**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE
DEPARTAMENTO DE ECONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

Gender in Higher Education

Gênero no Ensino Superior

Bruna Pugialli da Silva Borges

Orientadora: Profa. Dra. Fernanda Gonçalves de La Fuente Estevan

**São Paulo
2021**

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. Fábio Frezatti
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. José Carlos de Souza Santos
Chefe do Departamento de Economia

Prof. Dr. Ariaster Baumgratz Chimeli
Coordenador do Programa de Pós-Graduação em Economia

BRUNA PUGIALLI DA SILVA BORGES

Tese apresentada ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito parcial para a obtenção do título de Doutor em Ciências.

Área de concentração: Teoria Econômica

Orientadora: Profa. Dra. Fernanda Gonçalves de La Fuente Estevan

Versão Corrigida

(versão original disponível na Faculdade de Economia, Administração e Contabilidade)

São Paulo

2021

Catálogo na Publicação (CIP)
Ficha Catalográfica com dados inseridos pelo autor

Borges, Bruna Pugialli Da Silva.
Gender in Higher Education / Bruna Pugialli Da Silva Borges. - São Paulo, 2021.
214 p.

Tese (Doutorado) - Universidade de São Paulo, 2021.
Orientador: Fernanda Gonçalves de La Fuente Estevan.

1. gênero. 2. Ensino Superior. 3. escolha de curso de graduação. 4. efeito de pares. 5. exame de vestibular. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

Ao meu avô Daniel, que partiu e deixou uma saudade imensa.

AGRADECIMENTOS

Em primeiro lugar, agradeço às mulheres incríveis que encontrei ao longo do meu percurso em Economia. Com certeza elas foram essenciais no meu caminho. À minha orientadora Fernanda Estevan, pelo aprendizado, apoio, e enorme dedicação ao processo de orientação. À Lígia Vasconcellos, por ter despertado meu interesse em econometria, e me guiado e ensinado ao longo da carreira. À Professora Lena Edlund, por ter me recebido em Columbia e pelas sugestões. Às professoras do grupo EconomistAs, Fabiana Rocha, Maria Dolores Diaz, Paula Pereda e Renata Narita, pela confiança e ensinamentos. À Andrea Lepine, pela ajuda ao longo do doutorado. Muito obrigada às queridas amigas economistas que fiz ao longo dos anos! Em especial, à Julia, que me acompanha na academia e vida desde o primeiro minuto da graduação, e às amigas do mestrado Isabel, Luísa e Natália.

Meu maior agradecimento é ao Tales. Eu tenho a sorte de encontrar na mesma pessoa meu amor, melhor amigo, e colega de doutorado. Sem seu apoio e confiança, eu não teria concluído esse trabalho. Obrigada por tudo. Obrigada por tanto.

Um agradecimento muito especial à minha família. Aos meus pais, Sandra e Waldemar, pelo amor e apoio incondicionais. À minha irmã Carolina, minha melhor amiga e meu suporte. Aos meus sobrinhos Sebastián e Teca, por me trazerem alegria e leveza. Agradeço também à minha segunda família, Rozenfeld. Com carinho, agradeço Cibele, Henrique, João Fernando, Márcia, Tatiana e Therezinha, por terem, literalmente, me acolhido na casa de vocês.

Obrigada aos colegas de doutorado, em especial Assaf, Bruce, Lucas e Rafa, pelo apoio e amizade. Agradeço ao Leo Rosa pelo auxílio com o Censo Escolar, e pela amizade sua e da Bárbara. Agradeço imensamente aos queridos amigos que me acompanharam e escutaram ao longo do doutorado.

Por fim, agradeço o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - bolsa 2017/23593-8) e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES - código financeiro 001). Obrigada também aos professores e demais funcionários da Universidade de São Paulo, imprescindíveis para a execução da tese. Em particular, agradeço Naercio Menezes-Filho e Ricardo Madeira, pelas sugestões dadas nas avaliações de progresso.

RESUMO

Essa tese de doutorado contém três artigos que abordam questões de gênero no Ensino Superior. No primeiro artigo, investiga-se se a subrepresentatividade feminina na graduação em Economia afeta os resultados de alunas no mercado de trabalho. Utilizam-se dados administrativos do Departamento de Economia da Universidade de São Paulo, combinados com dados do mercado de trabalho formal. Para lidar com problemas de endogeneidade, explora-se a alocação aleatória de alunos para turmas no primeiro semestre da graduação. Encontramos que maiores percentuais de colegas mulheres aumentam a participação feminina no mercado de trabalho e maiores parcelas de professoras aumentam a probabilidade de mulheres trabalharem em cargos de direção. O segundo artigo utiliza dados do vestibular da UNICAMP para verificar se há diferenças de gênero na reação ao aumento da importância de exames. No vestibular da UNICAMP todos os candidatos realizam as mesmas provas, porém na segunda fase há um peso maior para algumas disciplinas prioritárias, variando conforme o curso de graduação escolhido. Verificamos que o desempenho de mulheres decai em disciplinas prioritárias. Ao analisarmos heterogeneidades, observamos que os efeitos não variam sistematicamente de acordo com as disciplinas, porém encontramos maior redução no desempenho para as mulheres de maior habilidade. Nossa evidência também sugere como potenciais mecanismos diferenças de gênero em perfeccionismo e autoconfiança. Por fim, o terceiro artigo verifica se o percentual de colegas mulheres no Ensino Médio influencia as escolhas de curso de graduação dos alunos. Utilizando dados do vestibular da UNICAMP e do Censo Escolar, a estratégia empírica explora mudanças idiossincráticas na composição de gênero de coortes de uma mesma escola. A parcela de pares mulheres no Ensino Médio reduz a probabilidade de seleção de cursos intensivos em Matemática. Ademais, o percentual de colegas do mesmo gênero aumenta a propensão de escolha de cursos de maioria masculina e reduz a escolha de graduações com percentuais balanceados de gênero. Por fim, apenas para as mulheres, uma maior representatividade feminina no Ensino Médio gera escolhas de carreiras mais competitivas.

Palavras-chave: gênero, Ensino Superior, teto de vidro, Economia, mercado de trabalho, exames importantes, escolha de curso de graduação, efeito de pares.

ABSTRACT

This doctoral dissertation comprises three research papers on gender issues at the higher education level. In the first paper, we investigate if women underrepresentation in Economics influences female students' labor outcomes. We use administrative data from the Department of Economics at the University of São Paulo, linked with formal labor market data. To deal with endogeneity problems, we exploit the random assignment of students to sections in first-semester courses from the undergraduate program. We find that higher shares of female classmates increase female participation in the labor market. Also, higher percentages of female professors increase the likelihood of a female student working as a top manager. In the second paper, we use data from the UNICAMP admission exam to assess gender differences in reaction to an increase in the examination's stakes. At the UNICAMP admission exam, all applicants answer the same exam in both stages. Still, in Phase 2, there is a higher weight to priority subjects, defined according to the applicant's major choice. We find that women underperform in priority subjects. When we look at heterogeneous effects, we do not verify systematic heterogeneity in the impacts across subjects, but we see more substantial gender gaps for high-achieving students. The evidence suggests that gender differences in perfectionism and confidence could explain our results. Finally, the third paper examines if higher shares of female schoolmates in high school affect the students' major choices. Using UNICAMP admission data combined with data from the Brazilian School Census, the identification strategy relies on idiosyncratic changes in cohorts' gender composition within schools. The share of female peers during high school reduces the likelihood of selecting math-intensive majors. Besides, the percentage of own-gender peers increases the probability of choosing male-dominated fields and reduces the likelihood of choosing gender-balanced majors. Lastly, women exposed to higher shares of female schoolmates choose more competitive majors.

Keywords: gender, higher education, glass-ceiling, Economics, labor market, high-stakes assessments, major choice, peer effects.

LIST OF FIGURES

Figure 1.3.1 Career outcomes in the years after expected graduation, by gender	43
Figure 1.6.1 Unconditional mean performance, by student and professor gender	60
Figure 1.A.1 Share of female students admitted into the Economics major	69
Figure 1.A.2 Share of female professors in compulsory courses	70
Figure 1.A.3 Distribution of admission scores, by gender	71
Figure 3.4.1 Histogram of the proportion of female students in high-school (Year 3)	149
Figure 3.4.2 Density for changes in the proportion of females between conse- cutive years	151
Figure 3.4.3 Year-to-year variation in the proportion of female high school students	153
Figure 3.A.1 Distribution of female students' share in our sample of Brazilian high-schools (Year 3), by graduation year	187
Figure 3.A.2 Year-to-year variation in the proportion of female high school students, control for school FE and linear trend	188
Figure 3.A.3 Robustness check – Exclude majors: Impacts on the likelihood of selecting a math-intensive field	189
Figure 3.A.4 Robustness check – Exclude majors: Impacts on the likelihood of selecting a STEM major	190
Figure 3.A.5 Robustness check – Exclude majors: Impacts on the likelihood of selecting a female-dominated field	191
Figure 3.A.6 Robustness check – Exclude majors: Impacts on the likelihood of selecting a male-dominated field	192
Figure 3.A.7 Robustness check – Exclude majors: Impacts on the likelihood of selecting a gender-balanced field	193
Figure 3.A.8 Robustness check – Exclude majors: Impacts on the share of female applicants in the chosen major	194
Figure 3.A.9 Robustness check – Exclude majors: Impacts on major choice's average number of applicants per slot	195
Figure 3.A.10 Robustness check – Exclude majors: Impacts on major choice's average admission cutoff	196

LIST OF TABLES

Table 1.2.1	Randomization balance check	31
Table 1.2.2	Sorting regressions Fairlie, Hoffmann e Oreopoulos (2014) , First semester course sections	32
Table 1.2.3	Gender specific selection	33
Table 1.2.4	Students' compliance to assigned sections in compulsory courses, by undergraduate program term	34
Table 1.2.5	Balance test - Female classmates and professors	35
Table 1.3.1	Summary statistics	39
Table 1.3.2	Summary statistics: Long-term students' outcomes	42
Table 1.5.1	Long-term effects of women's representation on labor force parti- cipation	50
Table 1.5.2	Long-term effects of women's representation on occupational and industry choice	52
Table 1.5.3	Long-term effects of women's representation on career progression	54
Table 1.5.4	Long-term effects of women's representation on normalized real hourly wages (R\$ 2002 Brazilian <i>reais</i>)	56
Table 1.6.1	Impact of having an own-gender professor on first-semester course performance	58
Table 1.6.2	Effects of female classmates' share on normalized first-semester GPA	61
Table 1.6.3	Long-term effects of women's representation on academic perfor- mance	63
Table 1.6.4	Long-term effects of women's representation on elective courses' choice	65
Table 1.6.5	Long-term effects of women's representation on the probability of working during the undergraduate program	67
Table 1.A.1	Comparing students from daytime and evening streams	72
Table 1.A.2	Randomization balance check excluding 2008 admission year . .	72
Table 1.A.3	Testing whether students' characteristics correlate with their initials	73
Table 1.A.4	Comparing course-sections 1 and 2	74

Table 1.A.5 Subsample of students without individual taxpayer number (CPF) at USP data: Comparing students with or without CPF at our final data	75
Table 1.A.6 Brazilian Classification of Occupations and Industries	76
Table 1.A.7 Summary statistics: Long-term students' outcomes, years after expected graduation	77
Table 1.A.8 Share of female professors, 2SLS Regressions: First stage estimations	78
Table 1.A.9 IV estimates: Long-term effects of the predicted share of female professors on labor force participation	79
Table 1.A.10 IV estimates: Long-term effects of the predicted share of female professors on occupational and industry choice	80
Table 1.A.11 IV estimates: Long-term effects of the predicted share of female professors on career progression	81
Table 1.A.12 IV estimates: Long-term effects of the predicted share of female professors on normalized real hourly wages (R\$ 2002 Brazilian <i>reais</i>)	82
Table 1.A.13 Variation in the share of female classmates	82
Table 1.A.14 Heckman Selection Two-Step Procedure: Long-term effects of the predicted share of female professors on the probability of working in a consulting firm	83
Table 1.A.15 Heckman Selection Two-Step Procedure: Long-term effects of high-performing female classmates on the probability of working as an economist	84
Table 1.A.16 Heckman Selection Two-Step Procedure: Long-term effects of predicted female professors' share in compulsory courses on the probability of working as a top manager	85
Table 1.A.17 Heckman Selection Two-Step Procedure: Long-term effects of the predicted share of female professors on normalized real hourly wages two years after graduation	86
Table 2.3.1 Descriptive statistics	95
Table 2.4.1 Priority and P_1 subject-specific performance	103
Table 2.5.1 Priority subjects and gender performance gap	105

Table 2.5.2	Phase 2 normalized subject-specific scores: Heterogeneity	108
Table 2.5.3	Phase 2 normalized subject-specific scores: Regression results by subject	110
Table 2.5.4	Interaction term: Relative performance ENEM	112
Table 2.5.5	Alternative dependent variable: Normalized Phase 2 - Phase 1 scores	113
Table 2.6.1	Phase 2 score coefficient of variation	114
Table 2.6.2	Priority subjects and number of ‘perfect scores’	115
Table 2.6.3	Priority subjects, question’s difficulty and performance	117
Table 2.6.4	Priority subjects and within-exam performance (Early vs. Late questions)	118
Table 2.6.5	Priority subjects and number of omitted questions (2000-2002) .	120
Table 2.6.6	Priority subjects and number of zeros, excluding omitted questi- ons (2000-2002)	121
Table 2.A.1	Priority subjects by undergraduate program	125
Table 2.A.2	Priority subjects and gender performance gap (Raw Phase 2 scores)	126
Table 2.A.3	Priority subjects and gender performance gap (Including Portu- guese)	127
Table 2.A.4	Priority subjects of third and second major choices	128
Table 2.A.5	Priority, very difficult questions and P_2 performance	129
Table 2.A.6	Gender differences in Phase 2 normalized scores	130
Table 2.A.7	Probability of choosing priority subjects	131
Table 2.A.8	Only UNICAMP $\times$ Other admission processes	131
Table 3.3.1	Outcome variables	146
Table 3.4.1	Variation in the gender composition after removing school fixed- effects and linear trends	150
Table 3.4.2	Balancing test - Correlation between the share of female peers and students’ characteristics	152
Table 3.4.3	School-level autocorrelation in the percentage of female students	155
Table 3.5.1	Impact of high-school female peers on the likelihood of choosing a math-intensive field	157

