for 2002 to 2018. Therefore, for all cohorts (2001–2004), we observe individual labor market outcomes for up to 14 years after they applied to UNICAMP.

4 Empirical Model

We 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 score. Applicant i's performance on a specific subject, 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 applicants' ability and reaction to a priority subject to differ across subjects. Finally, ε_i^s is a purely random performance shock. Note that, other than $\mathbb{P}_i^s$, none of the terms on the right-hand side of equation (1) are observed.

Given that we are interested in 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 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.[10]

Equation (2) suggests the following regression equation:

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

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'

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

average general academic ability, and $\Delta\pi$ is the gender gap in general academic ability $(\pi^f - \pi^m)$.[11] 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 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., $\beta_{4,s} = \beta_4, \forall s$).[12] 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, 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 more challenging 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.

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)).[13] 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.

[11] Also, $\beta_1 \equiv \pi^m$, $\beta_2 \equiv \Delta\pi$, $\beta_{3,s} \equiv \phi^m + \omega^{s,m}$, and $\beta_{4,s} \equiv \Delta\phi + \Delta\omega^s$.

[12] 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.

[13] See, e.g., Ellison and Swanson (2010), Niederle and Vesterlund (2010), and Ellison and Swanson (forthcoming) for discussions on gender performance differences in mathematics.

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 ($\mathbb{P}_i^s\mathbb{F}_i$). However, given that applicants choose their major and, therefore, their priority subjects, such an assumption is unlikely to hold. It is entirely 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 gender. 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 P_1 and P_2. Importantly, in P_1, each subject is equally weighted to determine who will move on to the second phase, and overall performance in P_1 has a small impact on the likelihood of being admitted (its weight is equivalent to one priority subject), conditional on making it to P_2.[14] There is no incentive to do well in particular subjects. Hence, we can use applicants' performance on each P_1 subject to control for their relative subject-specific ability ($\widetilde{\gamma}_i^{s,g}$).

Finally, to control for applicants' overall and subject-specific relative 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 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:

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

where $\widetilde{P1}_i^{s,g}$ is the applicant's P_1 relative performance in subject s, and $\widetilde{ENEM}_i^g$ is the ap-

[14]The fact that P_1 performance has only a weight of 2 in the calculation of the final 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 a differential would lead to a biased estimation of β_4.

plicant'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.[15] Subject and individual fixed effects will capture ρ^s and π_i, while gender-subject interactions will capture $\gamma^{s,g}$. Note 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. Our performance measure y_i^s is the applicant's P_2 score in subject s, normalized by subject and admission year.[16]

4.1 Validity of the Empirical Model

A potential concern would arise if, for instance, males consider their comparative advantage when selecting their major and females 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. Even if our empirical strategy deals with gender differences in comparative advantages, we now show that there is no evidence of gender-specific selection differences into priority subjects, minimizing concerns of biases in our estimated coefficients.

Specifically, we investigate whether we observe gender performance differences in 'future' priority subjects in P_1 when 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. If, instead, females were to concentrate more than males on P_1 disciplines that will be priorities in P_2, then one could argue that females cannot improve as much as males in priority subjects between P_1 and P_2. Such a situation would have consequences for interpreting our results from estimating equation (5) since we control for P_1 subject-specific performance. More focused female priority-subject specialization in P_1 (compared to males) could lead to a negative estimate for our parameter of interest, despite females being the ones reacting more to increased rewards.

Our data allow us to investigate potential gender differences in subject specialization before P_2. As all subjects are equally weighted in Phase 1, we can use P_1 performance to run a placebo test. Table 3 presents 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, a 'Female $\times$ Future Priority' interaction term, and the same regressors as in equation (5) (but without P_1-scores controls). Two findings come out of Table 3. First, applicants do better in subjects that will be a priority in P_2. This finding suggests that they apply to majors in which they have a comparative advantage and that

[15] We use quartic functions (e.g., $G(\widetilde{P1}_i^{s,g}) \equiv \sum_{j=1}^{4} \alpha_j (\widetilde{P1}_i^{s,g})^j$) in our main specifications.

[16] As a robustness test, we use P_2 raw scores as our dependent variable, and the results are very similar. We report these estimates in Online Appendix Table O.1.

we should control for P_1 scores in our P_2 regressions to control for such selection. Second, and importantly, females do not concentrate more than males, 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 P_1 scores within-applicant standard deviation (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 3: 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	0.299***	0.357***	0.324***	0.341***
	(0.005)	(0.006)	(0.006)	(0.007)
Female × Future priority	0.025***	0.007	-0.009	-0.007
	(0.008)	(0.009)	(0.009)	(0.009)
ENEM	0.385***	0.384***		
	(0.003)	(0.003)		
Number of observations	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275
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 such that they have a mean 0 and a standard deviation of 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' is the applicant's ENEM relative performance, i.e., the applicant's normalized ENEM score minus her/his gender-year group average normalized ENEM. 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 ENEM 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%.

5 Results

Table 4 presents results from estimating 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.

Table 4: Priority Subjects and Gender Performance Gap

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.197***	-0.085***				
	(0.006)	(0.008)				
Priority	0.479***	0.578***	0.543***	0.480***	0.496***	0.514***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	0.000	-0.049***	-0.063***	-0.061***	-0.053***	-0.051***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ENEM	0.543***	0.541***				
	(0.003)	(0.003)				
Female × Chemistry		-0.175***	-0.181***	-0.190***	-0.191***	-0.189***
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Geography		-0.060***	-0.074***	-0.086***	-0.086***	-0.087***
		(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
Female × History		-0.043***	-0.051***	-0.059***	-0.059***	-0.060***
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Female × Mathematics		-0.098***	-0.117***	-0.149***	-0.149***	-0.147***
		(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Female × Physics		-0.195***	-0.216***	-0.245***	-0.244***	-0.243***
		(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Number of observations	253,650	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275	42,275
Subject 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 they have a mean 0 and a standard deviation of 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 0 and the standard deviation is 1 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%.

The specification in column (1) of Table 4 does not attempt to control for applicants' selection into majors other than through a linear control for overall relative ability. Since Table 2 suggests gender differences in major selection, it is unlikely that the β_4 estimate in column (1) captures a causality link unless our measure of ability completely captures this selection. Therefore, the following specifications sequentially attempt to control for 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' captures the gender performance gap in biology. The 'Female × Priority' estimate is sizable as it represents about 8 percent of the within-applicant standard deviation (0.62). The female-subject interaction terms suggest that, once we control for

ENEM, females perform worse than males in all six subjects, with the largest gender gaps being 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. The main impact of using fixed effects 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 further control 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 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 fairly flexibly, we still find that applicants' performance increases significantly when facing priority subjects. Still, females do not react as much as males when facing increased rewards.

5.1 Heterogeneity

The imposition of a common effect of increased rewards across subjects in Table 4 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 priority-subject effects across subjects, we estimate equation (5), allowing β_3 and β_4 to vary across subjects. Table 5 presents the results. We focus our discussion on specification (6), in which we have our full set of controls. We do not find gender differences in reaction to increased rewards in biology or chemistry. However, we observe large reactions (around -0.11, or 18% of a within-applicant s.d.) when rewards increase in other subjects. Looking back at our descriptive statistics (Table 2), 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.[17]

[17] 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

Table 5: Heterogeneity Across Subjects

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.197***	-0.125***				
	(0.006)	(0.009)				
Priority	0.647***	0.741***	0.373***	0.312***	0.312***	0.333***
	(0.011)	(0.012)	(0.010)	(0.010)	(0.010)	(0.011)
Female × Priority	0.035***	-0.036**	0.001	0.008	0.015	0.011
	(0.013)	(0.016)	(0.013)	(0.013)	(0.013)	(0.013)
Female × Priority × Chemistry	-0.078***	0.088***	0.036**	0.029	0.023	0.026
	(0.017)	(0.021)	(0.018)	(0.018)	(0.018)	(0.018)
Female × Priority × Geography	-0.102	-0.083	-0.114*	-0.107*	-0.130**	-0.126**
	(0.066)	(0.068)	(0.064)	(0.063)	(0.063)	(0.063)
Female × Priority × History	-0.100***	-0.113***	-0.111***	-0.120***	-0.115***	-0.113***
	(0.023)	(0.028)	(0.025)	(0.024)	(0.024)	(0.024)
Female × Priority × Mathematics	-0.125***	-0.020	-0.092***	-0.102***	-0.101***	-0.094***
	(0.017)	(0.026)	(0.022)	(0.021)	(0.021)	(0.021)
Female × Priority × Physics	-0.261***	-0.122***	-0.119***	-0.120***	-0.118***	-0.110***
	(0.018)	(0.025)	(0.021)	(0.020)	(0.020)	(0.020)
Priority × Chemistry	-0.013	-0.138***	-0.171***	-0.145***	-0.147***	-0.149***
	(0.013)	(0.015)	(0.013)	(0.013)	(0.013)	(0.013)
Priority × Geography	0.155***	-0.041	0.573***	0.548***	0.589***	0.580***
	(0.045)	(0.046)	(0.043)	(0.042)	(0.042)	(0.042)
Priority × History	0.022	-0.067***	0.571***	0.544***	0.557***	0.550***
	(0.017)	(0.020)	(0.018)	(0.017)	(0.017)	(0.017)
Priority × Mathematics	-0.311***	-0.321***	0.215***	0.225***	0.235***	0.223***
	(0.013)	(0.019)	(0.016)	(0.016)	(0.016)	(0.016)
Priority × Physics	-0.148***	-0.198***	0.214***	0.202***	0.209***	0.197***
	(0.013)	(0.018)	(0.015)	(0.014)	(0.014)	(0.014)
ENEM	0.537***	0.536***				
	(0.003)	(0.003)				
Number of observations	253,650	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275	42,275
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 they have a mean of 0 and a standard deviation of 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 0 and the standard deviation is 1 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%.

Next, we consider whether the effect of increasing rewards on the gender performance gap varies with applicant ability. 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 increasing rewards on the gender performance gap increases or decreases with ENEM.[18]

Table 6 suggests that the impact of increasing 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 is close to what we found in Table 4 (-0.057 versus -0.051). 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 0.05 s.d. more than a male applicant with an ENEM score at the mean. This difference would be 0.019 s.d. (0.050-0.031) for females. This difference is statistically significant at a 1% level. So, as we compare females and males with larger 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 0.088 (14% of the within-applicant s.d.) in favor of males.[19] Given that admission to UNICAMP is competitive, our findings suggest that the gender difference in reaction to increased rewards could affect the gender representation of admitted students.[20]

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. The results in Online

applicants' performance is 0.085 s.d. lower (13% of the within-applicant s.d.). Looking at heterogeneous impacts by subject, we still observe no gender differences in biology and a relatively better female performance in chemistry when the discipline is a priority (by 0.057 s.d. and statistically significant at 5%). We present these results in the Online Appendix Tables O.2 and O.3.

[18] Using a quartic instead of a linear function has little impact on our 'Female × Priority' and 'Female × Priority × ENEM' parameter estimates (Online Appendix Table O.4).

[19] Using P_1 subject-specific performance (instead of ENEM) as ability measure yields similar results. See Online Appendix Table O.5.

[20] We have run the regressions separately for female and male applicants to analyze the pattern of high-achieving applicants in priority subjects by gender (Online Appendix Tables O.6 and O.7). Males and females perform better in priority subjects, and the reaction to rewards is larger for the 'higher-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 females react less than males to an increase in rewards.

Table 6: Heterogeneity Across Academic Ability

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Female	-0.196***	-0.085***			
	(0.006)	(0.008)			
Priority	0.478***	0.577***	0.540***	0.477***	0.499***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.003	-0.050***	-0.061***	-0.059***	-0.057***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Priority × ENEM	-0.008	-0.013**	-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***	0.040***	0.050***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
ENEM	0.522***	0.522***			
	(0.005)	(0.005)			
Number of observations	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275
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 they have a mean of 0 and a standard deviation of 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 0 and the standard deviation is 1 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%.

Appendix Table O.8 are similar. If anything, the magnitude of our coefficient estimate of interest increases slightly.

Our analytical framework suggests using equation (5) as our main econometric model and controlling for relative performance in P_1 scores in our regressions. However, one could model performance changes between P_1 and P_2 instead to capture applicants' improvement, as in Cai et al. (2019) and Schlosser et al. (2019). 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).[21] As a robustness check, we use the difference between the P_2 and P_1 normalized scores as the dependent variable and estimate our model without controls for $G(\widetilde{P1}_i^{s,g})$. In Online Appendix Table O.9, we show that our coefficient estimates for 'Female × Priority' are very close to those in Table 4 once we

[21]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 4 (in specifications where we control for P_1).

control for subject-gender fixed effects. They vary between -0.047 and -0.056 in Table O.9 and between -0.049 and -0.063 in Table 4.[22] The only notable difference is that the coefficient estimates for 'Priority' are significantly smaller when we do not control for P_1 scores. Thus, our findings are not sensitive to whether we consider the P_2 scores in level while controlling for P_1 or our model in difference.

6 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 channels, like males not caring about non-priority subjects (admission is competitive), applicants' reaction to their performance in other subjects (applicants do not know their rank before the end of the admission process), or changes in the competition level (the pool of applicants is the same throughout Phase 2).

We leverage the richness of our data and analyze gender performance differences across exam subjects and at the question level. Recall that, for each subject, we observe applicants' performance on each of the 12 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 examine the plausibility of gender disparities in exam strategies as an explanation. Second, we discuss mechanisms we can rule out, such as gender gaps in performance in difficult questions, time management, and information.

6.1 Exam Strategy

For the 2001 and 2002 admission years, we can distinguish omitted questions from attempted questions that received a score of zero. Column (1) of Table 7 suggests that females leave more questions blank when rewards increase (the estimate represents an 11% increase relative to the average number of omitted questions). However, the results are reversed in column (2), as increased rewards lower females' number of attempted questions with a score of zero relative to males'. Finally, treating omitted questions as zeros, we observe a negligible effect of increased rewards on the gender gap in zeros (column (3)). These results suggest that females are more likely to shy away from answering questions they are uncertain about the answer.

As the admission exam does not penalize wrong answers, answering a question is a dominant strategy if time constraint is not binding. If females had tried (e.g., guessed an

[22] The standard deviation of the difference between the normalized P_2 and P_1 scores is slightly below 1 (0.959). Therefore, in terms of standard deviations, our estimates are even closer than Online Appendix Table O.9 suggests.

Table 7: Priority Subjects, Omitted Questions and Zeroes (2001-2002)

	Omissions	Zero scores	Zero + Omissions
	(1)	(2)	(3)
Priority	-0.512***	-0.563***	-1.075***
	(0.017)	(0.020)	(0.024)
Female × Priority	0.082***	-0.080***	0.002
	(0.019)	(0.022)	(0.026)
Mean dependent variable	0.76	2.09	2.85
Std.dev. dependent variable	1.54	2.10	2.74
Number of observations	120,270	120,270	120,270
Number of applicants	20,045	20,045	20,045
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 omitted questions (column (1)), the number of zeros on attempted questions, excluding omissions (column (2)), and the total number of zeros (column (3)), obtained in a given P_2 subject. The sample is restricted to 2001-2002. '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 0 and the standard deviation is 1 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%.

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. To investigate how much of the gender gap in priority subjects can be explained by the decision to skip questions, we use an Item Response Theory (IRT) Graded Response Model to estimate each applicant's predicted score on an omitted question based on their subject-specific ability and the question's difficulty level.[23] Online Appendix Table O.11 suggests that if applicants had answered all questions and achieved their predicted score, the gender gap in priority subjects would have been 4.2% instead of 4.7% in our main regression exercise using the 2001-2002 sample (Online Appendix Table O.10). Thus, while the omission pattern seems to play a role, it cannot account for a considerable proportion of the observed gender gap in reaction to higher rewards.

Next, we investigate whether females spread their effort more equally than males across

[23] One could argue that using the predicted score based on ability overestimates the score one would have obtained on a question since they decided to omit that question. However, the Graded 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 performance information in that specific subject. Since we find that females perform worse than expected on priority subjects, the estimated ability of female applicants will be underestimated for these subjects.

and within subjects in Phase 2. To do so, we compute the coefficient of variation for each applicant over all eight P_2 subject scores. We then regress the applicant's coefficient of variation on their ENEM score, a female dummy, and its interaction with the ENEM score.

Conditional on ENEM, female applicants seem to equalize their effort across subjects more than males, incurring a lower coefficient of variation, as shown in Table 8. Moreover, the correlation is stronger (more negative) and highly statistically significant for higher-ability females, the group with the largest gender performance gap in priority subjects, as shown in Table 6.

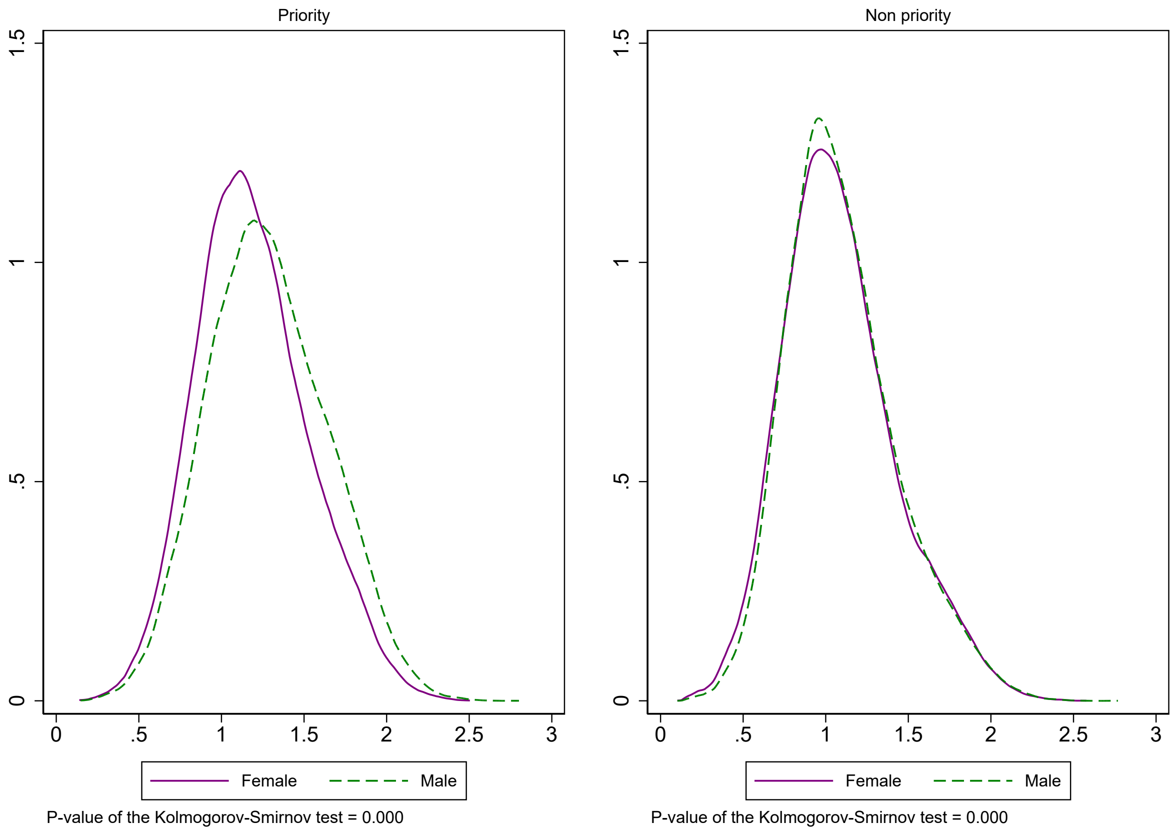
Table 8: Coefficient of Variation across Subjects