Table 3.5.2	Impact of high-school female peers on the probability of selecting a male-dominated major	159
Table 3.5.3	Impact of high-school female peers on the choice of competitive majors	161
Table 3.5.4	School-Cohort FE: Impact of high-school female peers on the probability of choosing a math-intensive field	162
Table 3.5.5	School-Cohort FE: Impact of high-school female peers on the likelihood of selecting a male-dominated major	163
Table 3.5.6	School-Cohort FE: Impact of high-school female peers on the choice of competitive majors	164
Table 3.5.7	Impact of high-school female majority on the probability of choosing a math-intensive field	168
Table 3.5.8	Impact of high-school female majority on the likelihood of selecting a male-dominated major	169
Table 3.5.9	Impact of high-school female majority on the choice of competitive majors	170
Table 3.5.10	Female peers $> 66\%$ and $\leq 33\%$ dummies: Impact on the probability of choosing a math-intensive field	172
Table 3.5.11	Female peers $> 66\%$ and $\leq 33\%$ dummies: Impact on the likelihood of selecting a male-dominated major	173
Table 3.5.12	Female peers $> 66\%$ and $\leq 33\%$ dummies: Impact on the choice of competitive majors	174
Table 3.6.1	Heterogeneous impacts of high-school female peers on the probability of choosing a math-intensive field, by maternal education .	176
Table 3.6.2	Heterogeneous impacts of high-school female peers on the likelihood of selecting a male-dominated major, by maternal education	177
Table 3.6.3	Heterogeneous impacts of high-school female peers on the choice of competitive majors, by maternal education	178
Table 3.7.1	Impact of high-school female peers on performance	181
Table 3.7.2	Reasons: UNICAMP Choice	183
Table 3.7.3	Reasons: Career Choice	184
Table 3.A.1	UNICAMP data: Summary statistics	197

Table 3.A.2 Brazilian School Census: Summary statistics	198
Table 3.A.3 Performance: Summary statistics	199
Table 3.A.4 Reasons UNICAMP choice: Summary statistics	199
Table 3.A.5 Reasons major choice: Summary statistics	200
Table 3.A.6 Gender composition of field of study: Higher Education Census .	201
Table 3.A.7 Heterogeneous impacts of high-school female peers on the proba- bility of choosing a math-intensive field, by school type	202
Table 3.A.8 Heterogeneous impacts of high-school female peers on the like- lihood of selecting a male-dominated major, by school type . . .	203
Table 3.A.9 Heterogeneous impacts of high-school female peers on choice of competitive majors, by school type	204
Table 3.A.10 Impacts of high-school female peers on Phase 1 performance, by subject	205

TABLE OF CONTENTS

1 Does exposure to more women in male-dominated fields render female students more career-oriented?	23
1.1 Introduction	23
1.2 Institutional background	27
1.2.1 Students' assignment to course sections	29
1.3 Data	35
1.3.1 Variables and descriptive statistics	38
1.4 Empirical strategy	44
1.4.1 Female professors	44
1.4.2 Female classmates	47
1.5 Main results	48
1.6 Potential channels	57
1.6.1 Academic achievement	57
1.6.2 Elective courses' choice	64
1.6.3 Work during undergraduate studies	64
1.7 Conclusion	66
Appendices	69
2 Gender differences in reaction to high-stakes	87
2.1 Introduction	87
2.2 UNICAMP admission exam	90
2.3 Data	93
2.4 Empirical model	97
2.4.1 Validity	101
2.5 Main results	104
2.5.1 Heterogeneity	107
2.5.2 Robustness checks	111
2.6 Potential channels	113
2.6.1 Gender differences in exam strategy	114
2.6.2 Discussion	119
2.7 Conclusion	123
Appendices	125

3	Gender peer effects on major choice	133
3.1	Introduction	133
3.2	Institutional background	140
3.3	Data	141
3.3.1	Descriptive statistics	143
3.4	Empirical strategy	146
3.4.1	Evidence on identification strategy	148
3.5	Results	154
3.5.1	Robustness checks	160
3.5.2	Non-linear effects	166
3.6	Heterogeneous impacts	171
3.7	Mechanisms	179
3.8	Conclusion	182
	Appendices	187
	REFERENCES	207

1 Does exposure to more women in male-dominated fields render female students more career-oriented?

With *Fernanda Estevan*^{1.0.1}

1.1 Introduction

Despite women’s recent progress in labor market outcomes, women remain underrepresented among top earners (BERTRAND, 2018). Female participation in work and study environments provides role models that could indirectly modify perceptions and give women access to mentorship and specific knowledge, helping them find better career opportunities.^{1.1.1} Therefore, the exposure to successful women may improve female labor market outcomes by prompting women to aspire to high-status job positions, reducing the so-called glass ceiling.^{1.1.2}

This paper tests for gender composition effects in Economics, a historically male-dominated field of study.^{1.1.3} We investigate the impact of female peers and

^{1.0.1}We gratefully acknowledge funding from the Sao Paulo Research Foundation (FAPESP - grant # 2015/21640-3), CNPq, the British Academy, and the Newton Fund (Newton Advanced Fellowship, AF140079). Bruna received financial support from the Sao Paulo Research Foundation (FAPESP - grant # 2017/23593-8) and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES - Finance Code 001). We are indebted to FUVEST, USP’s admission office, and the Faculty of Economics, Business, and Accounting (FEA) at USP for providing the data and assistance during the project. This research has benefited from discussions with Raphael Corbi, Lena Edlund, Bruno Ferman, Claudio Ferraz, Claudia Goldin, Gabriel Leite, Andrea Lépine, Cecilia Machado, Ricardo Madeira, Naercio Menezes-Filho, Daniele Paserman, and Rodrigo Soares. We thank Viviani Kasahara for her assistance in data collection, Kelly Gonçalves dos Santos and Lucas Amici Della Rocca for providing us with additional data, and Michael L. Anderson for sending us a sample Stata code for the GLS summary index. All remaining errors are ours.

^{1.1.1}Mani e Riley (2019) review the literature on social networks and suggest that female role models and peers play a relevant role in the decision-making process, influencing women’s beliefs and aspirations.

^{1.1.2}Glass ceiling is a term to define difficulties in career advancement.

^{1.1.3}Worldwide, Economics is a research field with low women representation. In the United States, female faculty share in tenure track positions is 20%, and in Brazil, this percentage is 25% (CSWEP, 2017; BWE, 2018). In our setting, we verify low proportions of female undergraduate students (23%) and professors (18%). Besides, the female shares have not increased over the period under analysis, 2000-2008 (Figures 1.A.1 and 1.A.2).

professors on women economists' career decisions by exploiting an environment, the prestigious University of Sao Paulo (USP) in Brazil, which randomly assigns Economics students to first-semester course-sections. This setting generates an exogenous variation in classmates' and professors' gender, allowing us to address threats to internal validity, such as self-selection of students into courses and professors. By focusing on compulsory classes, we also deal with issues of gender-specific selection of students into subjects.

In addition to the low proportion of women, some studies show a hostile environment for women in Economics, reinforcing the importance of female representation in the field. [Wu \(2018\)](#) measures gender-biased language in an anonymous professional Economics forum using text-analysis techniques. She finds that while economists describe their male colleagues in professional terms, comments on women focus on their physical appearance. In Chilean universities, [Paredes, Paserman e Pino \(2020\)](#) find that not only sexist students self-select into Economics, but the training received throughout the Economics major boosts gender-biased perceptions. Importantly, they show that higher shares of female professors and peers lead to less gender-biased students' views.

Our paper estimates the long-term impacts of female peers and professors during Economics undergraduate studies on career outcomes. We combine detailed administrative data from USP admission and undergraduate records with two administrative datasets from the Brazilian federal government, a matched employer-employee formal labor market data (RAIS), and publicly available firm ownership data. Our database allows us to investigate impacts on female labor force participation rates, occupational choice, career progression, and wages up to five years after (expected) graduation.

Our results suggest that female students react to a broader women representation in Economics by becoming more career-driven. Higher shares of female classmates lead to increases in female labor force participation rates. A one standard deviation rise in female classmates' percentage results in a female labor force participation of 0.21 and 0.15 standard deviations higher two and five years after graduating. A one standard deviation higher percentage of high-performing female peers increases female labor force participation two and five years after graduating

by 0.17 and 0.11 standard deviations. Overall, we find little impact on career choice (occupation and industry).

Notably, our findings suggest that female role-models affect career advancement. In particular, female professors' share increases the probability that a female student works as a top manager after graduation: a one standard deviation increase in this share raises by 0.16 standard deviations the likelihood that a female student works as a top manager. These impacts are robust once we correct for selection into the formal labor market using a Heckman Two-Step procedure.

We then investigate the potential channels underlying our results. We start by examining the impacts of female professors and peers on performance. In general, the previous literature found that female peers ([HAN; LI, 2009](#); [OOSTERBEEK; EWIJK, 2014](#); [BOSTWICK; WEINBERG, 2018](#)), and professors ([CARRELL; PAGE; WEST, 2010](#); [GRIFFITH, 2014](#)) have positive impacts on female students' performance. We do not find similar effects of female role models on women's educational outcomes. For the average female student, the share of female classmates or high-performing peers does not impact achievement. In first-semester courses, female students perform relatively better in sections taught by female vs. male instructors, having higher grades and pass rates than their male colleagues. However, these results arise because male students perform worse in course sections taught by female professors, as in [Hoffman e Oreopoulos \(2009\)](#). In the long run, female faculty shares do not influence female students' performance. Our analysis also shows that female professors and peers do not affect elective courses' choice, the propensity to choose a female advisor ([KATO; SONG, 2018](#); [GAULE; PIACENTINI, 2018](#); [FUNK; IRIBERRI; SAVIO, 2019](#)), or to pursue an academic career in Economics ([GAULE; PIACENTINI, 2018](#); [PORTER; SERRA, 2020](#)).

Finally, we analyze if professors' and classmates' gender composition impacts female students' likelihood of working during undergraduate studies. The share of female professors does not differentially affect this job decision for women. However, if female classmates' share rises by one standard deviation, female students are 0.15 standard deviations more likely to work during their undergraduate studies. Given that most USP Economics undergraduates work or do internships in firms and banks, we can interpret this effect as an early career-oriented option. Female students

exposed to more female peers seem to focus on their future careers by working while in undergraduate studies, leading to a higher degree of labor market attachment, even though this could potentially jeopardize their academic performance [see, e.g., [Stinebrickner e Stinebrickner \(2003\)](#)].

A valuable contribution of our paper is to look at the longer-term effects of exposure to both female peers and professors on labor market outcomes. Most studies analyzing the impact of female representation in math-intensive fields concentrate on educational outcomes, such as college performance, course-taking behavior, and major choice ([HOFFMAN; OREOPOULOS, 2009](#); [CARRELL; PAGE; WEST, 2010](#); [OOSTERBEEK; EWIJK, 2014](#)).^{1.1.4} The few papers examining gender composition effects on career outcomes focus on instructor effects. [Mansour et al. \(2020\)](#) estimate the impacts of female STEM professors on occupational choice and postgraduate studies for the United States Air Force Academy students. For high-achieving female students, an increase in female professors' share results in a higher likelihood of obtaining a STEM degree, working in a STEM occupation, and earning a master's degree in a STEM field. [Kato e Song \(2018\)](#) analyze same-gender advisor impacts at a liberal arts university on undergraduate studies' outcomes (GPA and degree completion) and career outcomes (employment and graduate studies). While having an own-gender advisor positively impacts performance (graduation rates and grades), there is no impact on labor market outcomes. For lower ability students, a same-gender advisor increases the probability of enrolling in a graduate school. [Mouganie e Wang \(2020\)](#) estimate the effects of high-achieving female peers at high school in China on the decision to pursue a STEM career. Female students exposed to higher percentages of high-performing female peers are more likely to choose a STEM course-track. Like our paper, [Brodaty e Gurgand \(2016\)](#) examine peer and instructor effects in an undergraduate program in Economics from a French university. However, their article estimates ability effects on educational outcomes, while ours focus on gender composition and look at long-term results.

We organize the remainder of the paper as follows. In Section 1.2, we present background information on the undergraduate program in Economics at the

^{1.1.4}As Brazilian students choose their major in the university application process, the major choice is not directly affected by faculty or university peers.

University of Sao Paulo. Section 1.3 describes our data and the main variables used in our analysis. Section 1.4 discusses identifying assumptions and the econometric model. Section 1.5 reports our results, and Section 1.6 explores possible channels driving our effects. Section 1.7 concludes.

1.2 Institutional background

In Brazil, public (federal, state, or municipal) universities enroll approximately 25% of tertiary education students ([INEP/MEC, 2008](#)). Since they do not charge tuition fees and have a good reputation, their admission processes are typically competitive. The remaining students attend private institutions, which can be for-profit or non-profit, with varying degrees of reputation/quality.

In contrast to the U.S. higher education system, in Brazil, students choose the major when they apply for each college/university. At the end of high school, students take admission exams for each institution they are interested in attending (decentralized system).^{1.2.1} Changing majors is costly in this educational context, as students typically have to take another admission exam.^{1.2.2}

The University of Sao Paulo (USP) is a large research-intensive state university located in Sao Paulo, the country's wealthiest state. USP is a selective Brazilian university, one of the most prestigious universities in Latin America.^{1.2.3}

All students interested in attending an undergraduate program at the University of Sao Paulo must take an admission exam called *vestibular*, held once a year, administered by *Fundação Universitária para o Vestibular* (FUVEST). The admission exam has two phases, held in November/December (Phase 1) and January (Phase 2). The Economics undergraduate program offers 180 places, 90 in the daytime stream and 90 in the evening stream, so the 180 best-ranked students,

^{1.2.1}Until 2010, most public (federal and state) universities administered their admission exam. More recently, federal and some state universities joined SISU, a centralized system ([MACHADO; SZERMAN, 2016](#)).

^{1.2.2}In our sample, few students (38) transferred to other USP majors, and only two students transferred to other universities. In both cases, we cannot observe the new major or university. For students who dropped out, we have no information if they enrolled in different universities/fields.