	(1)	(2)	(3)
Dependent variable: Phase 2 score coefficient of variation			
Female	0.031***	-0.005***	-0.006***
	(0.002)	(0.001)	(0.001)
Norm. ENEM scores		-0.083***	-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	42,275	42,275	42,275
Exam year FE	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 such that they have a mean of 0 and a standard deviation of 1 for each year. Heteroskedasticity-robust standard errors are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Females also exhibit lower performance variation within subjects than males in priority subjects. To account for each question's difficulty level and the applicant's subject-specific ability, we again use an IRT 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 1 presents distributions of IRT-residual standard deviations. 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.

Figure 1: Standard Deviations of IRT Residuals



6.2 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 females do relatively worse on more difficult questions (assuming that most applicants answer the easier questions correctly). 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 preparation 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 each question's average score. Then, we classify a question in a specific subject as *difficult* if the average performance on this question is below the median of questions' average scores for that subject.[24] We interact

[24] We have run an alternative specification where we define the difficulty level by gender. The results are similar and available in Online Appendix Table O.13. As robustness checks, we have also looked at different

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 72 observations per applicant (12 observations per applicant-subject).

Table O.12 suggests that females' performance on difficult questions does not drive our results. While there is a large change in the gender performance gap in reaction to increasing rewards (e.g., -0.155 points or around 9% of a standard deviation in specification (6)) on easier questions, it is not the case when priority questions are difficult. If anything, the gender gap changes slightly in favor of females. For example, in specification (6), the estimated change in the gender gap on difficult priority-subject questions is +0.027 (0.182-0.155). While this change in the gender gap is small, it is statistically significant at the 5% level in all specifications.

Therefore, our main findings (in Table 4) are not driven by females' relative performance on difficult priority-subject questions. However, females' 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 rewards on the gender performance gap comes mainly from their reaction to more straightforward priority-subject questions, i.e., those that most applicants answer correctly.

Mental Fatigue

Mental fatigue could also drive our results. 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 O.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). Next, 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.

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 similar regardless of the difficulty-level definition we use (see Online Appendix Tables O.14 and O.15).

Table O.16 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 perform worse on later questions. Table O.16 suggests that while males' performance worsens more when facing priority subjects as they move to questions toward the end of the exam, females' relative performance remains unchanged whether they face a priority subject or not. The sum of the 'Priority' and 'Female $\times$ Priority' parameter estimates in column (6) is close to zero and statistically insignificant. Our results are consistent with the conclusions from Balart and Oosterveen (2019). Their paper provides evidence that women have a higher capacity than men to sustain their performance during long exams and that this pattern holds even for male-stereotyped domains.

Information

In theory, gender differences in information about the UNICAMP admission exam's rules could explain our results. Suppose male applicants have better access to information than females. Females could, in that case, not focus on priority subjects simply because they are not fully aware 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 and 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 to better prepare for university admission exams. Also, UNICAMP is located in the municipality of Campinas, 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 to those 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 female performance in priority subjects, and the coefficients' sizes are also similar, suggesting that potential gender differences in exam information cannot explain our findings (see Online Appendix Tables O.17-O.21).

7 Does the Gender Gap in Exam Performance Matter?

7.1 University admission

Our results so far show that females' performance increases less than males' when facing larger rewards and that this difference increases with applicants' ability. 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.

We perform a counterfactual exercise to assess the potential relevance of gender differences in reaction to increased rewards in explaining university admission. In a nutshell, we simulate admissions by considering applicants' first-choice major and 'eliminating' the gender gap in priority subjects. We then compare the simulated and actual admissions.

More precisely, we first run a regression to obtain the 'Female × Priority' and 'Female × Priority × ENEM' coefficient estimates in a specification analogous to column (5) of Table 6, where we replace the normalized subject-specific scores by the raw P_2 scores. Next, we adjust female applicants' scores in priority subjects by eliminating the gender gap in priority subjects, considering the coefficient estimates obtained. Spending more time on omitted questions from priority subjects can reduce women's time on other questions. Therefore, we assume that better performance in priority subjects would reduce female performance in non-priority subjects.[25] Therefore, we redistribute the gender gap in priority subjects proportionally across non-priority subjects.[26] Then, we simulate admissions based on the adjusted scores, taking the total number of available slots per major as given.

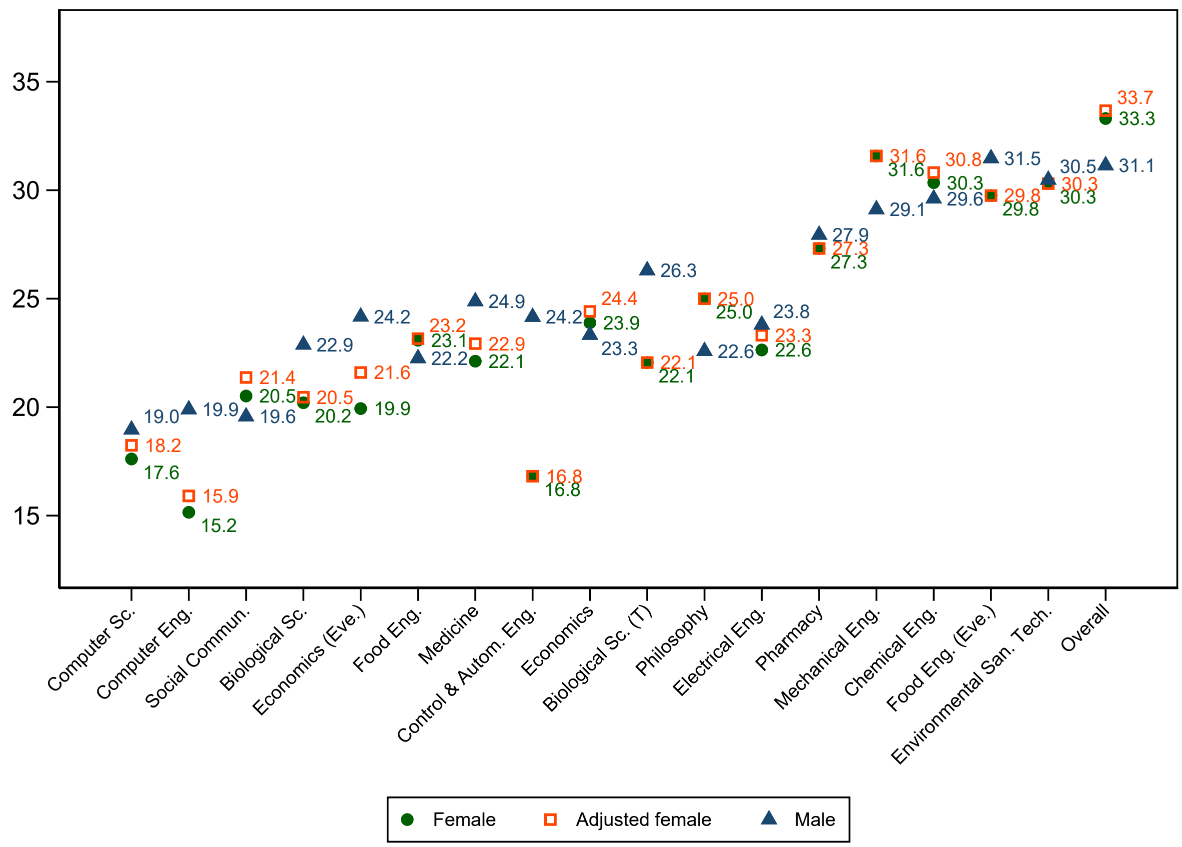
Figure 2 presents the actual admission rates for male and female applicants, considering all applicants registered for the admission exam. Overall, we find a modest impact on admission, corresponding to 1.2% of the actual admission rates. However, the effect is more prominent for competitive majors with lower-than-average admission rates (those presented in Figure 2). Men have higher (unconditional) admission rates than females in most of these majors. Eliminating the gender performance gap would notably impact female applicants' admission rates for some majors, as shown by the adjusted rates. For medicine and computer engineering, the adjusted admission rates are around 4-5% larger than the actual admission rates. We observe sizable effects in economics (evening), with a 9% increase in admission

[25] As an alternative specification, we assume that females' performance in non-priority subjects is unaffected. We interpret this exercise as an upper bound of the impacts on admission rates. We report this simulation in Online Appendix Figure O.2.

[26] For priority-subjects, we eliminate the gender gap in reaction to priority subjects using the 'Female × Priority' and 'Female × Priority × ENEM' coefficient estimates. In return, we reduce female performance in non-priority subjects by a proportional equivalent of the gender gap in priority subjects $\left(\frac{\#priority}{6-\#priority}\right)$.

rate. Thus, our estimated gender performance gap is sufficiently large to affect admission to some competitive majors meaningfully.

Figure 2: Admission Rates with Actual and Adjusted scores



Notes: The figure presents admission rates for males and females, considering all applicants registered for the admission exam. We show the overall admission rate, and those for majors characterized by lower-than-average admission rates.

7.2 Labor market outcomes

One of the main motivations of the literature on gender differences in reaction to increased exam competition or stakes is that these differences could explain part of the observed gender wage gap, especially at the top of the wage distribution. However, data limitations prevented most studies from directly linking their findings to labor market outcomes.

Our data allow us to investigate whether the gender differences in reaction to priority subjects are a significant driver of the gender wage gap. First, we estimate our main specification (column (6) in Table 4), excluding the 'Female × Priority' interaction, and obtain the residuals.[27] Then, we calculate the difference between the average residuals in priority and in non-priority subjects for each applicant. This difference informs us on whether an individual performs particularly well in priority subjects. Finally, we normalize this difference such that its mean is zero and its standard deviation is one.

Second, we estimate wage regressions to check whether controlling for our measure of priority-subject performance influences wages and the gender pay gap. In these wage regressions, we include both P_2 applicants who were admitted and those who were not. However, we must restrict our sample to applicants for whom we can observe formal labor market outcomes during the analyzed period (78% of individuals observed in our exam-performance regressions).[28]

The RAIS information is available for 2002 to 2018. Since our most recent cohort wrote the admission exam in 2004, we can observe applicants' labor market outcomes up to 14 years after the exam. We use the year applicants wrote the admission exam as the reference period since we do not observe the graduation year for those who were not admitted at UNICAMP (and some applicants may have not graduated from university).[29] We focus on applicants' labor market outcomes ten to fourteen years after the admission exam to give enough time for applicants to enter the labor market fully—the expected duration of undergraduate programs at UNICAMP is between 4 to 6 years.[30]

Panel A of Table 9 shows how the gender gap in log average annual wages changes when we

[27] As robustness tests, we use alternative regressions to construct the residuals. First, we estimate a simpler model controlling linearly for P_1 scores (Online Appendix Table O.22). Second, we estimate our main specification excluding the 'Female × Priority' interaction and the 'ENEM × Priority' and 'Phase 1 scores × Priority' polynomials (Online Appendix Table O.23). Lastly, we construct the residuals by estimating our main specification and adding the 'Priority' coefficients for priority subjects for male students and the 'Priority' and 'Female × Priority' coefficients for female students (Online Appendix Table O.24). Regardless of how we estimate our residuals, the results remain quite similar in the wage regressions once we control for major fixed effects.

[28] Conditional on major choice, females have lower participation 13 and 14 years after (around 2 p.p. gap). However, there is no gender gap in formal labor market participation from 10 to 12 years after the admission exam. The residuals do not influence labor force participation. We present these results in Online Appendix Table O.25.

[29] Based on the RAIS information, most applicants in our sample obtain a higher education degree, regardless of admission into UNICAMP. In our analysis period (10 to 14 years after the admission exam), more than 94% of them earned a higher education degree from any institution. In Online Appendix Table O.26, we present regressions in which the dependent variable is a binary variable equal to 1 if the applicant obtained a higher education diploma and zero otherwise. Women are slightly more likely to earn a diploma, as well as applicants with higher ENEM scores. A one s.d. increase in mean residual difference relates with a three percentage point higher likelihood of holding a higher education degree between 10 and 14 years later. Although statistically significant at a 5% level, the small coefficient size suggests that degree completion is not a relevant channel in explaining the results observed in the wage regressions.

[30] As a robustness check, we look at labor market outcomes 6 to 14 years after the exam, and estimates are similar (Online Appendix Table O.27). We also examine earlier career wages (between 6 and 9 years after the exam) in Online Appendix Table O.28 and find parallel results.

Table 9: Log (Annual Wages) 10 to 14 Years After the Admission Exam

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average annual wages (10-14 years after admission exam)								
Female	-0.270***	-0.269***	-0.196***	-0.195***	-0.122***	-0.122***	-0.095***	-0.096***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.016***		0.016***		0.026***		0.026***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.166***	0.166***			0.102***	0.102***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes
Panel B: Maximum annual wages (10-14 years after admission exam)								
Female	-0.262***	-0.261***	-0.185***	-0.184***	-0.121***	-0.122***	-0.095***	-0.096***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.015***		0.014***		0.026***		0.026***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.173***	0.173***			0.099***	0.099***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.622	10.622	10.622	10.622	10.622	10.622	10.622	10.622
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the log of the average or maximum annual wages between 10 and 14 years after the applicant took the UNICAMP admission exam. 'Normalized residuals' is the difference between the applicant's average residuals in priority and non-priority subjects in our main specification, excluding the 'Female × Priority' interaction. We normalize the difference in average residuals to have a mean of 0 and a standard deviation of 1. Standard errors based on 999 applicant-level cluster-bootstrap replications are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

include our priority-subject performance measure (i.e., the 'normalized residuals').[31] Column (1) presents the gender wage gap when we control only for exam-year fixed effects, which will serve as a baseline. In columns (2) to (4), we add controls for our measure of priority-subject performance and ENEM (a measure of overall academic ability) to compare their relative contribution to the gender wage gap. Columns (5) to (8) present the same specifications, but where we also control for (applied-) major fixed effects. As a robustness check, Panel B repeats the exercise presented in Panel A using (the log of) the maximum annual wage as the dependent variable (the maximum wage could be a better measure of career progress than the average).

To be concise, since the results for maximum annual wages are very similar, we only

[31] Wages are measured in 2002 Brazilian *reais*. Since we use a generated regressor, we compute standard errors by bootstrapping the whole estimation procedure (i.e., estimating the P_2 performance regressions and the wage regressions). Standard errors presented in Table 9 are based on 999 applicant-level cluster-bootstrap replications.

discuss the average annual wages. The baseline gender gap is 27 log points (Column (1)). Controlling for major fixed effects greatly reduces this gap: 12 log points (Column (6)). Our findings suggest that applicants who do better than expected on priority subjects earn more. A one s.d. increase in mean residual difference is associated with 1.6–2.6% higher wages. The 'normalized residuals' parameter is statistically significant at a 1% level in all specifications. While the estimate might seem small, one should remember that our performance regressions include individual fixed effects, and they are therefore excluded from the residuals.

One striking result from Table 9 is that when we compare the gender wage gaps, we can see that controlling for priority-subject performance barely affects the gender gap. As a comparison, we note that the gender gap is quite similar in columns (1) and (2), while column (3) suggests that controlling for academic ability reduces the gender gap significantly (by more than a quarter), and controlling for the major individuals applied to is even more critical. So, while performing well on priority subjects correlates positively with wages, it does not explain much of the gender wage gap.

To analyze how the performance in higher rewarding exams influences wages all over the distribution, we present quantile regression results for the 10th, 25th, 50th, 75th, and 90th percentiles in the Online Appendix Table O.29. As we would expect, the gender wage gap increases as we move to upper quantiles. While there is no gender gap at the 10th percentile, we observe a large wage gap at the 90th percentile (around 15 log points, controlling for major fixed effects and ENEM scores). We note that the normalized residuals influence wages all over the distribution, although the coefficient estimates are larger for the bottom percentiles (10th and 25th). As in our main specification, including the normalized residuals does not affect the estimated gender wage gap 10 to 14 years after the exam.

8 Conclusion

This paper provides evidence that females and males react differently to increased rewards, proposes channels through which this difference might operate, and whether it translates to the labor market. We observe males and females taking identical exams but with varying rewards 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. Moreover, our data allows us to follow applicants in the formal labor market up to 14 years after the admission exam.

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

Our evidence suggests that females and males adopt distinct approaches when uncertain about a question's correct answer. Females become relatively more likely to omit questions when these questions are associated with priority subjects, while males are more likely to try to answer the questions and obtain zero scores. Also, we observe that females 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 exam strategy are behind our findings.

Our findings suggest that applicants who perform better on priority subjects earn higher annual wages at the formal labor market up to 14 years after the UNICAMP admission exam. Surprisingly, it does not seem to relate to the gender wage gaps in the labor market meaningfully. Therefore, our results cast doubt on whether gender differences in behavior in an exam environment can explain gender gaps in the labor market.