^{1.2.3}See, for example, [Times Higher Education \(2019\)](#).

in terms of Phase 1 and 2 scores, are invited to enroll in the first admission list. USP publishes additional admission lists until there are no more remaining places. Students approved in the admission process will start to attend the university in February. Students may apply and enroll in more than one major at USP sequentially but not concurrently.

Although USP offers programs in many cities, our data is from the main campus located in Sao Paulo city, which offers the most prestigious Economics undergraduate degree. Since all Brazilian universities, including USP, also offer Business undergraduate degrees, the Economics degree at USP is highly specialized in economics courses, not including business-related disciplines as core courses. At the admission process, students selecting the Economics degree can opt for daytime or evening classes. The coursework is the same in both streams. Students have the same compulsory courses and need to earn the same amount of total credits to graduate. However, the expected duration is longer for the evening stream (four years for the daytime and five years for the evening) since they have fewer courses per semester.^{1.2.4}

Different from most U.S. universities, the coursework structure is relatively rigid. About 57% of credits are in compulsory (core) courses, and introductory courses are pre-requisites for more advanced courses, implying that students typically follow courses with their admission cohorts. Although professors from the Department of Economics teach most of the compulsory courses (73%), subjects such as Calculus, Accounting, Law, etc., are lectured by professors from other USP departments. USP hires professors through civil service examinations, and they have tenure nearly from the beginning. The university rarely relies on temporary lecturers, implying that the faculty composition changes slowly.^{1.2.5} To graduate, students also have to write a bachelor thesis under a professor's supervision.

^{1.2.4}Students admitted into the daytime and evening streams have different characteristics (Table 1.A.1). Daytime students have better admission scores, are younger, less likely to have enrolled at other USP majors before, and a higher percentage of them are female.

^{1.2.5}Over the period 2000-2013, there were 72 professors at the Department of Economics on average. Turnover was between three and four per year on average, and 46% of departures were due to retirement or death. Even temporary lecturers are subject to a civil service examination and appointed for two years.

As in other Brazilian universities, it is not uncommon for students to work during their undergraduate studies, either in internships or regular jobs. In our sample, we verify that 82% of students work at some point in their undergraduate education (60% do internships and 53% as regular workers at the formal labor market). Unlike the U.S., students typically work for firms or banks during their studies, in occupations closely related to their future careers. Since 2005, students can earn course credits for internships. Internships require a three-party contract, signed by the worker, firm, and educational institution, allowing students to work between 20 and 30 hours per week. When we analyze internship contracts from USP students between 2006 and 2016, we observe that 57% of the students are interns at banks or other financial institutions, 19% work in research institutions, 10% at consulting firms, and 11% in private firms from different sectors. As mentioned above, the internships correlate with the industry choices over their career paths: 53% of graduates work in the financial industry, 20% at consulting firms, 17% in the public sector, and 9% in research institutions. Regarding occupational choices, 28% work as economists, while the remaining occupations are typically business-related or in other fields of study.^{1.2.6}

1.2.1 Students' assignment to course sections

Students admitted into the Economics degree at USP are split into two alternative time-slots: daytime or evening classes. Each period has two sections, each containing around 45 students (hereafter, sections 1 and 2). Conditional on the period, the assignment criterion of students to course sections is alphabetical order, which mimics the random assignment of students (SCOPPA; PAOLA, 2010).^{1.2.7}

^{1.2.6}Following the Brazilian occupation classification, we categorize an individual as an economist if her principal occupation in the labor market data is economist, researcher, or professor in Economics. Students interested in business-related occupations typically opt for a Business undergraduate degree offered at USP, as explained in Section 1.2.

^{1.2.7}In all admission years but one, all Economics students were assigned to classes by alphabetical order, except those at the waiting list or who transferred from other majors or universities. Only in 2008, student allocation did not follow the alphabetical order, and the criterion used to assign students to classes is unknown. We tested for whether USP used observable characteristics to assign students to sections in 2008, like age, admission scores, admission list, etc., and rejected those alternative criteria. We present in Table 1.A.2 the randomization balance checking excluding

We test whether the first letter of a student’s given name correlates with a host of observables: gender, age, normalized admission scores, city of residence, previous experience at USP, and first admission list. Overall, results suggest orthogonality in our sample: the regressions of students’ characteristics on the first letter of their given names lead to low R-squared and F-Statistics. In general, the coefficients of the regressions are not statistically significant at a 10% level. However, we find a correlation between the first letter of a student’s name and gender, as some letters are associated with higher or lower probabilities of female names. Nevertheless, those gender differences spread across the alphabet. Table 1.A.3 reports those results.

In Table 1.2.1, we test for whether there were differences in observable pretreatment characteristics for students assigned to section 1 or 2.^{1.2.8} As students’ assignment to sections is by admission year and stream, we control for cohort-stream fixed effects. There are no statistically significant differences between students assigned to section 1 compared to section 2 for most characteristics. However, in section 1, there is a lower percentage of female students (4 percentage points). This difference relates to Brazilian female first names being less concentrated in the first half of the alphabet (LEPINE; ESTEVAN, 2020).

Students’ assignment to sections determines in which class the student attends compulsory courses. In first-semester courses, enrollment is centralized, and students cannot choose courses or professors. Since course attendance is mandatory and instructors/evaluations typically differ by course section, attending a course in the other section is unfeasible. Students may ask to change sections from the second semester on, but approvals depend on slot availability.

Thus, in the first semester, all students must take the same compulsory courses and attend lectures in randomly assigned course sections. To test whether

the 2008 admission year, and we find evidence of orthogonality, as in Table 1.2.1. We also confirm that excluding the 2008 admission cohort from our sample does not change our estimates.

^{1.2.8}In Table 1.A.4, we test for differences between course sections 1 and 2 of all mandatory courses that our sample of students attended over the undergraduate studies, for each stream. In general, there are no differences between sections, although characteristics diverge for streams. The only noticeable difference between sections 1 and 2 is that there is a higher share of female students in course sections 2.

Table 1.2.1 – Randomization balance check

	Female	Admission scores	Previous USP enrollment	SP city	Age	First admission list
Section 1	-0.040*	0.016	0.009	0.009	0.066	0.018
	(0.019)	(0.038)	(0.018)	(0.020)	(0.165)	(0.016)
Observations	1576	1576	1576	1576	1576	1576
F Statistics	4.19	0.18	0.29	0.21	0.16	1.31
Mean of dependent variable	0.23	-0.00	0.14	0.84	19.46	0.85
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents coefficients from separate OLS regressions. The dependent variables are: Female indicator; Normalized admission scores (mean zero, standard deviation one); Previous enrollment at USP dummy; Sao Paulo city of residence dummy; Student's age at admission; First admission list dummy. The key explanatory variable is a dummy variable for whether the student was assigned to section 1. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

the assignment to first-semester course sections is independent of the professor's gender, in Table 1.2.2, we follow [Fairlie, Hoffmann e Oreopoulos \(2014\)](#) and present regressions to check for differential sorting of high-achieving female students into course sections taught by female professors. The explanatory variables are a gender dummy, a female professor indicator, and the professor's and student's gender interaction. The dependent variables are students' average characteristics by course section and gender, to measure students' features related to their ability. We use the following students' characteristics: age, admission scores, and a dummy variable for whether they had previously enrolled at a USP major. In column (1), we do not insert control variables. In column (2), we add cohort fixed-effects, in column (3), we add a stream dummy variable, and in column (4), we include cohort-stream fixed-effects. The correlation between the instructor's gender and students' characteristics is small and statistically insignificant. We find no evidence of differential sorting by gender.

Moreover, to test for gender-specific selection into first-semester courses, in Table 1.2.3, we regress female students' share in the course section on a female instructor dummy variable. The professor's gender does not affect the percentage of female students enrolled in a first-semester course section.

We use the first-semester assignment to sections to determine exogenous variation in the gender composition of peers and professors and address potential selection issues. When evaluating the impacts of female classmates' share, the

Table 1.2.2 – Sorting regressions Fairlie, Hoffmann e Oreopoulos (2014), First semester course sections

	(1)	(2)	(3)	(4)
<i>Dependent variable: Student's average age</i>				
Female professor $\times$ female	-0.024 (0.171)	-0.038 (0.176)	-0.004 (0.162)	-0.018 (0.157)
Female professor	-0.048 (0.186)	-0.037 (0.196)	0.123 (0.110)	0.122 (0.102)
Female	-0.883*** (0.100)	-0.896*** (0.103)	-0.752*** (0.094)	-0.745*** (0.091)
Mean of dependent variable	20.05			
Standard deviation of dependent variable	1.09			
<i>Dependent variable: Student's average admission scores</i>				
Female professor $\times$ female	0.958 (2.561)	-1.005 (2.208)	-1.692 (2.204)	-1.684 (2.176)
Female professor	8.930 (6.928)	3.655 (3.146)	0.405 (0.917)	0.953 (0.863)
Female	1.974 (1.461)	-0.150 (1.326)	-3.073** (1.277)	-3.129** (1.243)
Mean of dependent variable	646.69			
Standard deviation of dependent variable	34.45			
<i>Dependent variable: Share of students with previous USP enrollment</i>				
Female professor $\times$ female	0.017 (0.021)	0.014 (0.021)	0.015 (0.021)	0.014 (0.020)
Female professor	0.016 (0.013)	0.011 (0.012)	0.017 (0.012)	0.014 (0.010)
Female	-0.085*** (0.011)	-0.086*** (0.011)	-0.081*** (0.011)	-0.079*** (0.011)
Mean of dependent variable	0.18			
Standard deviation of dependent variable	0.09			
# Observations	6713	6713	6713	6713
# Students	1576	1576	1576	1576
Cohort fixed-effects	No	Yes	Yes	No
Stream fixed-effects	No	No	Yes	No
Cohort-stream fixed-effects	No	No	No	Yes

Notes: We estimate OLS regressions. The dependent variables are the students' average age, admission scores and the share of students with previous enrollment at USP, by gender and course section. The explanatory variables are Female indicator; Female Professor dummy; and interaction term of the previous variables. Standard errors in parentheses, clustered at the course section level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.2.3 – Gender specific selection

	(1)	(2)	(3)
<i>Dependent variable: Share of female students</i>			
Female professor	-0.001 (0.012)	-0.010 (0.010)	-0.009 (0.010)
Daytime stream		0.074*** (0.010)	0.075*** (0.009)
Mean of dependent variable	0.23		
Observations	137	137	137
F Statistics	0.01	28.28	35.06
Cohort fixed-effects	No	No	Yes

Notes: We estimate OLS regressions. The dependent variable is the Share of female students in first-semester (compulsory) course sections. The explanatory variables are Female professor dummy; Daytime stream dummy. Robust standard errors in parentheses. P-values * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

relevant peers for us are students from the same admission-cohort and stream, assigned to the same section in the first semester. Indeed, we expect friendships to be formed right after admission, a period of intense social activity at USP.^{1.2.9} We calculate the student's share of female peers as the leave-me-out mean of female students in the assigned section. To analyze female professors' effects, we use the predicted percentage of female faculty members (we will discuss the specifications in more details in Section 1.4).

In Table 1.2.4, we analyze whether students comply with their assigned sections in compulsory courses.^{1.2.10} In the first term, as expected, students enroll in their designated course sections (99%). In the subsequent semester terms, there is a decrease in students' compliance with the assigned sections. Nevertheless, adherence to the sections remains relatively high over the undergraduate program. Overall, in more than 85% of our student-course observations, students enroll in their initially assigned section.

^{1.2.9}In particular, freshmen students do not attend classes in the first week of the term but instead participate in several activities designed to promote integration.

^{1.2.10}We exclude unified course sections from this analysis and restrict the sample to students attending classes in the regular schedule, excluding cases where the student failed a course or postponed the course enrollment to the following semesters.

Table 1.2.4 – Students’ compliance to assigned sections in compulsory courses, by undergraduate program term

<i>Term</i>	Section 1		Section 2	
	Mean	Obs	Mean	Obs
1	0.99	2681	0.99	2534
2	0.90	2065	0.89	2052
3	0.87	2064	0.78	2397
4	0.79	1477	0.79	1352
5	0.74	778	0.78	638
6	0.77	343	0.71	427
7	0.55	119	0.55	181
Total	0.88	9527	0.85	9581

Notes: This table reports the percentage of students assigned to sections 1 or 2 enrolled in the designated section in compulsory courses by undergraduate program term (semester). We exclude unified course-sections to compute these descriptive statistics (courses with a single section, where students from both sections 1 and 2 enroll). We restrict the sample to students attending classes in the regular schedule, excluding cases where they failed a course or postponed the course enrollment to the following semesters. The number of observations decreases by term as the number of compulsory (elective) decreases (increases) over the undergraduate studies’ terms. According to the ideal coursework plan, students from both streams should complete their core courses in the 7th term (semester).

To formally test the exogeneity of the share of assigned classmates and professors, in Table 1.2.5, we regress students’ characteristics on those explanatory variables, conditional on admission year and stream (cohort-stream fixed-effects). The correlation between students’ variables and the gender composition of peers and instructors is negligible for most variables and statistically insignificant in all cases. To avoid the negative mechanical correlation between female students and female classmates’ share, we also control for the percentage of female peers in the cohort-stream, as suggested by [Guryan, Kroft e Notowidigdo \(2009\)](#).

Table 1.2.5 – Balance test - Female classmates and professors

	Female	Admission scores	Previous USP enrollment	SP city	Age	First admission list
Predicted % female professors	-0.012	0.313	-0.074	0.219	-0.129	0.128
F Statistics	(0.207) 0.00	(0.293) 1.15	(0.156) 0.22	(0.127) 2.98	(1.035) 0.02	(0.115) 1.25
Share of female classmates	-0.009	-0.365	-0.039	-0.179	-1.415	-0.015
F Statistics	(0.009) -	(0.442) 0.68	(0.185) 0.04	(0.171) 1.10	(2.167) 0.43	(0.200) 0.01
% high-performing female classmates	-0.108	-0.012	0.008	-0.022	0.179	-0.089
F Statistics	(0.108) 1.00	(0.186) 0.00	(0.090) 0.01	(0.081) 0.08	(0.507) 0.12	(0.053) 2.86
Mean of dependent variable	0.23	-0.00	0.14	0.84	19.46	0.85
Standard deviation	0.42	1.00	0.34	0.37	3.09	0.36
Observations	1576	1576	1576	1576	1576	1576
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents coefficients from separate OLS regressions. The dependent variables are Female indicator; Normalized admission scores; Previous enrollment at USP dummy; Sao Paulo city of residence dummy; Student's age at admission; First admission list dummy. The key explanatory variables are the Predicted Share of female professors in compulsory courses in the first panel, the share of female classmates in the second panel, and the share of high-achieving female classmates in the third panel. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Following [Guryan, Kroft e Notowidigdo \(2009\)](#), when analyzing the correlation between the share of female classmates and the gender dummy, we also control for the share of female peers in the student's admission year and stream ("Leave-me-out share of female, urn"). In this specification, we do not include the F Statistics, which is large, as we include in the regression the "Leave-me-out share of female, urn".