Acknowledgements

This research has benefited from discussions with Ghazala Azmat, Sonia Bhalotra, Pierre Brochu, Abel Brodeur, Bladimir Carrillo, Catherine Deri Armstrong, Hülya Eraslan, Raissa Fabregas, Edoardo Gallo, Jason Garred, Paola Giuliano, Daniel Haanwinckel, Adam Lavecchia, Cecilia Machado, Mario Macis, Ricardo Madeira, Philip Oreopoulos, Daniele Paserman, Aureo de Paula, Cristine Pinto, Christopher R. Taber, and Eric Verhoogen, and comments from seminar participants at PUC-Rio, McMaster, U. República, SBE, PIMES/UFPE, the 48th ANPEC Meeting, the 2nd Brazilian Meeting in Family and Gender Economics, the 44th SBE Meeting, the 27th SOLE Annual Meetings, the 2022 AASLE Conference, the WEAI 17th International Conference, and the XVII RIDGE FORUM/WELAC (LACEA) Workshop on Gender and Household Economics. Financial support from the SSHRC (435-2020-0081), São Paulo Research Foundation (FAPESP) (grants # 2015/21640-3 and # 2019/25033-5), CALDO, and CNPq are gratefully acknowledged. Bruna received funding from FAPESP (grants # 2017/23593-8 and # 2021/08880-6) and from CAPES (Finance Code 001). We are grateful to COMVEST for providing the data and assistance during the project. All remaining errors are ours.

References

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.

Balart, P. and Oosterveen, M. (2019). Females show more sustained performance during test-taking than males. *Nature communications*, 10(1):3798.

Bandiera, O., Fischer, G., Prat, A., and Ytsma, E. (2021). Do women respond less to performance pay? Building evidence from multiple experiments. *American Economic Review: Insights*, 3(4):435–54.

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

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3):789–865.

Buser, T., Niederle, M., and Oosterbeek, H. (2021). Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures. Working Paper 28916, National Bureau of Economic Research.

Buser, T. and Peter, N. (2012). Multitasking. *Experimental Economics*, 15:641–655.

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. (2019). Redistribution without distortion: Evidence from an affirmative action program at a large Brazilian university. *Economic Journal*, 129(619):1182–1220.

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.

Jurajda, Š. and Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to Czech universities. *American Economic Review: Papers and Proceedings*, 101(3):514–518.

Morin, L.-P. (2015). Do men and women respond differently to competition? Evidence from a major education reform. *Journal of Labor Economics*, 33(2):443–491.

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.

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

Reuben, E., Sapienza, P., and Zingales, L. (2015). Taste for competition and the gender gap among young business professionals. Working Paper 21695, National Bureau of Economic Research.

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.

Online Appendix (For Online Publication)

Gender Differences in Prioritizing Rewarding Tasks

Bruna Borges, Fernanda Estevan and Louis-Philippe Morin

O.1 UNICAMP Admission

In this Online Appendix, we provide more details on UNICAMP's admission for the majors we consider in our empirical analysis.

UNICAMP's admission exam has two sequential phases, referred to as P_1 and P_2. Applicants write P_1 in November, P_2 in January, and the academic year begins in late February.

P_1 comprises twelve written-answer questions on biology, chemistry, geography, history, mathematics, and physics. Each question is worth 2.5 points. Applicants must also write an essay worth 30 points in P_1. UNICAMP computes the applicants' P_1 score using two different formulas and selects the most favorable one for each applicant. The first formula sums the essay and the general question scores. In the second formula, the sum of the essay and general question scores 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%.[32] Thus, the maximum possible P_1 score is 60 points, under both formulas.[33]

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 was 84% for medicine and 43% for statistics.

[32] ENEM (*Exame Nacional do Ensino Médio*) is a national end-of-high-school exam used by some universities as their only admission criterion, or as part of their admission process (like UNICAMP). Between 2001 and 2004, ENEM comprised 63 multiple-choice questions based on high-school subjects and an essay. UNICAMP considers only the ENEM score based on the multiple-choice questions. Applicants must have taken the ENEM in the previous two years and provided UNICAMP with their ENEM ID to be eligible for the second formula. The formula that considers the applicant's ENEM score is advantageous for 85% of applicants. In our analysis, we drop applicants without a valid ENEM score. 96% of individuals who passed P_1 (the individuals we will focus on in our empirical analysis) submitted a valid ENEM score.

[33] 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.

P_2 covers the same six high-school subjects plus Portuguese and a foreign language (English or French). There are twelve questions per subject, each receiving the same weight. 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 have four hours to submit their answers for both subjects. Depending on the applicant's major choice, one or two subjects are considered *priority subjects*, receiving a weight of two (instead of one) in the final score calculation.

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 applicants who participated in the exam. In all admission years, the standardization of the P_1 scores only considers the scores of applicants who passed P_1.

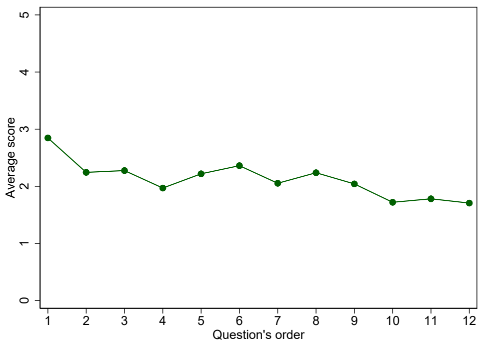
An applicant's final admission score is the weighted average of her normalized: 1) P_1 score, with a weight of two; 2) P_2 priority-subject scores, each with a weight of two, and; 3) P_2 non-priority subject scores, each with a weight of one.

UNICAMP ranks applicants based on their final score and major choice. 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 their second or third option.

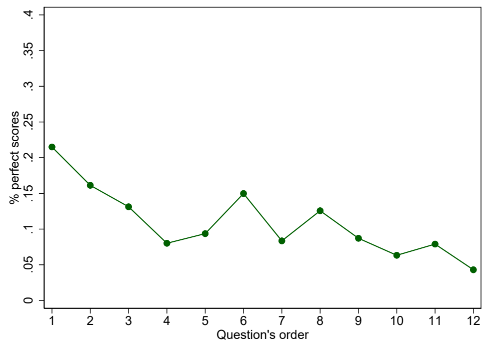
O.2 Appendix Figures and Tables

Figures

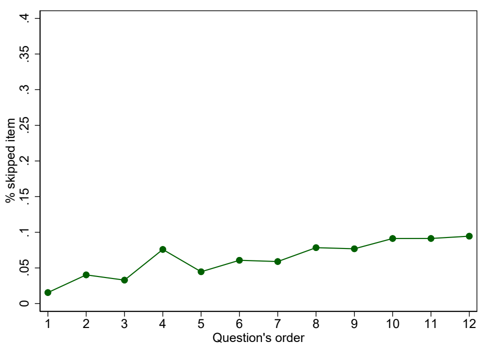
Figure O.1: Performance by Question Order



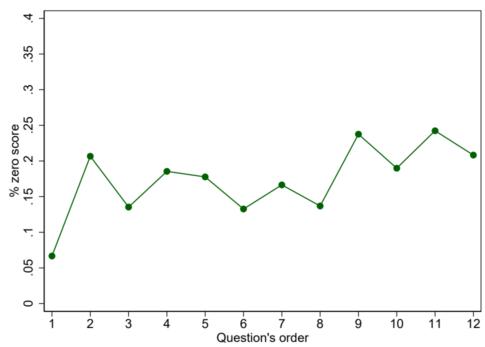
(a) Average Score



(b) Share of 'Perfect' Scores



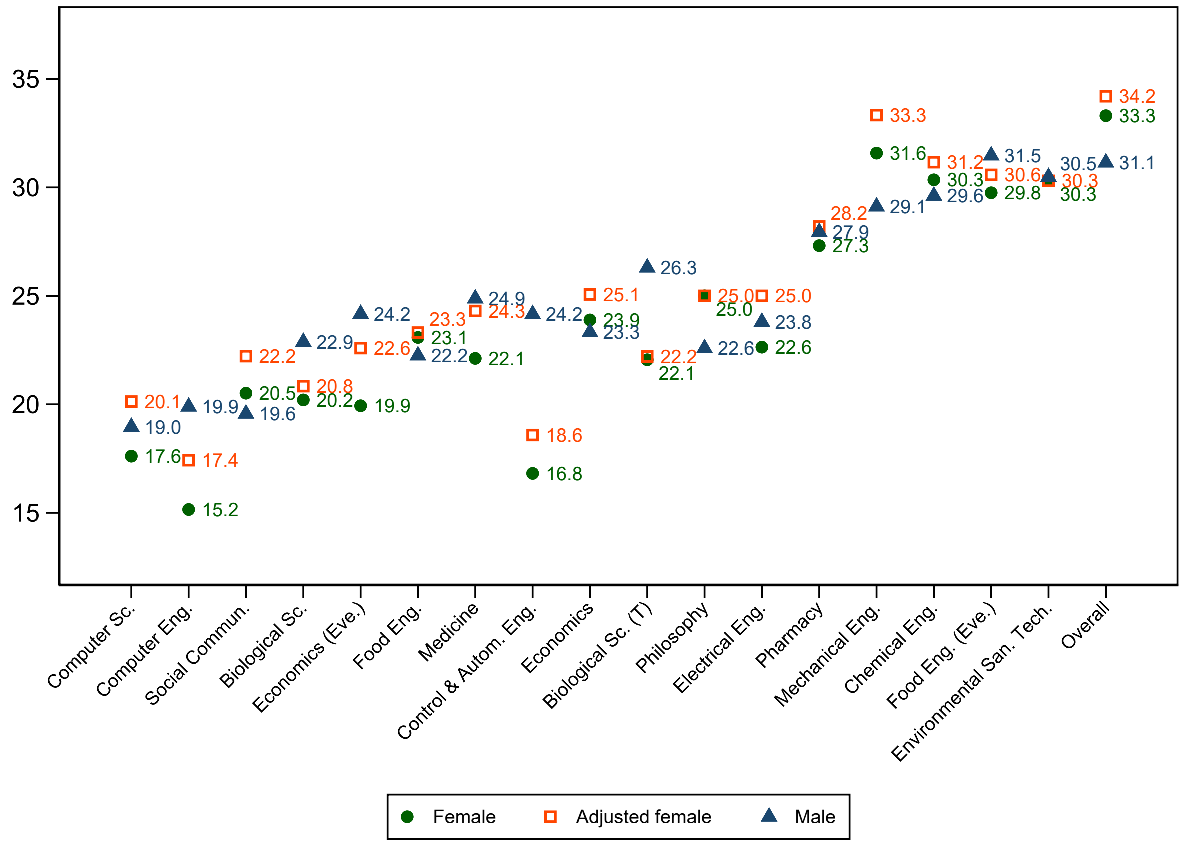
(c) Share of Omissions



(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.

Figure O.2: Admission Rates with Actual and Adjusted scores, Upper Bound



Notes: The figure presents actual admission rates for males and females, 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 6, 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.

Tables

Table O.1: Priority Subjects and Gender Performance Gap (P_2 Raw Scores)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 subject-specific raw scores						
Female	-2.074***	-0.761***				
	(0.055)	(0.083)				
Priority	4.933***	6.158***	5.689***	4.936***	5.055***	5.295***
	(0.051)	(0.059)	(0.055)	(0.052)	(0.058)	(0.068)
Female × Priority	-0.439***	-0.787***	-0.916***	-0.834***	-0.735***	-0.683***
	(0.076)	(0.081)	(0.076)	(0.072)	(0.074)	(0.074)
ENEM	0.431***	0.429***				
	(0.002)	(0.002)				
Mean dependent variable	25.45					
Stand dev dependent variable	10.61					
Number of observations	253,650	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275	42,275
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, P_1 and P_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 O.2: Priority Subjects and Gender Performance Gap (Excluding Medicine)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.226***	-0.122***				
	(0.006)	(0.008)				
Priority	0.495***	0.665***	0.628***	0.561***	0.568***	0.573***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Female × Priority	-0.020***	-0.105***	-0.091***	-0.090***	-0.086***	-0.085***
	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ENEM	0.476***	0.473***				
	(0.003)	(0.003)				
Number of observations	222,336	222,336	222,336	222,336	222,336	222,336
Number of applicants	37,056	37,056	37,056	37,056	37,056	37,056
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 0 and the standard deviation is 1 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 Subjects (Excluding Medicine)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.226***	-0.134***				
	(0.006)	(0.009)				
Priority	0.417***	0.451***	0.407***	0.349***	0.367***	0.374***
	(0.017)	(0.018)	(0.015)	(0.014)	(0.014)	(0.015)
Female × Priority	0.173***	0.081***	-0.022	-0.017	-0.021	-0.021
	(0.020)	(0.022)	(0.018)	(0.017)	(0.017)	(0.017)
Female × Priority × Chemistry	-0.141***	0.026	0.076***	0.063**	0.057**	0.057**
	(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)	(0.066)	(0.064)	(0.064)	(0.064)
Female × Priority × History	-0.216***	-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***	-0.245***	-0.083***	-0.100***	-0.089***	-0.087***
	(0.023)	(0.031)	(0.025)	(0.025)	(0.025)	(0.025)
Female × Priority × Physics	-0.377***	-0.285***	-0.102***	-0.105***	-0.093***	-0.091***
	(0.024)	(0.030)	(0.024)	(0.023)	(0.023)	(0.023)
Priority × Chemistry	0.076***	-0.048*	-0.121***	-0.101***	-0.106***	-0.105***
	(0.023)	(0.025)	(0.020)	(0.020)	(0.020)	(0.020)
Priority × Geography	0.415***	0.268***	0.506***	0.478***	0.503***	0.501***
	(0.046)	(0.047)	(0.044)	(0.043)	(0.043)	(0.043)
Priority × History	0.310***	0.295***	0.528***	0.492***	0.482***	0.482***
	(0.021)	(0.023)	(0.021)	(0.020)	(0.020)	(0.020)
Priority × Mathematics	-0.020	0.260***	0.225***	0.245***	0.233***	0.231***
	(0.018)	(0.024)	(0.020)	(0.019)	(0.019)	(0.019)
Priority × Physics	0.148***	0.307***	0.229***	0.207***	0.189***	0.188***
	(0.019)	(0.022)	(0.018)	(0.017)	(0.017)	(0.018)
ENEM	0.476***	0.472***				
	(0.003)	(0.003)				
Number of observations	222,336	222,336	222,336	222,336	222,336	222,336
Number of applicants	37,056	37,056	37,056	37,056	37,056	37,056
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 0 and the standard deviation is 1 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 (ENEM × Priority - Quartic Function)

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Female	-0.196***	-0.084***			
	(0.006)	(0.008)			
Priority	0.441***	0.540***	0.562***	0.494***	0.512***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.012*	-0.058***	-0.053***	-0.052***	-0.050***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Female × Priority × ENEM	-0.016**	-0.018***	-0.020***	-0.024***	-0.022***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Female × ENEM	0.019***	0.021***			
	(0.006)	(0.006)			
ENEM	0.522***	0.522***			
	(0.005)	(0.005)			
Priority × ENEM	0.099***	0.099***	0.049***	0.059***	0.069***
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Number of observations	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275
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 0 and the standard deviation is 1 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.5: Heterogeneity Across Academic Ability (Phase 1 scores)

	(1)	(2)	(3)	(4)
Dependent variable: Phase 2 normalized subject-specific scores				
Female	-0.194***	-0.057***		
	(0.005)	(0.007)		
Priority	0.390***	0.456***	0.480***	0.498***
	(0.004)	(0.005)	(0.005)	(0.005)
Female × Priority	-0.024***	-0.047***	-0.052***	-0.044***
	(0.006)	(0.007)	(0.007)	(0.007)
Female × Priority × Phase 1 score	0.003	-0.002	-0.017***	-0.016**
	(0.007)	(0.007)	(0.006)	(0.006)
Female × Phase 1 score	-0.011***	-0.008**	-0.016***	-0.017***
	(0.004)	(0.004)	(0.004)	(0.004)
Priority × Phase 1 score	-0.016***	-0.022***	-0.015***	-0.030***
	(0.005)	(0.005)	(0.004)	(0.004)
Phase 1 score	0.361***	0.359***	0.208***	0.212***
	(0.003)	(0.003)	(0.003)	(0.003)
Number of observations	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275
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 0 and the standard deviation is 1 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.6: Heterogeneity Across Academic Ability (Females)

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Priority	0.475***	0.526***	0.479***	0.421***	0.444***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)
Priority × ENEM	0.040***	0.028***	0.005	0.005	0.020***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
ENEM	0.541***	0.543***			
	(0.004)	(0.004)			
Number of observations	108,906	108,906	108,906	108,906	108,906
Number of applicants	18,151	18,151	18,151	18,151	18,151
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 0 and the standard deviation is 1 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: Heterogeneity Across Academic Ability (Males)

	(1)	(2)	(3)	(4)	(5)
Dependent variable: Phase 2 normalized subject-specific scores					
Priority	0.478***	0.577***	0.540***	0.474***	0.497***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Priority × ENEM	0.048***	0.041***	0.036***	0.040***	0.051***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)
ENEM	0.522***	0.522***			
	(0.005)	(0.005)			
Number of observations	144,744	144,744	144,744	144,744	144,744
Number of applicants	24,124	24,124	24,124	24,124	24,124
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 0 and the standard deviation is 1 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.8: 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	0.496***	0.568***	0.548***	0.483***	0.494***	0.509***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.065***	-0.069***	-0.066***	-0.065***	-0.057***	-0.056***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
ENEM	0.536***	0.535***				
	(0.003)	(0.003)				
Number of observations	295,925	295,925	295,925	295,925	295,925	295,925
Number of applicants	42,275	42,275	42,275	42,275	42,275	42,275
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 0 and the standard deviation is 1 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: 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***	0.027***			
	(0.005)	(0.009)			
Priority	0.180***	0.221***	0.219***	0.223***	0.417***
	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)
Female × Priority	-0.025***	-0.056***	-0.055***	-0.047***	-0.049***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)
ENEM	0.158***	0.157***			
	(0.002)	(0.002)			
Number of observations	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275
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: The dependent variable is Phase 2 normalized score minus Phase 1 normalized score in each subject. 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 0 and the standard deviation is 1 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: 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	-0.227***	-0.113***				
	(0.008)	(0.011)				
Priority	0.464***	0.557***	0.525***	0.469***	0.491***	0.500***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
Female × Priority	-0.014	-0.057***	-0.063***	-0.061***	-0.049***	-0.047***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
ENEM	0.574***	0.572***				
	(0.004)	(0.004)				
Number of observations	120,270	120,270	120,270	120,270	120,270	120,270
Number of applicants	20,045	20,045	20,045	20,045	20,045	20,045
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 0 and the standard deviation is 1 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: Normalized Phase 2 Scores, Omissions = Predicted IRT Score