1.3 Data

Our study uses detailed student, instructor, and course administrative data from the Department of Economics at the University of Sao Paulo. This dataset also includes admission data for these students.

Our data cover students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. The original sample contains 1,619 students. We exclude students who enrolled but did not attend any course from the Economics degree, which corresponds to 1.5% of the original sample. Finally, we drop 18 students with missing values on students' covariates (1% of the original sample). Our final sample consists of 1,576 students.

The university academic records contain information on course outcomes, including grades, whether the student passed the course, and credits obtained for

every class offered by and every student enrolled in Economics at the University of Sao Paulo. We match each course section to the corresponding professor’s full name, which allows us to assign the instructor’s gender based on their given name.^{1.3.1} This dataset contains several student identifiers, such as full name and date of birth and, for 65% of the sample (1,020 individuals), also the unique Brazilian tax identifier number, CPF.

We also use USP admission data that contain student-level pretreatment data. The data include information on gender, date of birth, city of residence, and students’ scores at the admission process.

We obtain additional instructor data linking our administrative data with academic career measures, collected from each instructor’s curriculum vitae, available at Lattes Platform (CNPQ, 2020).^{1.3.2} For each instructor, we extract her publications, years since Ph.D., and Ph.D. granting institution (USP, another university in Brazil, or abroad). To assess the quality of the published papers, we use the *Qualis* ranking created by the Ministry of Education (CAPES, 2020).^{1.3.3} *Qualis* ranks journals on a categorical scale (A1, A2, B1, B2, B3, B4, B5, C), where A1 defines papers published in the best journals and C papers published in the worst journals.

For labor market outcomes, we use *Relação Anual de Informações Sociais* (RAIS) from 2002 to 2017, a matched employer-employee dataset of the Brazilian formal labor market. RAIS contains information on the universe of formal employees and firms in Brazil, including individuals’ tax identifiers. We first merge RAIS with our university academic records using students’ full names and dates of birth to search for the tax identifier number for the remaining 35% of the sample (556 individuals). This procedure allows us to recover the tax identifier number for 98% of our initial sample. As shown in Table 1.A.5, the success of the merge cannot be predicted by any student predetermined characteristics such as age, admission

^{1.3.1}It is trivial to identify gender associated with most given names used in Brazil. In the few cases where there was some ambiguity, we checked the instructors’ university webpage to confirm their gender.

^{1.3.2}The Lattes platform is a nationwide platform administered by the Brazilian Government, containing all Brazilian professors’ and researchers’ academic curricula.

^{1.3.3}We use the 2013-2016 *Qualis* ranking, which is the most recent available at Sucupira Platform.

scores, stream, etc.

Using the tax identifier number, we then merge university academic records with each RAIS edition, allowing us to obtain yearly information on students' labor force participation, occupational and sector choice, career progression, and wages after graduation.

Instead of working as an employee at the formal labor market, some individuals may be partners in companies. To allow for this possibility when evaluating labor force participation rates, we include publicly-available information on firm partnership (ownership) gathered by the Brazilian Revenue Service ([Receita Federal, 2020](#)). The data contain partners' full name and six out of eleven digits of the tax identifier number, which allow us to merge it with the students' data.^{1.3.4}

To identify students working during undergraduate studies, we combine RAIS with USP administrative dataset. RAIS provides us information on all workers at the Brazilian formal labor market but does not record data on internships. To obtain information on internships, we use administrative data from the Department of Economics, indicating that the student enrolled in an 'internship' discipline.^{1.3.5}

Finally, we also obtain information on students' interest in attending a Brazilian Master's program in Economics by linking our administrative information to data from the Brazilian Association of Graduate Programs in Economics (ANPEC). As Brazil adopts a centralized admission for graduate studies in Economics, nearly all students pursuing a Master's program take the national exam. Thus, a dummy for whether the student took the ANPEC exam is a good measure of students' interest in graduate studies.

^{1.3.4}The Brazilian Revenue Service updates the partnership database roughly every three months and does not keep previous versions online. For this study, we use all the updates we could obtain, starting from December 27, 2017, until February 2, 2020. Since the datasets include partners' entry date in a company, we can build panels for the duration of partnerships, except for those no longer partners on December 27, 2017.

^{1.3.5}The creation of the internship subject took place in 2005, and students from the earlier cohorts (2000 and 2001) were only able to enroll in this course by the end of their undergraduate program. On average, students register for internships in their third year. Therefore, we restrict the sample to students admitted between 2002 and 2008 in this analysis.

1.3.1 Variables and descriptive statistics

Table 1.3.1 provides summary statistics for students and professors. Panel A presents student-level characteristics for our sample of incoming students. Twenty-three percent of the students enrolled in Economics are female. Compared with their male counterparts, female students are younger and more likely to enroll in daytime classes. A higher percentage of male students have enrolled in another USP major before admission into the Economics degree.^{1.3.6}

Panel B reports descriptive statistics for the sample of professors teaching mandatory courses to undergraduate students in Economics from 2000 to 2016. Female faculty members are more likely to have earned a Ph.D. degree from the University of Sao Paulo. Male professors are more likely to have received a Ph.D. from a foreign research institution. Differences in other instructors' characteristics are indistinguishable from zero.

Our main goal is to estimate the effects of gender composition on long-term outcomes related to career choices. When looking at career outcomes, we analyze labor force participation rates, career choice (occupation and industry), career progression, and wages after graduation. To investigate potential channels, we also evaluate the impact of female faculty members' and female classmates' percentages on academic performance, elective courses' selection, and work during undergraduate studies.

To assess labor force participation, we create dummies indicating whether the student works at the formal labor market or is a firm partner (owner). We consider that a student is part of the labor force if she is an employee in the formal sector or a firm partner.

To investigate whether female professors and peers influence career occupations and industries, we focus on those outcomes most related to Economics. We define a dummy for whether the student worked as an economist, i.e., has

^{1.3.6}Figure 1.A.3 plots the distribution of the admission exam scores, a pretreatment measure of academic performance, by gender. The distribution of admission scores is similar for both genders, although male students outperform female students in Phase 1 and female students outperform male students in Phase 2.

Table 1.3.1 – Summary statistics

	Female	Male	Difference
Panel A: Student level			
Normalized admission scores	0.02 (0.97)	-0.01 (1.01)	0.03
Age	18.86 (2.14)	19.63 (3.30)	-0.78***
Daytime classes	0.59 (0.49)	0.48 (0.50)	0.10***
First admission list	0.84 (0.36)	0.85 (0.36)	-0.01
Previous USP enrollment	0.09 (0.28)	0.15 (0.36)	-0.06**
Sao Paulo city	0.86 (0.35)	0.83 (0.38)	0.03
Sao Paulo state	0.99 (0.11)	0.99 (0.12)	0.00
N	357	1,219	1,576
	Female professor	Male professor	Difference
Panel B: Professor level (Compulsory courses)			
No Lattes CV profile	0.06 (0.23)	0.05 (0.22)	0.01
Ph.D. in Economics	0.41 (0.50)	0.58 (0.49)	-0.17
Ph.D. in other math-intensive field	0.35 (0.49)	0.28 (0.45)	0.07
Ph.D. in other fields	0.24 (0.43)	0.14 (0.34)	0.10
Ph.D. at USP	0.79 (0.41)	0.52 (0.50)	0.27**
Ph.D. abroad	0.18 (0.39)	0.42 (0.49)	-0.24**
Years since Ph.D. graduation	22.94 (9.82)	23.79 (11.30)	-0.85
# published papers	19.65 (19.76)	25.19 (25.70)	-5.54
# published A papers	2.59 (4.62)	4.86 (8.82)	-2.27
# published B papers	12.29 (13.40)	13.51 (15.62)	-1.21
# published C papers	0.59 (1.69)	0.86 (3.54)	-0.27
N	34	132	166

Notes: The table presents descriptive statistics of students' and professors' characteristics. The column "Difference" reports the coefficient of a t-test of mean differences between groups. Panel A - Student-level variables: Normalized admission scores; Student's age at admission; Daytime stream dummy; First admission list dummy; Previous enrollment at USP dummy; Sao Paulo city and state of residence dummies. Panel B - Professor-level variables: Dummies for whether the professor earned a Ph.D. degree in Economics, other math-intensive fields, or other non-math-intensive fields; Dummies for whether the professor earned a Ph.D. degree from USP or a foreign institution; Dummy for whether the professor did not have the CV registered at Lattes Platform; Years since Ph.D. graduation; Number of published papers until 2016, by *Qualis* ranking. Standard deviations in parentheses. P-values: * p<0.10, ** p<0.05, *** p<0.01.

economist, researcher, or professor in Economics as occupation, according to the students' occupation at the formal labor market, using the Brazilian Classification of Occupations (CBO - *Classificação Brasileira de Ocupações* 2002). In Panel A from Table 1.A.6, we present all CBO codes considered as economists. Besides, we use the Brazilian Classification of Industries (CNAE - *Classificação Nacional de Atividades Econômicas*) to obtain information on which sector the student works at the formal labor market. We create a dummy for the following industries that typically employ economists: financial services, consulting firms, research institutions, and the public sector (see Panel B Table 1.A.6).

Subsequently, we analyze gender composition effects on career advancement by investigating the impacts on the probability of holding managerial job positions. We look at two outcome variables: indicator variables for Top Manager and Middle Manager. We define the job position using the Brazilian Classification of Occupations. In Panel A, from Table 1.A.6, we report the respective codes.

We also examine the effects of gender composition on wages one, two, or five years after expected graduation. Our wage measure is the average hourly salary in the months that the person had an active job contract at the formal labor market, considering all contracts that the individual holds in the month. We use real hourly wages in 2002 Brazilian *reais*. We normalize wages to have an average of zero and a standard deviation of one.

We analyze the following long-term educational outcomes: the probability of obtaining a bachelor's degree in Economics at USP; time to graduation (conditional on degree completion); grade point average in compulsory courses (GPA)^{1.3.7}; the percentage of elective courses the student enrolled in the fields of Microeconomics, Macroeconomics, Finance and Humanities; a dummy variable for a female bachelor's thesis advisor; and a dummy for students' interests in graduate studies in Economics (i.e., whether they took the ANPEC exam). At the Department of Economics from USP, course grades and GPA vary from 0 to 10. The minimum course grade to pass a course is 5. We normalize the GPA, such as the variable has a mean of zero and a standard deviation of one.

^{1.3.7}As a robustness check, we run the regressions replacing GPA in core courses with GPA in all disciplines (including electives), and results remain similar.

Another intermediate outcome that we analyze is the likelihood of working during the undergraduate program. As explained above, an undergraduate can have a regular job in the labor market or undertake an internship. As an outcome variable, we include an indicator variable that equals one if the undergraduate work in the formal labor market or as an intern while at university, and zero otherwise.

Table 1.3.2 reports summary statistics for these outcome variables. The variables after graduation consider students' decisions at any moment of their career. For instance, the master's indicator variable equals one if the students applied for a Brazilian master's program in Economics at any moment (since admission into USP). Female students have a higher GPA, complete the undergraduate studies in fewer years, and have higher degree completion rates (87% for female students graduate versus 72% for the male students). Related to elective course choices, female students choose a higher percentage of Macroeconomics or Microeconomics courses. Female students also have a higher probability of having a female advisor (6 percentage points higher). There are no gender differences in the likelihood of applying for a Master's in Economics or labor force participation rates. However, female students are less likely to be firm partners and work in the public sector and more likely to work as economists. The share of students that hold managerial job positions after graduation is not different across genders.

We also evaluate the effects on labor market outcome variables (labor force participation rates and normalized wages), one, two, and five years after expected graduation.^{1.3.8} We report descriptive statistics for such outcome variables in Table 1.A.7.

Figure 1.3.1 presents career outcomes for female and male students from one to five years after the expected graduation. We verify that both participation in the Brazilian formal labor market and firm ownership increase after (expected) graduation. Although entrepreneurship rates are higher for male students five years after predicted graduation, we observe that labor force participation (the combi-

^{1.3.8}As the share of female peers and professors could affect time to graduation, we use the expected date of graduation to create those variables, instead of the actual date of graduation. The expected graduation year is when students would finish the undergraduate program if they graduate on time. Note that the expected graduation date varies by admission year and stream.