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Phase 2 normalized subject-specific scores						
Female	-0.245***	-0.117***				
	(0.008)	(0.011)				
Priority	0.450***	0.528***	0.501***	0.445***	0.468***	0.476***
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)
Female × Priority	-0.022**	-0.049***	-0.058***	-0.055***	-0.044***	-0.042***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)
ENEM	0.582***	0.580***				
	(0.004)	(0.004)				
Number of observations	120,270	120,270	120,270	120,270	120,270	120,270
Number of applicants	20,045	20,045	20,045	20,045	20,045	20,045
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 0 and the standard deviation is 1 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, 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	0.589***	0.690***	0.653***	0.601***	0.611***	0.620***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Female × Priority	-0.130***	-0.159***	-0.167***	-0.165***	-0.156***	-0.155***
	(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	0.445***	0.443***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev dependent variable	1.65					
# observations	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800
# applicants	42,275	42,275	42,275	42,275	42,275	42,275
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 the average performance on this question is below the median of question average scores for that 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 0 and the standard deviation is 1 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, Question's Difficulty (by Gender) and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.165***	-0.054***				
	(0.006)	(0.007)				
Priority	0.591***	0.693***	0.656***	0.604***	0.613***	0.623***
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Female × Priority	-0.131***	-0.160***	-0.168***	-0.165***	-0.156***	-0.156***
	(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	0.445***	0.443***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev dependent variable	1.65					
# observations	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800
# applicants	42,275	42,275	42,275	42,275	42,275	42,275
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 the average performance on this question is below the median of question average scores for that subject and gender. '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 0 and the standard deviation is 1 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, Very Difficult Questions and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.178***	-0.067***				
	(0.005)	(0.007)				
Priority	0.505***	0.606***	0.569***	0.517***	0.527***	0.536***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.083***	-0.112***	-0.120***	-0.117***	-0.109***	-0.108***
	(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	0.445***	0.443***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev dependent variable	1.65					
# observations	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800
# applicants	42,275	42,275	42,275	42,275	42,275	42,275
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). 'Very difficult question' is a dummy variable equal to 1 if the average performance on this question is among the bottom 25% of question average scores for that 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 0 and the standard deviation is 1 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.15: Priority Subjects, Most Difficult Questions and Performance

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Questions' raw scores (Phase 2)						
Female	-0.183***	-0.073***				
	(0.005)	(0.007)				
Priority	0.434***	0.536***	0.498***	0.447***	0.456***	0.466***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Female × Priority	-0.047***	-0.076***	-0.084***	-0.081***	-0.072***	-0.072***
	(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	-0.292***	-0.292***	-0.292***	-0.292***	-0.292***	-0.292***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Most difficult question	-1.204***	-1.204***	-1.204***	-1.204***	-1.204***	-1.204***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Female × Most difficult question	0.053***	0.053***	0.053***	0.053***	0.053***	0.053***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
ENEM	0.445***	0.443***				
	(0.003)	(0.003)				
Mean dependent variable	2.12					
Std.dev dependent variable	1.65					
# observations	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800	3,043,800
# applicants	42,275	42,275	42,275	42,275	42,275	42,275
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). 'Most difficult question' is a dummy variable equal to 1 if the average performance on this question is the lowest question average score for that 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 0 and the standard deviation is 1 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.16: 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	0.135***	-0.059***				
	(0.006)	(0.011)				
Priority	0.814***	0.014**	0.019**	0.010	0.044***	0.044***
	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)	(0.010)
Female × Priority	-0.665***	-0.037***	-0.045***	-0.042***	-0.034***	-0.031***
	(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	253,650	253,650	253,650	253,650	253,650	253,650
Number of applicants	42,275	42,275	42,275	42,275	42,275	42,275
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 0 and the standard deviation is 1 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.17: 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	-0.210***	-0.108***					-0.191***	-0.075***				
	(0.010)	(0.014)					(0.007)	(0.010)				
Priority	0.461***	0.569***	0.533***	0.480***	0.491***	0.498***	0.488***	0.583***	0.548***	0.482***	0.501***	0.523***
	(0.008)	(0.010)	(0.010)	(0.009)	(0.010)	(0.011)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)
Female × Priority	-0.023**	-0.065***	-0.061***	-0.060***	-0.052***	-0.049***	0.007	-0.046***	-0.069***	-0.066***	-0.057***	-0.056***
	(0.012)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ENEM	0.557***	0.555***					0.535***	0.534***				
	(0.005)	(0.005)					(0.004)	(0.004)				
# observations	77,778	77,778	77,778	77,778	77,778	77,778	174,078	174,078	174,078	174,078	174,078	174,078
# applicants	12,963	12,963	12,963	12,963	12,963	12,963	29,013	29,013	29,013	29,013	29,013	29,013
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 0 and the standard deviation is 1 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.18: 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	-0.198***	-0.076***					-0.199***	-0.111***				
	(0.007)	(0.009)					(0.011)	(0.015)				
Priority	0.488***	0.583***	0.549***	0.485***	0.503***	0.523***	0.460***	0.569***	0.529***	0.470***	0.483***	0.496***
	(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	0.008	-0.043***	-0.065***	-0.064***	-0.055***	-0.053***	-0.024**	-0.072***	-0.070***	-0.063***	-0.056***	-0.055***
	(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	0.539***	0.537***					0.549***	0.547***				
	(0.004)	(0.004)					(0.006)	(0.006)				
# observations	177,852	177,852	177,852	177,852	177,852	177,852	74,154	74,154	74,154	74,154	74,154	74,154
# applicants	29,642	29,642	29,642	29,642	29,642	29,642	12,359	12,359	12,359	12,359	12,359	12,359
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 0 and the standard deviation is 1 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.19: 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	-0.185***	-0.061***					-0.190***	-0.075***				
	(0.012)	(0.017)					(0.007)	(0.009)				
Priority	0.402***	0.588***	0.551***	0.488***	0.505***	0.525***	0.497***	0.568***	0.535***	0.473***	0.488***	0.505***
	(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	0.057***	-0.054***	-0.056***	-0.055***	-0.052***	-0.050***	-0.016**	-0.049***	-0.066***	-0.063***	-0.054***	-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***	0.484***					0.542***	0.541***				
	(0.006)	(0.006)					(0.004)	(0.004)				
# observations	51,174	51,174	51,174	51,174	51,174	51,174	185,610	185,610	185,610	185,610	185,610	185,610
# applicants	8,529	8,529	8,529	8,529	8,529	8,529	30,935	30,935	30,935	30,935	30,935	30,935
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 0 and the standard deviation is 1 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.20: 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***					-0.179***	-0.069***				
	(0.011)	(0.015)					(0.006)	(0.009)				
Priority	0.467***	0.666***	0.622***	0.538***	0.557***	0.572***	0.481***	0.547***	0.517***	0.462***	0.476***	0.495***
	(0.009)	(0.011)	(0.011)	(0.010)	(0.011)	(0.013)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)
Female × Priority	0.002	-0.085***	-0.072***	-0.067***	-0.058***	-0.057***	-0.007	-0.045***	-0.062***	-0.061***	-0.053***	-0.051***
	(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	0.475***	0.471***					0.549***	0.548***				
	(0.005)	(0.005)					(0.004)	(0.004)				
# observations	64,176	64,176	64,176	64,176	64,176	64,176	188,976	188,976	188,976	188,976	188,976	188,976
# applicants	10,696	10,696	10,696	10,696	10,696	10,696	31,496	31,496	31,496	31,496	31,496	31,496
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 0 and the standard deviation is 1 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.21: 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	-0.172***	-0.060***					-0.232***	-0.117***				
	(0.007)	(0.010)					(0.009)	(0.013)				
Priority	0.491***	0.550***	0.519***	0.463***	0.474***	0.491***	0.451***	0.622***	0.582***	0.507***	0.532***	0.549***
	(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	-0.017**	-0.045***	-0.060***	-0.059***	-0.052***	-0.049***	0.026**	-0.068***	-0.073***	-0.068***	-0.059***	-0.059***
	(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***					0.504***	0.502***				
	(0.004)	(0.004)					(0.005)	(0.005)				
# observations	162,456	162,456	162,456	162,456	162,456	162,456	88,704	88,704	88,704	88,704	88,704	88,704
# applicants	27,076	27,076	27,076	27,076	27,076	27,076	14,784	14,784	14,784	14,784	14,784	14,784
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 0 and the standard deviation is 1 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.22: Alternative Residuals: Linear P1-Score Function, Log (Annual Wages) 10 to 14 Years After Admission Exam