Table 1.3.2 – Summary statistics: Long-term students' outcomes

	Full sample	Female	Male	Difference
Normalized GPA compulsory courses	0.00 (1.00) [1576]	0.26 (0.84) [357]	-0.07 (1.03) [1219]	0.33***
Time to graduation (in years)	5.23 (1.19) [1189]	5.10 (1.07) [309]	5.28 (1.22) [880]	-0.18*
% elective courses - Microeconomics	0.24 (0.13) [1576]	0.26 (0.12) [357]	0.23 (0.14) [1219]	0.03**
% elective courses - Macroeconomics	0.20 (0.12) [1576]	0.22 (0.11) [357]	0.20 (0.12) [1219]	0.02**
% elective courses - Finance	0.13 (0.10) [1576]	0.13 (0.07) [357]	0.13 (0.10) [1219]	-0.00
% elective courses - Humanities	0.28 (0.21) [1576]	0.26 (0.18) [357]	0.28 (0.22) [1219]	-0.02
Work during undergraduate studies	0.87 (0.34) [1227]	0.90 (0.30) [286]	0.86 (0.35) [941]	0.04
Degree completion	0.75 (0.43) [1576]	0.87 (0.34) [357]	0.72 (0.45) [1219]	0.14***
Female advisor - Bachelor Thesis	0.16 (0.37) [1080]	0.20 (0.40) [272]	0.14 (0.35) [808]	0.06*
Applied for a Master in Economics	0.18 (0.38) [1576]	0.21 (0.41) [357]	0.17 (0.37) [1219]	0.04
Worker at the formal labor market	0.91 (0.28) [1576]	0.92 (0.27) [357]	0.91 (0.28) [1219]	0.01
Firm partner	0.35 (0.48) [1576]	0.29 (0.46) [357]	0.37 (0.48) [1219]	-0.07*
Labor force participation	0.94 (0.23) [1576]	0.95 (0.22) [357]	0.94 (0.23) [1219]	0.00
Economist	0.28 (0.45) [1441]	0.35 (0.48) [329]	0.26 (0.44) [1112]	0.09**
Financial services	0.53 (0.50) [1441]	0.50 (0.50) [329]	0.54 (0.50) [1112]	-0.04
Public sector	0.17 (0.38) [1441]	0.13 (0.33) [329]	0.19 (0.39) [1112]	-0.06*
Consulting firms	0.20 (0.40) [1441]	0.24 (0.43) [329]	0.19 (0.40) [1112]	0.04
Research institutions	0.09 (0.28) [1441]	0.08 (0.27) [329]	0.09 (0.28) [1112]	-0.01
Top manager	0.07 (0.25) [1441]	0.06 (0.24) [329]	0.07 (0.25) [1112]	-0.01
Middle manager	0.30 (0.46) [1441]	0.31 (0.46) [329]	0.30 (0.46) [1112]	0.01

Notes: The table presents descriptive statistics of students' outcome variables. All outcomes presented in the table consider variables at any point in students' careers. For example, the Economist dummy equals one if a student worked as an economist at any moment from 2002 to 2017. The column "Difference" reports the coefficient of a t-test of mean differences between groups. Variables: Normalized GPA: average grade in compulsory courses weighted by course-credits, normalized to have a mean of zero and standard deviation of one; Time to graduation: years from admission to graduation in Economics at USP, conditional on degree completion; % electives courses in Microeconomics, Macroeconomics, Finance, and Humanities: share of elective courses from each field that the student enrolled in; Work during undergraduate studies: dummy for whether the student was at the Brazilian formal labor market dataset before graduation (RAIS) or undertook an internship; Degree completion: dummy variable for whether the student majored in Economics at USP; Female advisor: dummy for whether the bachelor thesis' advisor is a female professor; Applied for a Master's in Economics dummy; Worker at the formal labor market: dummy for whether the student was at the Brazilian formal labor market dataset (RAIS). Firm partner: indicator variable equals one if a person is a firm partner/owner. Labor force participation: work at the formal labor market or own a firm. Economist: economist occupation at the formal labor market (conditional on being at the formal labor market). Sector dummies for financial services, public sector (government), consulting firms, and research institutions (CNAE codes). Top and middle manager, according to CBO 2002. Standard deviations in parentheses and the number of observations in square brackets. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

nation of formal labor work and firm partnership) is similar across genders. The percentage of students applying for master's programs in Economics is higher in the first two years following degree completion and decreases in the next years. Female students are more likely to apply for a Master's in Economics after graduating. The share of students that enroll in a master's program increases in the first three years after graduation and decreases in the following years, while the percentage of Ph.D. students steadily increases. A small percentage of students enroll in Ph.D. programs. A higher share of female students enroll in master's degrees, but the pattern is the opposite in doctorate programs.

Figure 1.3.1 – Career outcomes in the years after expected graduation, by gender

Notes: Labor force participation combines data on formal labor market participation and firm ownership (see Section 1.3 for more details). We use RAIS data for formal labor market participation, the Brazilian Revenue Service for firm ownership, ANPEC for students that applied for a Master's in Economics, and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) for Master and Ph.D. students in Brazil. We complement the Ph.D. student information with web scraped data on students enrolled in Ph.D. programs abroad collected by Santos (2020).

1.4 Empirical strategy

1.4.1 Female professors

Long-term outcomes

In our main specification, we investigate long-term effects, looking at students' outcomes after graduation. One advantage of focusing on long-term results is that professors cannot directly manipulate these outcomes. Therefore, they are more consistent with passive teacher effects (such as role-model effects) than with active teacher effects (FAIRLIE; HOFFMANN; OREOPOULOS, 2014).

Students may attend lectures in a course section other than the one initially assigned starting in the second semester. To address this potential selection issue, we use the predicted share of female instructors, i.e., the percentage of female professors the student would have if she remained in the assigned course sections as our key explanatory variable.^{1.4.1}

The econometric model to estimate the impacts of the instructor's gender is given by equation (1.4.1):

$$y_{ik} = \beta_0 + \beta_1 female_i + \beta_2 p_share_femaleprof_k + \beta_3 female_i \times p_share_femaleprof_k + \alpha_{tm} + Z_i\omega + X_k\gamma + W_p\phi + \epsilon_{ik} \quad (1.4.1)$$

where y_{ik} is the outcome variable of a student i randomly assigned to a class k (section, stream, and admission-cohort level). Our main response variables are labor force participation rates, occupational and industry choice, career development, and normalized hourly wages. To investigate potential channels, we also estimate equation (1.4.1) for the following outcome variables: graduation rates, normalized students' GPA, time to graduation, elective courses' choice (share of elective courses

^{1.4.1}As a robustness check, we use an instrumental variable estimator. The Two-Stage Least Squares (2SLS) estimations use female instructors' predicted share as an instrument for the students' actual percentage of female instructors in core courses. We report the first stage in Table 1.A.8 and the estimates of professor effects in Tables 1.A.9, 1.A.10, 1.A.11, and 1.A.12.

from Microeconomics, Macroeconomics, Finance, and Humanities fields), a dummy for a female advisor, whether the student applied for a Master' in Economics, and the probability of working during the undergraduate program. $p_share_femaleprof_k$ is the predicted share of female instructors teaching mandatory courses to section k , and $female_i$ is a gender dummy variable. Since our econometric models are at the student level, we will not include individual or course section fixed-effects. Still, we will control for a host of observables.^{1.4.2} The vector Z_i includes students' variables, X_k contains sections' characteristics, and W_p contains professors' characteristics. The vector Z_i contains Normalized admission scores, Previous USP enrollment, Student's age at admission, and Sao Paulo city of residence dummy. The variables included in X_k are Class size, Share of female classmates, Average peers' ability (Admission scores), Share of peers with previous USP enrollment, Peers' average age. Finally, in W_p , we add the following covariates: Percentage of professors with Ph.D. abroad, Professors' Experience, Share of A papers, Share of classified papers. We also control for cohort-stream fixed-effects (α_{tm}). The cohort-stream dummies are particularly relevant, since students from the daytime and evening streams differ in many observable characteristics (Table 1.A.1). The fixed-effects inclusion allows us to control for any cohort-by-stream confounders as we compare students within the same cohort and stream (time-slot). Standard errors are clustered at the cohort-stream level. We estimate the model by Ordinary Least Squares (OLS).

Our main coefficient of interest is β_3 , which measures the differential effect across female and male students of having more female versus male professors in compulsory courses. β_1 reflects average gender differences in outcome variables, and β_2 measures the effect on male students of higher shares of female instructors.

Since we test for several hypotheses simultaneously, both for outcomes and potential channels, we deal with multiple hypothesis tests using two approaches. The first is the randomization test based on [Young \(2019\)](#), which tests the null hypothesis that each treatment, i.e., gender composition of professors, classmates, and high-performing peers, does not affect any outcome. We apply this test for each

^{1.4.2}We may worry that the inclusion of professors' characteristics as control variables partially captures the impacts of higher female faculty members' shares. Reassuringly, our results are not sensitive to the inclusion of covariates. Results of the regressions without controls are available upon request.

table (group) of outcomes. The second approach consists of building a summary index of families of results that becomes the dependent variable of our specification, as suggested by [Anderson \(2008\)](#). To do so, we group findings that share a similar interpretation by computing the index by each table of results.^{1.4.3} We then construct an aggregated index by combining data on multiple outcomes weighted by their covariance so that the index provides a maximum amount of information.

First-semester courses

We also estimate own-gender instructor effects on first-semester course section level outcomes when students are randomly assigned to course sections. The main econometric model is given by:^{1.4.4}

$$y_{ic} = \delta_0 + \beta female_i \times female_instructor_c + \alpha_i + \theta_c + \epsilon_{ic} \quad (1.4.2)$$

where y_{ic} is the outcome variable for a student i in course section c , $female_i$ is a dummy variable for whether a student is female, $female_instructor_c$ is an indicator variable for whether the course section professor is female, α_i are student fixed-effects, and θ_c are course section fixed-effects. By including course section fixed-effects θ_c , we are implicitly controlling for the professor, section, course, and term fixed-effects. We cluster standard errors at the course section level. We run OLS regressions.

The coefficient of interest is β , which measures the extent to which gender gaps depend on students' assignment to a female or male instructor. We will estimate equation (1.4.2) for two different outcome variables: a dummy variable for whether the student passed the course and normalized course grades.

^{1.4.3}We do not construct the summary index for occupational and industry choice, elective course choice, or female advisor, as the definition of a 'better' outcome is unclear. We are also unable to include course duration, which is conditional on degree completion.

^{1.4.4}Similar econometric models have been widely used by the economic literature, see e.g., [Dee \(2005\)](#), [Dee \(2007\)](#), [Fairlie, Hoffmann e Oreopoulos \(2014\)](#), and [Paredes \(2014\)](#).

1.4.2 Female classmates

The random assignment of students to classmates and the use of a pre-determined peer characteristic (gender) allow us to overcome empirical challenges of peer effects measurement, such as the endogenous formation of the peer group and the reflection problem (SACERDOTE, 2001; FELD; ZÖLITZ, 2017).

To investigate gender peer effects, we estimate the following specification, both for short-term and long-term outcomes:

$$y_{ik} = \beta_0 + \beta_1 female_i + \beta_2 share_femaleclassmates_{ik} + \beta_3 female_i \times share_femaleclassmates_{ik} + \alpha_{tm} + Z_i\omega + X_k\gamma + W_p\delta + \epsilon_{ik} \quad (1.4.3)$$

where i indexes student, k sections, and p professors. y_{ik} are the response variables, $female_i$ is a gender indicator, and $share_femaleclassmates_{ik}$ is the percentage of female peers. As aforementioned, we define a student's classmates as the students assigned to the same section in first-semester courses. Indeed, course section composition is quite stable throughout the undergraduate program, as shown in Table 1.2.4, and friendships are likely to be built at the beginning of the undergraduate degree, a period of intense social activities.

The model also includes individual-level control variables Z_i and classmates' and professors' covariates X_k and W_p . We also control for the predicted share of female professors. α_{tm} are cohort-stream fixed effects.^{1.4.5} We estimate our model by Ordinary Least Squares (OLS). We cluster standard errors at the cohort-stream level.

Our primary variable of interest is the interaction term $female_i \times share_femaleclassmates_{ik}$. The coefficient β_3 indicates the differential effect of having higher shares of female classmates for women. β_2 captures overall differences in outcome variables for classes with higher shares of female peers, and β_1 measures gender differences in response variables.

^{1.4.5}Both for peers' and professors' effects, we have also run regressions allowing for nonlinear impacts. We found no evidence of nonlinearity. Results are available upon request.

The undergraduate studies and career outcomes are the same analyzed when measuring professor effects. For academic results, we analyze impacts on elective course-taking behavior, normalized GPA, degree completion, time to graduate, advisor gender, students' likelihood of applying for a Master's in Economics, and working during the undergraduate studies. After graduation, the outcomes are labor force participation, probability of working at the formal labor market or being a firm partner, occupational choice, career progression, and normalized hourly wages at the formal labor market. For first-semester courses, the dependent variable y_{ik} is the normalized first-semester GPA (average course-grade in first-semester courses, weighted by course credits).^{1.4.6} When measuring gender peer effects, we also perform the multiple hypothesis tests mentioned above.

Table 1.A.13 examines the extent of variation in gender composition that is left after removing fixed effects and controlling for other explanatory variables, following [Bifulco, Fletcher e Ross \(2011\)](#) and [Lepine e Estevan \(2020\)](#). Removing cohort-stream fixed effects reduces the total variation by 35%. The inclusion of students' and classes' covariates reduces the total variation by 45%. The inclusion of all control variables reduces the total variation by 52%. Compared to the previous studies, it seems that the remaining variation in our sample is sufficient to estimate peer effects.

We also estimate our main specification using an alternative measure of female classmates. Instead of using the share of female peers $share_femaleclassmates_{ik}$, we replace with the percentage of females among high-performing students $HP_femaleclassmates_{ik}$. We consider a student as high-performing if she ranked amongst the top 25% at the admission exam (by admission year).

1.5 Main results

We now investigate the impact of higher shares of female faculty members and classmates on future career choices, looking at outcomes after graduation. First,

^{1.4.6}Students will have the same classmates in all course sections. Therefore, when estimating gender peer effects, we cannot use equation (1.4.2), as our key explanatory variable will not vary by course section.

we analyze the effects of the gender composition on female labor force participation rates, and then we look at occupational and industry choice, career progression, and wages. As explained in Section 1.3, our labor force participation measure combines formal labor market participation with firm partnership data. Besides looking at labor force participation at any point in time, we also evaluate it one, two, or five years after expected graduation.

Panel A from Table 1.5.1 provides the results from estimating the effect of professor gender, measured by the proportion of mandatory courses taught by female faculty, on labor force participation. In general, we find no statistically significant impact on female students' labor force participation. Most coefficients are positive, but none of them is statistically significant at a 5% level. Both the Summary Index Test and the Westfall-Young suggest no professors' effects on female labor force participation.

Table 1.5.1 (Panel B) also displays estimates of the impacts of female classmates' share on female labor force participation rates. Higher percentages of female peers increase the likelihood that a female student works at the formal labor market two and five years after expected graduation. If female peers' share increases by seven percentage points (one standard deviation), female students relatively increase labor force participation rates by ten percentage points two years after expected graduation (0.21 standard deviations) and six percentage points after five years (0.15 standard deviations). The coefficient measuring effects on female labor participation in the year after students' expected graduation is also positive but smaller and not statistically significant.^{1.5.1} Our multiple inference tests (summary index and randomization-t p-values) confirm that female classmates have a statistically significant impact on female labor force participation.

In Panel C from Table 1.5.1, we verify that high-performing classmates' gender composition positively impacts female labor force participation two and five years after graduation. An increase by one standard deviation (12 percentage points)

^{1.5.1}This pattern may emerge because students, on average, take longer to graduate than the coursework schedule recommends. The predicted duration of the undergraduate program in Economics is four years for the daytime and five for evening classes. Our sample's average time to graduation is five years for daytime students and 5.6 years for evening students.

Table 1.5.1 – Long-term effects of women’s representation on labor force participation

	Formal	Formal 1 +	Formal 2 +	Formal 5 +	Work	Work 1 +	Work 2 +	Work 5 +
Panel A: Predicted share of female professors in compulsory courses								
Female × Predicted % female professors	-0.154 (0.177)	0.427 (0.354)	0.474* (0.243)	0.146 (0.235)	0.016 (0.152)	0.358 (0.438)	0.449 (0.347)	0.427* (0.213)
Predicted % female professors	0.111 (0.158)	-0.126 (0.326)	0.201 (0.299)	0.206 (0.328)	0.145 (0.115)	-0.259 (0.247)	0.111 (0.217)	0.132 (0.279)
Female	0.044 (0.030)	-0.063 (0.052)	-0.061 (0.062)	-0.001 (0.047)	0.006 (0.023)	-0.061 (0.062)	-0.050 (0.068)	-0.063 (0.044)
Observations	1576	1576	1576	1576	1576	1576	1576	1576
Summary Index Test (p-value)	0.297							
Westfall-Young multiple testing (randomization-t p-value)	0.282							
Panel B: Share of female classmates								
Female × % female classmates	0.363 (0.307)	0.290 (0.432)	1.419*** (0.460)	1.164** (0.402)	0.312 (0.270)	0.164 (0.407)	1.407*** (0.438)	0.896** (0.382)
Share of female classmates	0.114 (0.152)	0.036 (0.282)	-0.366* (0.193)	-0.354 (0.296)	0.006 (0.130)	0.015 (0.198)	-0.456** (0.161)	0.026 (0.250)
Female	-0.066 (0.039)	-0.052 (0.109)	-0.297** (0.120)	-0.239** (0.103)	-0.062 (0.064)	-0.034 (0.101)	-0.288** (0.117)	-0.190* (0.097)
Observations	1576	1576	1576	1576	1576	1576	1576	1576
Summary Index Test (p-value)	0.039							
Westfall-Young multiple testing (randomization-t p-value)	0.027							
Panel C: Share of high-performing female classmates								
% high-performing female classmates × female	0.292* (0.156)	0.651* (0.320)	0.792*** (0.247)	0.596** (0.211)	0.240 (0.142)	0.439 (0.287)	0.639*** (0.208)	0.366* (0.190)
% high-performing female classmates	-0.044 (0.079)	-0.078 (0.128)	-0.255* (0.121)	-0.274* (0.154)	-0.047 (0.065)	-0.026 (0.088)	-0.190* (0.100)	-0.071 (0.125)
Female	-0.043 (0.039)	-0.112 (0.067)	-0.127* (0.064)	-0.089* (0.049)	-0.039 (0.035)	-0.081 (0.056)	-0.089 (0.058)	-0.060 (0.046)
Observations	1576	1576	1576	1576	1576	1576	1576	1576
Summary Index Test (p-value)	0.008							
Westfall-Young multiple testing (randomization-t p-value)	0.034							
Mean of dependent variable	0.91	0.52	0.64	0.70	0.94	0.56	0.69	0.79
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students’ covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Classes’ covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Professors’ covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me-out counts). High-performing: Top quartile (highest 25% admission scores). The dependent variables are Formal: formal labor market participation dummy, and Work: labor force participation, also considering firm ownership. 1+, 2+, 5+ state that the variable refers to one, two, or five years after expected graduation. The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2000 and 2008. Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (Admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

in the share of high-performing female classmates results in an eight percentage point higher labor force participation (0.17 standard deviations) two years after graduation and a four percentage point larger participation (0.11 standard deviations) five years later. The multiple hypothesis tests reject the null hypothesis of no impacts of high-performing female peers at a 5% significance level (p-values 0.008 for summary index test and 0.034 for the randomization-t).

We also analyze whether women underrepresentation in Economics influences female students' occupations and activity in the formal labor market in Table 1.5.2. We only have information on students' occupations and sectors conditional on their Brazilian labor market participation. Therefore, the sample used in regressions presented in Table 1.5.2 is restricted to students that we observe at the administrative formal labor market data (RAIS dataset). Panel A presents the impacts of the share of female faculty members on career choice. Overall, we find little influence of female professors on women's occupation and activity choices. However, female instructors' share leads to a higher probability that a female student works at a consulting firm. If the female instructors' share increases by ten percentage points (one standard deviation), women are nine percentage points (0.21 standard deviations) more likely to work at a consulting company. However, the joint significance p-value does not reject the null hypothesis of the irrelevance of female professors' effects on occupational and industry choice at a 10% level.

Information on students' activity is only available for students working in the formal labor market (i.e., on RAIS data). As the impact of female faculty shares on labor force participation is limited, sample selection probably does not influence our estimates. However, we deal with the possible sample selection by running the regressions using a Heckman Two-Step Procedure. The instrument we use in the selection equation is the formal employment rate of higher education workers at Sao Paulo city by gender-age-graduation year cell.^{1.5.2} We present results correcting

^{1.5.2}To construct the instrument, we merge formal labor market data (RAIS), the Brazilian census, and municipal population estimates. We use RAIS data from Sao Paulo city between 2002 and 2017 to obtain information on the number of formal workers with a higher education degree, segmented by gender, age, and year. We use Census data to gather information on the city's percentage of people with a higher education degree by gender and age. As we only have Census data each decade, we used population estimates to obtain a yearly number of citizens with a higher education diploma (we multiply the shares by the population sizes). Finally, we

Table 1.5.2 – Long-term effects of women’s representation on occupational and industry choice

	Occupation	Industry			
	Economist	Finance	Government	Consulting	Research
Panel A: Predicted share of female professors in compulsory courses					
Female × Predicted % female professors	-0.091 (0.577)	0.039 (0.303)	0.174 (0.258)	0.862** (0.383)	-0.234 (0.166)
Predicted % female professors	-0.263 (0.329)	0.151 (0.109)	-0.166 (0.188)	0.075 (0.289)	0.201 (0.256)
Female	0.090 (0.118)	-0.050 (0.073)	-0.074 (0.048)	-0.116 (0.081)	0.037 (0.035)
Observations	1441	1441	1441	1441	1441
Westfall-Young multiple testing (randomization-t p-value)	0.175				
Panel B: Share of female classmates					
Female × % female classmates	0.723 (0.521)	0.606* (0.347)	-0.373 (0.371)	-0.014 (0.674)	-0.161 (0.267)
Share of female classmates	0.456 (0.360)	0.292 (0.219)	-0.201 (0.201)	0.444 (0.308)	-0.070 (0.212)
Female	-0.091 (0.115)	-0.181** (0.084)	0.042 (0.081)	0.041 (0.150)	0.032 (0.067)
Observations	1441	1441	1441	1441	1441
Westfall-Young multiple testing (randomization-t p-value)	0.409				
Panel C: Share of high-performing female classmates					
% high-performing female classmates × female	0.712*** (0.233)	0.008 (0.194)	-0.206 (0.176)	-0.405 (0.349)	0.053 (0.117)
% high-performing female classmates	-0.135 (0.154)	0.077 (0.089)	0.076 (0.087)	0.053 (0.175)	-0.186** (0.080)
Female	-0.074 (0.054)	-0.050 (0.044)	0.002 (0.046)	0.110 (0.086)	-0.015 (0.022)
Observations	1441	1441	1441	1441	1441
Westfall-Young multiple testing (randomization-t p-value)	0.038				
Mean of dependent variable	0.28	0.53	0.17	0.20	0.09
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes
Students’ covariates	Yes	Yes	Yes	Yes	Yes
Classes’ covariates	Yes	Yes	Yes	Yes	Yes
Professors’ covariates	Yes	Yes	Yes	Yes	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me out counts). High-performing: Top quartile (highest 25% admission scores). The dependent variables are occupation dummies for economists, sector dummies for financial services, public sector (government), consulting firms, and research institutions. The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2000 and 2008, conditional on working in the formal labor market (we only observe occupational choice for students who have RAIS information). Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

for selection in Table 1.A.14. Heckman estimates suggest lower coefficients, but the impact is still statistically significant at a 10% significance level.

Panel B in Table 1.5.2 presents estimates of female peer effects on career choice. We find that the share of female classmates has no impact on female students' choice of occupation or industry, and we observe that the joint-significance test confirms the absence of impacts. Nevertheless, we find significant effects of high-performing peers' gender composition on career choice (Panel C). A 12 percentage higher percentage of high-ability peers (one standard deviation) increases by 8.5 percentage points the probability of a female student working as an economist (0.19 standard deviations). The joint significance randomization-t p-value rejects at a 5% significance level the hypothesis of no effects of high-performing female peers on occupational and sector choice. We checked if those results are robust to sample selection corrections (results shown in Table 1.A.15). Coefficients are larger when we use the Heckman Two-Step Procedure.

To assess the effects of higher women's representation on career progression, we look at the probability of working as a top or middle manager at the formal labor market. As in the previous table, our sample is restricted to students observed in the formal labor market. Results shown in Panel A from Table 1.5.3 indicate that higher female faculty shares increase female students' likelihood of working as top managers. A ten percentage points higher share of female professors (one standard deviation) increases by four percentage points the likelihood of working as a top manager (0.16 standard deviations). The joint significance and summary index tests reinforce that female professors influence women's career progression (p-values 0.034). Results remain similar after correcting for sample selection, as shown in Table 1.A.16. There is no impact of female classmates on women's career advancement (Panels B and C).

We also evaluate whether female professors and classmates impact future wages. We show results in Table 1.5.4. As explained in Section 1.3, we normalize real hourly wages (2002 prices in Brazilian *reais*), such as the wage variables have an average of zero and a standard deviation of one. We find no impact of the gender

divide the number of formal workers by population in the gender-age-year cell to obtain formal workers' share.

Table 1.5.3 – Long-term effects of women’s representation on career progression

	Top Manager	Middle Manager
Panel A: Predicted share of female professors in compulsory courses		
Female × Predicted % female professors	0.400** (0.154)	0.296 (0.289)
Predicted % female professors	-0.251 (0.228)	-0.198 (0.241)
Female	-0.075** (0.033)	-0.037 (0.062)
Observations	1441	1441
Summary Index Test (p-value)	0.034	
Westfall-Young multiple testing (randomization-t p-value)	0.034	
Panel B: Share of female classmates		
Female × % female classmates	0.229 (0.216)	0.279 (0.499)
Share of female classmates	-0.168 (0.270)	0.305 (0.197)
Female	-0.056 (0.060)	-0.048 (0.124)
Observations	1441	1441
Summary Index Test (p-value)	0.314	
Westfall-Young multiple testing (randomization-t p-value)	0.518	
Panel C: Share of high-performing female classmates		
% high-performing female classmates × female	-0.015 (0.112)	0.041 (0.292)
% high-performing female classmates	-0.053 (0.097)	0.201 (0.119)
Female	0.001 (0.027)	0.003 (0.061)
Observations	1441	1441
Summary Index Test (p-value)	0.978	
Westfall-Young multiple testing (randomization-t p-value)	0.986	
Mean of dependent variable	0.07	0.30
Cohort-stream fixed-effects	Yes	Yes
Students’ covariates	Yes	Yes
Classes’ covariates	Yes	Yes
Professors’ covariates	Yes	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me-out counts). High-performing: Top quartile (highest 25% admission scores). The dependent variables are occupation dummies for top manager and middle manager. The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2000 and 2008, conditional on working in the formal labor market (we only observe occupational choice for students who have RAIS information). Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

composition of classmates or high-performing female peers on women’s relative real hourly wages after graduation. The predicted percentage of female instructors in compulsory courses does not impact hourly wages one or five years after graduation. Still, two years after, there is a positive and statistically significant impact on wages. An increase of ten percentage points (one standard deviation) results in a 0.2 standard deviation increase in normalized hourly wage. The multiple inference tests suggest that female professors’ share impacts hourly wages (significant at a 5% level). Besides, once we consider selection into the formal labor market, using a Heckman Two-Step Procedure, the coefficients are similar and still statistically significant at a 5% level (Table 1.A.17).^{1.5.3}

Finally, we estimate heterogeneity effects for female students with different initial ability levels for our main response variables (results available upon request).^{1.5.4} Previous studies, such as [Hoffman e Oreopoulos \(2009\)](#), [Carrell, Page e West \(2010\)](#), and [Bottia et al. \(2015\)](#), found that same-gender instructor effects are larger for high-achieving students. While female professors’ share particularly influences top quartile female students to become top managers, our labor force participation effects are concentrated in the middle of the ability distribution.

In sum, a higher representation of women in a typically male-dominated environment, an Economics undergraduate program, influences female students in the job market. We observe that both an increase in the percentage of female classmates and high-performing female peers lead to higher female labor force participation rates. Moreover, female students who had higher shares of female instructors in core courses are more likely to be top managers in the formal labor market.

^{1.5.3}For the selection equation, we use a dummy variable for whether the student had information on the individual taxpayer number. As we explain in Section 1.3, for students that we do not have information on the individual taxpayer number, i.e., CPF, we used the full name and birth dates to link our administrative records to RAIS data, which reduces the probability of merging them.

^{1.5.4}To do so, we expand equations (1.4.1) and (3.4.1) to allow for heterogeneous impacts. We include a triple interaction term with new dummy variables *Top_25%_Admission_score_i* and *Bottom_25%_Admission_score_i*, which indicate if scores are in the top or bottom quartile of the admission-year scores.