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average annual wages (10-14 years after admission exam)								
Female	-0.270***	-0.270***	-0.196***	-0.196***	-0.122***	-0.122***	-0.095***	-0.095***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.007		-0.000		0.022***		0.018***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.166***	0.166***			0.102***	0.101***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes
Panel B: Maximum annual wages (10-14 years after admission exam)								
Female	-0.262***	-0.262***	-0.185***	-0.185***	-0.121***	-0.121***	-0.095***	-0.095***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.005		-0.002		0.021***		0.018***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.173***	0.173***			0.099***	0.098***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.622	10.622	10.622	10.622	10.622	10.622	10.622	10.622
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the logarithm of the average or maximum annual wages between 10 and 14 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in the simpler specification, where Phase 2 scores are the response variables and Phase 1 scores are the control variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.23: Alternative Residuals: Excluding Polynomials, Log (Annual Wages) 10 to 14 Years After Admission Exam

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average annual wages (10-14 years after admission exam)								
Female	-0.270***	-0.269***	-0.196***	-0.195***	-0.122***	-0.122***	-0.095***	-0.096***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.020***		0.013***		0.027***		0.023***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.166***	0.166***			0.102***	0.101***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes
Panel B: Maximum annual wages (10-14 years after admission exam)								
Female	-0.262***	-0.261***	-0.185***	-0.185***	-0.121***	-0.122***	-0.095***	-0.096***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.018***		0.010**		0.027***		0.023***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.173***	0.172***			0.099***	0.098***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.622	10.622	10.622	10.622	10.622	10.622	10.622	10.622
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the logarithm of the average or maximum annual wages between 10 and 14 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification excluding the 'Female × Priority' interaction, and the 'ENEM × Priority' and 'Phase 1 scores × Priority' polynomials, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.24: Alternative Residuals: Adding Coefficient Estimates, Log (Annual Wages) 10 to 14 Years After Admission Exam

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average annual wages (10-14 years after admission exam)								
Female	-0.270***	-0.269***	-0.196***	-0.195***	-0.122***	-0.122***	-0.095***	-0.096***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.017***		0.016***		0.026***		0.026***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.166***	0.166***			0.102***	0.102***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes
Panel B: Maximum annual wages (10-14 years after admission exam)								
Female	-0.262***	-0.261***	-0.185***	-0.184***	-0.121***	-0.121***	-0.095***	-0.095***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.015***		0.014***		0.026***		0.026***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.173***	0.173***			0.099***	0.099***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.622	10.622	10.622	10.622	10.622	10.622	10.622	10.622
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the logarithm of the average or maximum annual wages between 10 and 14 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification adding the 'Priority' coefficients for priority subjects for male students and the 'Priority' and 'Female × Priority' coefficients for female students, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.25: Labor Market Participation 10 to 14 Years After Admission Exam

	10		11		12		13		14	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: LFP 10 to 14 years after admission exam										
Female	0.001	0.001	-0.004	-0.004	-0.007	-0.007	-0.017***	-0.017***	-0.018***	-0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Norm. ENEM scores	-0.015***	-0.015***	-0.013***	-0.013***	-0.013***	-0.013***	-0.012***	-0.012***	-0.009***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Normalized residuals		0.001		-0.001		-0.003		-0.002		-0.003
		(0.002)		(0.002)		(0.002)		(0.002)		(0.002)
Number of observations	42,275	42,275	42,275	42,275	42,275	42,275	42,275	42,275	42,275	42,275
Mean dependent variable	0.675	0.675	0.679	0.679	0.675	0.675	0.662	0.662	0.648	0.648
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variables are dummy variables that indicate if we observe the applicant at the formal labor market between 10 and 14 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. We also include exam year and program (major-university) fixed effects. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification excluding the 'Female × Priority' interaction, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Columns (1) and (2) present the labor force participation 10 years after the exam, columns (3) and (4) after 11 years, columns (5) and (6) after 12 years, columns (7) and (8) 13 years later, and columns (9) and (10) after.

Table O.26: Higher Education Degree

	6-9		10-14		Any moment	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: Higher education degree						
Female	0.029***	0.029***	0.018***	0.018***	0.014***	0.014***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Norm. ENEM scores	0.014***	0.014***	0.010***	0.010***	0.008***	0.008***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Normalized residuals		0.008***		0.003**		0.004***
		(0.002)		(0.001)		(0.001)
Number of observations	31,645	31,645	33,179	33,179	36,713	36,713
Mean dependent variable	0.888	0.888	0.945	0.945	0.942	0.942
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variables are dummy variables that equal one if the applicant earned a higher education degree. Columns (1) and (2) present the results between 6 and 9 years after the exam; columns (3) and (4) report the results between 10 and 14 years after the exam; and columns (5) and (6) display the results considering any year we observe the applicant at the formal labor market. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification excluding the 'Female × Priority' interaction, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.27: Log (Annual Wages) 6 to 14 Years After Admission Exam

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average annual wages (6-14 years after admission exam)								
Female	-0.263***	-0.262***	-0.188***	-0.187***	-0.120***	-0.121***	-0.095***	-0.095***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Normalized residuals		0.011***		0.010**		0.022***		0.022***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.168***	0.168***			0.099***	0.099***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	35,717	35,717	35,717	35,717	35,717	35,717	35,717	35,717
Mean dependent variable	10.240	10.240	10.240	10.240	10.240	10.240	10.240	10.240
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes
Panel B: Maximum annual wages (6-14 years after admission exam)								
Female	-0.266***	-0.266***	-0.191***	-0.190***	-0.130***	-0.131***	-0.105***	-0.106***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Normalized residuals		0.010**		0.009**		0.021***		0.021***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.171***	0.171***			0.097***	0.097***
			(0.004)	(0.004)			(0.005)	(0.005)
Number of observations	35,717	35,717	35,717	35,717	35,717	35,717	35,717	35,717
Mean dependent variable	10.616	10.616	10.616	10.616	10.616	10.616	10.616	10.616
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the logarithm of the average or maximum annual wages between 6 and 14 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification excluding the 'Female × Priority' interaction, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.28: Log (Annual Wages) 6 to 9 Years After Admission Exam

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Average annual wages (6-9 years after admission exam)								
Female	-0.270***	-0.269***	-0.200***	-0.199***	-0.109***	-0.110***	-0.082***	-0.083***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Normalized residuals		0.021***		0.019***		0.025***		0.026***
		(0.004)		(0.004)		(0.004)		(0.004)
Norm. ENEM scores			0.149***	0.149***			0.103***	0.103***
			(0.005)	(0.005)			(0.005)	(0.005)
Number of observations	31,859	31,859	31,859	31,859	31,859	31,859	31,859	31,859
Mean dependent variable	9.963	9.963	9.963	9.963	9.963	9.963	9.963	9.963
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes
Panel B: Maximum annual wages (6-9 years after admission exam)								
Female	-0.277***	-0.276***	-0.207***	-0.207***	-0.118***	-0.119***	-0.092***	-0.093***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Normalized residuals		0.023***		0.021***		0.026***		0.027***
		(0.005)		(0.005)		(0.005)		(0.005)
Norm. ENEM scores			0.147***	0.147***			0.101***	0.101***
			(0.005)	(0.005)			(0.006)	(0.006)
Number of observations	31,859	31,859	31,859	31,859	31,859	31,859	31,859	31,859
Mean dependent variable	10.224	10.224	10.224	10.224	10.224	10.224	10.224	10.224
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	No	No	No	No	Yes	Yes	Yes	Yes

Notes: The dependent variable is the logarithm of the average or maximum annual wages between 6 and 9 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification excluding the 'Female × Priority' interaction, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table O.29: Quantile Regressions: Log (Annual Wages) 10 to 14 Years After Admission Exam

	10		25		50		75		90	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Average wages 10 to 14 years after admission exam										
Female	-0.034	-0.035	-0.076***	-0.077***	-0.109***	-0.110***	-0.116***	-0.116***	-0.146***	-0.147***
	(0.027)	(0.027)	(0.015)	(0.015)	(0.011)	(0.011)	(0.009)	(0.009)	(0.014)	(0.014)
Norm. ENEM scores	0.131***	0.131***	0.127***	0.127***	0.101***	0.101***	0.089***	0.089***	0.091***	0.091***
	(0.016)	(0.016)	(0.009)	(0.009)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Normalized residuals		0.030**		0.032***		0.021***		0.022***		0.016***
		(0.012)		(0.007)		(0.005)		(0.004)		(0.005)
Number of observations	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129	33,129
Mean dependent variable	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396	10.396
Exam year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the logarithm of the average annual wages between 10 and 14 years after the applicant took the UNICAMP admission exam. We compute real annual wages in Brazilian 2002 *reais*. We run recentered influence function (RIF) regressions to estimate quantile regressions. We present results for the 10th (Columns (1) and (2)); 25th (Columns (3) and (4)); 50th (Columns (5) and (6)); 75th (Columns (7) and (8)); and 90th (Columns (9) and (10)) quantiles. We control for a gender dummy and ENEM scores, normalized to mean 0 and standard deviation 1 for each year. In all columns, we include program (major-university) and year fixed effects. 'Normalized residuals' is the average difference between the residuals in priority and non-priority subjects in our main specification excluding the 'Female × Priority' interaction, where Phase 2 scores are the response variables. We normalize the variable to have a mean 0 and a unitary standard deviation. Bootstrapped standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.