Table 1.5.4 – Long-term effects of women’s representation on normalized real hourly wages (R\$ 2002 Brazilian *reais*)

	Norm. Wages +1	Norm. Wages +2	Norm. Wages +5
Panel A: Predicted share of female professors in compulsory courses			
Female × Predicted % female professors	0.892 (0.962)	2.126** (0.767)	0.680 (0.630)
Predicted % female professors	-0.274 (0.757)	-1.071 (0.694)	-0.360 (0.439)
Female	-0.144 (0.153)	-0.359** (0.133)	-0.078 (0.120)
Observations	787	992	1082
Summary Index Test (p-value)	0.032		
Westfall-Young multiple testing (randomization-t p-value)	0.027		
Panel B: Share of female classmates			
Female × % female classmates	-0.220 (1.154)	0.372 (0.869)	-0.208 (0.938)
Share of female classmates	0.831 (1.092)	0.434 (1.014)	0.004 (0.549)
Female	0.058 (0.254)	-0.058 (0.215)	0.091 (0.255)
Observations	787	992	1082
Summary Index Test (p-value)	0.611		
Westfall-Young multiple testing (randomization-t p-value)	0.958		
Panel C: Share of high-performing female classmates			
% high-performing female classmates × female	-0.183 (0.805)	0.286 (0.622)	-0.588 (0.483)
% high-performing female classmates	0.080 (0.418)	-0.104 (0.411)	-0.239 (0.235)
Female	0.034 (0.178)	-0.036 (0.153)	0.157 (0.157)
Observations	787	992	1082
Summary Index Test (p-value)	0.237		
Westfall-Young multiple testing (randomization-t p-value)	0.517		
Mean of dependent variable	0.00	-0.00	0.00
Std. dev. of dependent variable	1.00	1.00	1.00
Cohort-stream fixed-effects	Yes	Yes	Yes
Students’ covariates	Yes	Yes	Yes
Classes’ covariates	Yes	Yes	Yes
Professors’ covariates	Yes	Yes	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me-out counts). High-performing: Top quartile (highest 25% admission scores). The dependent variable is the Normalized Average Hourly Wages, considering all labor contracts. We normalize the variable to have an average of zero and standard deviation of one. Wages’ information is conditional on working in the formal labor market. 1+, 2+, 5+ state that the variable refers to one, two, or five years after expected graduation. The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2000 and 2008, conditional on working in the formal labor market one, two or five years after predicted graduation. Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

1.6 Potential channels

In this section, we investigate possible channels that could be driving our estimated effects. First, we analyze whether female students' academic performance could explain gender differences in labor market outcomes. In this analysis, we evaluate impacts on first-semester academic performance and throughout the undergraduate program. We also assess peers' and professors' effects on the likelihood of choosing a female advisor and the probability of pursuing graduate studies in Economics. Second, we evaluate if the share of female peers and professors influence female students' elective course choice. Finally, we check if women's representation influences the decision to work while at university.

1.6.1 Academic achievement

In this subsection, we analyze how the gender composition of professors and classmates affects female students' performance. First, we will present results for having a same-gender instructor on first-semester course performance. We estimate same-gender instructor effects on first-semester course performance, a sample that consists of 6,713 student-course section observations. We will use two student outcome variables at the student-course-section level: normalized course grades and a dummy for whether the student passed the course.

Table 1.6.1 presents the estimates of gender interaction effects on normalized course grades and pass rates. In column (1), we include only the variables on which the randomization is conditional, i.e., cohort-stream fixed-effects. In column (2), we add students' characteristics as explanatory variables. In column (3), we replace students' characteristics with individual fixed-effects. In column (4), we add professor fixed-effects. Finally, in column (5), we simultaneously include student and course section fixed-effects.

In the first panel, we present own-gender instructor impacts on course grades. The female professor dummy shows that students, on average, have lower grades in courses taught by female faculty members (0.3 standard deviation lower

Table 1.6.1 – Impact of having an own-gender professor on first-semester course performance

	(1)	(2)	(3)	(4)	(5)
<i>Dependent variable: Normalized course grade</i>					
Female professor × female	0.067 (0.060)	0.086 (0.059)	0.108* (0.055)	0.106** (0.041)	0.097** (0.040)
Female professor	-0.298*** (0.078)	-0.304*** (0.077)	-0.283*** (0.089)		
Female	0.177*** (0.037)	0.173*** (0.036)			
Mean of dependent variable	-0.00				
Stand. dev. of dependent variable	1.00				
	(1)	(2)	(3)	(4)	(5)
<i>Dependent variable: Pass rate</i>					
Female professor × female	0.031* (0.017)	0.034** (0.017)	0.042*** (0.015)	0.043*** (0.016)	0.041*** (0.015)
Female professor	-0.059*** (0.018)	-0.060*** (0.017)	-0.055*** (0.021)		
Female	0.036*** (0.008)	0.030*** (0.008)			
Mean of dependent variable	0.91				
Stand. dev. of dependent variable	0.29				
Observations	6713	6713	6713	6713	6713
Number of students	1576	1576	1576	1576	1576
Cohort-stream fixed-effects	Yes	Yes	No	No	No
Professor fixed-effects	No	No	No	Yes	No
Course section fixed-effects	No	No	No	No	Yes
Student fixed-effects	No	No	Yes	Yes	Yes
Students' covariates	No	Yes	No	No	No

Notes: We estimate OLS regressions. The dependent variables are Normalized course grades and Pass rate dummy for whether students passed a course. We normalize the course grade, such as the variable has an average of zero and standard deviation of one. The regressions include the sample of students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. The individual fixed-effect absorbs admission cohort-stream fixed-effects. Students' covariates: Normalized admission scores, Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy. Standard errors in parentheses, clustered at the course section level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

in column (2)).^{1.6.1} Besides, we note that female students have course grades 0.17 standard deviation higher than male students, on average. Once we include student fixed-effects, the interaction term indicates female students relatively gain in course-sections taught by female vs. male instructors. In our preferred specification (column (5)), we verify that female students have a 0.1 standard deviation higher grade in course sections with female instructors. Conversely, male students have a 0.1 standard deviation lower achievement in courses taught by female vs. male professors.

In the second panel from Table 1.6.1, we report the effects of female professors on pass rates. Female students have pass rates, on average, 3 to 4 percentage points higher than male students, and course sections taught by female instructors have pass rates six percentage points lower. Female students have relatively higher pass rates in course sections with female instructors: their likelihood of passing a first-semester course is 4.1 percentage points higher than male rates when a female rather than a male professor teaches the course (0.14 of a standard deviation).

We verify that the own-gender professor impacts are driven by male students performing relatively worse in courses taught by female instructors, rather than by an increase in female students' performance [results similar to [Hoffman e Oreopoulos \(2009\)](#)]. Figure 1.6.1 reports unconditional mean performance in courses by student and professor gender. Female students outperform male students, regardless of instructor gender. Students have lower performance in first-semester courses taught by female professors. However, this decrease in achievement is larger for male students.^{1.6.2}

^{1.6.1}Our empirical analysis shows that students from both genders have lower performance in courses taught by female instructors. In Table 1.3.1, we verify that female and male professors do not differ in many observable variables. Thus, differences in their characteristics do not seem to drive the results. At the Department of Economics from the University of Sao Paulo, the exams will differ by course section, and grading is not on a curve. A possible explanation is that female instructors hold students to higher standards than their male colleagues. Another possibility is the existence of students' behavioral responses to the professor's gender.

^{1.6.2}The pattern shown in the figure is qualitatively similar to additional regression estimates available upon request. We estimate the impacts of female professors for female and male students separately (in this specification, we include individual fixed-effects, analogous to column (3) from Table 1.6.1). We find that having a female professor in a first-semester course reduces male students' pass rates but has no impact on female students' rates. Female professors decrease the course grade for both genders, but the effect is more substantial for male students.

Figure 1.6.1 – Unconditional mean performance, by student and professor gender

Notes: Table displays information on students' average performance in first-semester courses. Normalized first-semester course grade is the normalization of course grade (0 to 10 scale) to have an average of zero and standard deviation of one and first-semester pass rate is the average proportion of courses students passed.

To examine the impact of the percentage of female classmates on average performance in first-semester courses, Table 1.6.2 presents estimates of model (1.4.3). In column (1), we control only for cohort-stream dummies. Column (2) includes students' characteristics. In column (3), we control for sections' and professors' characteristics. For the average student, the share of female classmates does not impact the first-semester GPA. In Panel B, we show estimates of the effects of females' share among high-performing peers (top admission scores quartile) on female students' first-semester GPA. Likewise, we find no impact of high-performing classmates' gender composition on students' performance in the first semester courses.

In Table 1.6.3, we present estimates of the impact of women representation in Economics on longer-term academic outcomes, looking at students' results

Table 1.6.2 – Effects of female classmates' share on normalized first-semester GPA

	(1)	(2)	(3)
Panel A: Share of female classmates			
Female $\times$ % female classmates	-0.586 (0.636)	-0.279 (0.582)	0.037 (0.577)
Share of female classmates	-0.335 (0.921)	-0.414 (0.804)	0.336 (0.877)
Female	0.366** (0.171)	0.297* (0.148)	0.251 (0.151)
Observations	1576	1576	1576
	(1)	(2)	(3)
Panel B: Share of high-performing female classmates			
% high-performing female classmates $\times$ female	-0.145 (0.420)	0.134 (0.472)	0.404 (0.378)
% high-performing female classmates	0.237 (0.482)	0.187 (0.480)	0.341 (0.344)
Female	0.267*** (0.092)	0.215** (0.087)	0.181** (0.071)
Observations	1576	1576	1576
Mean of dependent variable	-0.00		
Std.dev. of dependent variable	1.00		
Cohort-stream fixed-effects	Yes	Yes	Yes
Students' covariates	No	Yes	Yes
Classes' covariates	No	No	Yes
Professors' covariates	No	No	Yes

Notes: We estimate OLS regressions. In panel B, Share of high-performing female classmates = # female high-performing among peers $\div$ # high performing among peers (leave-me out counts). High-performing: Top quartile (highest 25% admission scores). The dependent variable is the GPA in first-semester courses, normalized to have a mean of zero and standard deviation of one. The regressions include the sample of students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. Students' covariates: Normalized admission scores, Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy, First admission list dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

throughout their undergraduate studies and after graduation. The undergraduate program variables are a dummy variable for degree completion (whether the student obtained a bachelor's degree in Economics), cumulative GPA, the duration of the undergraduate program, the gender of the bachelor thesis' advisor, and the probability of applying for a Master's in Economics. We control for cohort-stream fixed-effects and covariates for students, classes, and professors in all regressions.

In Panel A, we show the impact of the female faculty members' shares on academic performance. Female students are 17 percentage points more likely to graduate with an Economics degree and have a 0.27 standard deviation higher GPA. The share of female professors in core courses does not differentially impact long-term academic outcomes for female students. The percentage of female faculty members increases degree completion. A ten percentage points higher female instructors' share leads to a four percentage points higher graduation rate. Moreover, the percentage of female professors in compulsory courses does not affect female students' willingness to pursue a Master's in Economics. The share of female professors seems to reduce the probability of applying for a Master in Economics, but this coefficient is only marginally significant (at a 10% level). Furthermore, we cannot reject at a 10% significance level the hypothesis of no female professors' impacts on educational outcomes.

Panel B in Table 1.6.3 shows the impacts of female classmates' share on students' performance. The percentage of female peers decreases time to graduation for all students, but there is no additional effect for female students. We find no impacts of female classmates' share for the average female student for the other outcome variables. The percentage of high-performing female peers does not affect female students' educational outcomes differentially (Panel C).^{1.6.3}

Our results suggest that an increase in female representation in Economics does not improve women's performance throughout their undergraduate program. Female students have relatively better achievement in first-semester courses taught

^{1.6.3}We pooled the variables normalized GPA, degree completion, and applied for a Master dummy, and conducted a summary index test. The summary index tests suggest that the gender composition of professors, classmates, and high-achieving classmates does not impact performance in undergraduate studies (p-values 0.510, 0.191, and 0.358, respectively).

Table 1.6.3 – Long-term effects of women’s representation on academic performance

	Completion	Norm. GPA	Duration	Female advisor	Master
Panel A: Predicted share of female professors in compulsory courses					
Female × Predicted % female professors	-0.397 (0.243)	0.005 (0.701)	-0.170 (1.013)	0.009 (0.420)	0.002 (0.229)
Predicted % female professors	0.439*** (0.138)	0.146 (0.378)	0.753 (0.446)	-0.098 (0.184)	-0.324* (0.178)
Female	0.174*** (0.036)	0.267* (0.128)	-0.167 (0.166)	0.060 (0.071)	0.047 (0.050)
Observations	1576	1576	1189	1080	1576
Westfall-Young multiple testing (randomization-t p-value)	0.441				
Panel B: Share of female classmates					
Female × % female classmates	-0.114 (0.315)	0.589 (0.758)	-0.884 (1.185)	0.111 (0.563)	0.670* (0.320)
Share of female classmates	0.054 (0.248)	0.672 (0.468)	-3.836*** (0.374)	-0.019 (0.209)	0.287 (0.175)
Female	0.129 (0.083)	0.134 (0.218)	0.005 (0.293)	0.036 (0.131)	-0.105 (0.082)
Observations	1576	1576	1189	1080	1576
Westfall-Young multiple testing (randomization-t p-value)	0.236				
Panel C: Share of high-performing female classmates					
% high-performing female classmates × female	0.126 (0.198)	0.626 (0.532)	-0.882 (0.573)	0.179 (0.252)	0.135 (0.225)
% high-performing female classmates	-0.040 (0.155)	0.399 (0.329)	-1.231*** (0.377)	0.052 (0.100)	0.068 (0.090)
Female	0.078 (0.047)	0.140 (0.130)	0.027 (0.142)	0.027 (0.055)	0.015 (0.050)
Observations	1576	1576	1189	1080	1576
Westfall-Young multiple testing (randomization-t p-value)	0.529				
Mean of dependent variable	0.75	0.00	5.23	0.16	0.18
Std. dev. of dependent variable	0.43	1.00	1.19	0.37	0.38
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes
Students’ covariates	Yes	Yes	Yes	Yes	Yes
Classes’ covariates	Yes	Yes	Yes	Yes	Yes
Professors’ covariates	Yes	Yes	Yes	Yes	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me out counts). High-performing: Top quartile (highest 25% admission scores). The dependent variables are Degree completion rates, Normalized GPA in compulsory courses, Duration: Time to graduation (conditional on obtaining a bachelor degree), Female advisor: Bachelor thesis’ supervisor is a woman, Master dummy: applied for a Master’s in Economics at any moment. The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2000 and 2008. Duration and female advisor dummy are conditional on degree completion. Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

by female instructors, but a comparatively worse male performance drives the result. In the long term, female professors' share does not impact grades and degree completion rates for female students. Likewise, there is no female performance increase when assigned to classes with higher female classmates' shares. We find no impact of a broader women representation in Economics on the probability of having a female advisor or willingness to pursue graduate studies in Economics.

1.6.2 Elective courses' choice

In Table 1.6.4, we assess if higher female representation in Economics impacts elective courses' choice. As mentioned above, the coursework structure at the University of Sao Paulo is relatively rigid, as around 57% of credits are in mandatory courses. The remaining credits are elective courses from the Department of Economics (at least 29% of the total credits) and courses from other departments or universities (at most 13% of the credits). Our analysis's dependent variables are the shares of elective courses from the Department of Economics that the student enrolled in each field: Microeconomics, Macroeconomics, Finance, and Humanities (History, Sociology, and other humanities courses), considering all elective courses that the student enrolled as the denominator.

As in the previous tables, in Panel A, we present the effects of female instructors' share. In Panel B, we show the impacts of the percentage of female peers, and in Panel C, we display estimates for the proportion of high-achieving female classmates. Overall, we note that women's representation in Economics does not influence female students' course choices. None of the coefficients is statistically significant at a 5% level, and only one is significant at a 10% level (Finance coefficient in Panel B). Moreover, the Westfall-Young tests do not reject the null hypothesis of no impacts of the gender composition of classmates and professors.

1.6.3 Work during undergraduate studies

One of our main findings is that a greater women's representation in Economics increases labor force participation. This subsection investigates if the gender

Table 1.6.4 – Long-term effects of women’s representation on elective courses’ choice

	Microeconomics	Macroeconomics	Finance	Humanities
Panel A: Predicted share of female professors in compulsory courses				
Female × Predicted % female professors	0.013 (0.073)	-0.012 (0.057)	-0.056 (0.068)	0.097 (0.100)
Predicted % female professors	-0.025 (0.051)	0.108 (0.089)	0.059 (0.042)	-0.036 (0.157)
Female	0.016 (0.016)	0.018 (0.011)	0.006 (0.012)	-0.026 (0.021)
Observations	1576	1576	1576	1576
Westfall-Young multiple testing (randomization-t p-value)	0.782			
Panel B: Share of female classmates				
Female × % female classmates	-0.009 (0.083)	0.045 (0.104)	-0.010 (0.073)	0.094 (0.141)
Share of female classmates	-0.001 (0.077)	0.047 (0.095)	0.106* (0.059)	0.094 (0.124)
Female	0.020 (0.023)	0.005 (0.028)	-0.002 (0.018)	-0.030 (0.038)
Observations	1576	1576	1576	1576
Westfall-Young multiple testing (randomization-t p-value)	0.927			
Panel C: Share of high-performing female classmates				
% high-performing female classmates × female	-0.006 (0.067)	-0.046 (0.066)	0.041 (0.040)	-0.128 (0.103)
% high-performing female classmates	0.026 (0.057)	0.070 (0.068)	0.045 (0.028)	-0.068 (0.074)
Female	0.020 (0.017)	0.024 (0.017)	-0.013 (0.010)	0.014 (0.025)
Observations	1576	1576	1576	1576
Westfall-Young multiple testing (randomization-t p-value)	0.614			
Mean of dependent variable	0.24	0.20	0.13	0.28
Std.dev. of dependent variable	0.13	0.12	0.10	0.21
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes
Students’ covariates	Yes	Yes	Yes	Yes
Classes’ covariates	Yes	Yes	Yes	Yes
Professors’ covariates	Yes	Yes	Yes	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me-out counts). High-performing: Top quartile (highest 25% admission scores). The sample is restricted to students admitted into the Economics major between 2002 and 2008. The dependent variables are the share of elective courses that a student enrolled from each field: Microeconomics, Macroeconomics, Finance, and Humanities (History, Sociology, and other humanities). The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2000 and 2008. Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

composition influences female students to make more career-oriented decisions already during undergraduate studies. More specifically, we will analyze if exposure to higher shares of female professors and peers impacts the likelihood of working while at university, either as an intern or as a regular employee.

As explained in Section 1.3, we restrict this analysis to cohorts admitted between 2002 and 2008.^{1.6.4} We report the estimates in Table 1.6.5. In column (1), we control for cohort-stream dummies, and we gradually include students' (column (2)), instructors', and classmates' covariates (column (3)) in the subsequent columns.

In Panel A, we represent the impacts of the share of female professors in core courses. Although female instructors increase the probability of work during graduation, there is no differential impact on female students. In Panel B, we verify that female peers' share increases female students' likelihood of working during undergraduate studies. A six percentage point increase in the share of female classmates (one standard deviation in the 2002-2008 sample) results in a five percentage point higher likelihood of work during the undergraduate studies, which corresponds to a 0.15 standard deviations rise. Finally, in Panel C, we examine if the percentage of high-achievement female classmates influences females' decision to work during graduation and find no impact.

1.7 Conclusion

This paper estimates female peers' and instructors' effects in an undergraduate program in Economics. As Economics is a male-dominated and math-intensive field, higher women's representation in the profession is relevant for female students and could influence their future career outcomes.

We explore the random assignment of students to classes at the Department of Economics from the University of Sao Paulo (USP), a prestigious public Brazilian university. We merge detailed USP administrative data with Brazilian formal labor market and firm ownership data to obtain labor market outcomes.

^{1.6.4}As a robustness check, we verified that our conclusions remain the same using all admission cohorts, i.e., 2000 to 2008.

Table 1.6.5 – Long-term effects of women’s representation on the probability of working during the undergraduate program

	(1)	(2)	(3)
Panel A: Predicted share of female professors in compulsory courses			
Female × Predicted % female professors	-0.355 (0.337)	-0.337 (0.337)	-0.351 (0.333)
Predicted % female professors	0.371** (0.142)	0.351** (0.138)	0.582*** (0.120)
Female	0.108 (0.068)	0.100 (0.064)	0.108 (0.063)
Observations	1227	1227	1227
Panel B: Share of female classmates			
Female × % female classmates	0.793** (0.314)	0.805** (0.309)	0.817** (0.312)
Share of female classmates	-0.403** (0.174)	-0.384** (0.170)	0.042 (0.118)
Female	-0.142* (0.080)	-0.149* (0.077)	-0.142* (0.076)
Observations	1227	1227	1227
Panel C: Share of high-performing female classmates			
% high-performing female classmates × female	0.069 (0.214)	0.059 (0.220)	0.056 (0.224)
% high-performing female classmates	-0.103 (0.088)	-0.101 (0.087)	0.009 (0.077)
Female	0.030 (0.048)	0.028 (0.048)	0.031 (0.050)
Observations	1227	1227	1227
Mean of dependent variable	0.87		
Cohort-stream fixed-effects	Yes	Yes	Yes
Students’ covariates	No	Yes	Yes
Classes’ covariates	No	No	Yes
Professors’ covariates	No	No	Yes

Notes: In Panels A, B, and C, we run OLS regressions. In Panel C, Share of high-performing female classmates = # female high-performing among peers ÷ # high performing among peers (leave-me-out counts). High-performing: Top quartile (highest 25% admission scores). The regressions include the sample of students admitted through USP’s admission exam into the Economics undergraduate degree between 2002 and 2008. The dependent variable is a dummy that equals one if the student worked as an intern or regular worker during the undergraduate studies, and zero otherwise. Students’ covariates: Normalized admission scores, Previous USP enrollment, Student’s age at admission, Sao Paulo city of residence dummy; Classes’ covariates: Class size, Share of female classmates, Average peers’ ability (admission scores), Share of peers with previous USP enrollment, Peers’ average age. Professors’ covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Female students become more career-driven when exposed to higher shares of female classmates during their undergraduate studies, increasing their labor force participation rates five years after (expected) graduation. We also find suggestive evidence that female faculty members' share affects female students' career progression, raising their likelihood of working as top managers.

We investigate potential channels driving our results. First, we assess the effects of female professors and peers on students' achievement. Gender differences in academic performance, advisor's gender, or willingness to pursue graduate studies cannot explain our results. Second, we examine if the gender composition of classmates and instructors influences female students' choice of elective courses, and we find no impact. Lastly, we show that classmates' gender composition affects the probability of work during undergraduate studies (at the formal labor market or as an intern). Our results show that a higher percentage of female peers increases women's likelihood of working during undergraduate studies. Since undergraduates at USP typically work on their degree areas, we interpret this finding as suggestive evidence of an early labor market attachment. More broadly, it is possible that other channels that we cannot measure, such as ambition, self-confidence, and persistence, mediate our effects.

Appendix – Figures and Tables

Figures

Figure 1.A.1 – Share of female students admitted into the Economics major

Figure 1.A.2 – Share of female professors in compulsory courses

Figure 1.A.3 – Distribution of admission scores, by gender

Notes: Notes: We normalize Phase 1, Phase 2, and admission scores, such that all variables have a mean of zero and standard deviation of one.

Tables

Table 1.A.1 – Comparing students from daytime and evening streams

	Daytime	Evening	Difference
Normalized admission scores	0.31 (0.98)	-0.32 (0.92)	0.62***
Age	18.58 (2.07)	20.36 (3.67)	-1.77***
Female	0.26 (0.44)	0.19 (0.39)	0.07***
Previous USP enrollment	0.07 (0.26)	0.20 (0.40)	-0.13***
First admission list	0.88 (0.32)	0.82 (0.39)	0.07***
Sao Paulo city	0.83 (0.38)	0.85 (0.36)	-0.02
Sao Paulo state	0.98 (0.14)	0.99 (0.09)	-0.01
N	801	775	1,576

Notes: The table presents descriptive statistics comparing students from daytime and evening streams. The column “Difference” reports the coefficient of a t-test of mean differences between female and male students. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.2 – Randomization balance check excluding 2008 admission year

	Female	Admission scores	Previous USP enrollment	SP city	Age	First admission list
Section 1	-0.034 (0.021)	0.001 (0.041)	0.009 (0.019)	0.011 (0.022)	0.055 (0.187)	0.016 (0.017)
Observations	1401	1401	1401	1401	1401	1401
F Statistics	2.47	0.00	0.23	0.23	0.09	0.89
Mean of dependent variable	0.23	-0.00	0.13	0.83	19.49	0.84
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table presents coefficients from separate OLS regressions. The dependent variables are Female indicator; Normalized admission scores (mean zero, standard deviation one); Previous enrollment at USP dummy; Sao Paulo city of residence dummy; Student’s age at admission; First admission list dummy. The key explanatory variable is a dummy variable for whether the student was assigned to section 1. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.3 – Testing whether students' characteristics correlate with their initials

	Female	Admission scores	Previous USP enrollment	SP city	Age	First admission list
initial==B	-0.075 (0.067)	-0.182 (0.140)	-0.073 (0.045)	0.024 (0.058)	-1.058*** (0.373)	-0.010 (0.052)
initial==C	0.081 (0.064)	-0.263* (0.136)	0.005 (0.048)	0.028 (0.051)	-0.286 (0.442)	-0.016 (0.046)
initial==D	-0.172*** (0.053)	-0.306** (0.134)	-0.019 (0.045)	0.034 (0.049)	-0.506 (0.391)	-0.086* (0.049)
initial==E	-0.191*** (0.057)	-0.241 (0.159)	0.038 (0.056)	0.062 (0.053)	0.056 (0.489)	-0.016 (0.051)
initial==F	-0.197*** (0.050)	-0.127 (0.131)	0.040 (0.047)	0.077* (0.045)	-0.165 (0.415)	-0.053 (0.045)
initial==G	-0.256*** (0.046)	-0.108 (0.123)	-0.012 (0.044)	0.014 (0.048)	-0.770* (0.392)	-0.018 (0.043)
initial==H	-0.218*** (0.069)	0.147 (0.240)	0.091 (0.086)	-0.049 (0.087)	0.605 (0.777)	-0.009 (0.070)
initial==I	-0.111 (0.102)	-0.151 (0.271)	-0.045 (0.077)	0.088 (0.078)	-0.685 (0.544)	-0.240** (0.115)
initial==J	-0.004 (0.065)	0.066 (0.136)	-0.079* (0.041)	0.052 (0.051)	-0.702* (0.403)	-0.013 (0.048)
initial==K	0.133 (0.156)	-0.224 (0.233)	0.123 (0.139)	0.011 (0.122)	1.703 (1.528)	0.038 (0.092)
initial==L	0.000 (0.060)	0.003 (0.122)	-0.016 (0.044)	0.041 (0.048)	-0.736* (0.391)	0.021 (0.041)
initial==M	0.012 (0.056)	-0.095 (0.121)	0.003 (0.043)	0.033 (0.046)	-0.121 (0.444)	0.004 (0.040)
initial==N	0.366*** (0.123)	-0.155 (0.198)	-0.088 (0.068)	0.005 (0.104)	-0.093 (0.839)	-0.059 (0.102)
initial==O	-0.196 (0.124)	-0.392 (0.329)	-0.150*** (0.030)	0.193*** (0.034)	-0.968 (0.605)	0.004 (0.121)
initial==P	-0.124** (0.061)	-0.017 (0.142)	-0.018 (0.049)	-0.018 (0.058)	-0.185 (0.580)	-0.016 (0.050)
initial==R	-0.209*** (0.046)	-0.186 (0.120)	0.002 (0.041)	0.047 (0.043)	-0.551 (0.381)	-0.006 (0.038)
initial==S	0.058 (0.099)	-0.376** (0.183)	-0.116** (0.046)	-0.083 (0.090)	-0.188 (0.746)	-0.044 (0.076)
initial==T	-0.140** (0.059)	-0.179 (0.134)	-0.059 (0.045)	0.011 (0.056)	-0.960*** (0.358)	-0.027 (0.050)
initial==V	-0.081 (0.073)	-0.000 (0.164)	-0.070 (0.049)	0.013 (0.064)	-0.903** (0.444)	-0.051 (0.062)
initial==W	-0.321*** (0.040)	-0.850*** (0.248)	0.004 (0.105)	-0.038 (0.122)	3.311 (2.180)	-0.179 (0.132)
initial==Y	-0.036 (0.176)	-0.108 (0.409)	-0.150*** (0.030)	-0.093 (0.175)	-1.129*** (0.411)	-0.014 (0.136)
Mean of dependent variable	0.23	-0.00	0.14	0.84	19.46	0.85
Stand. dev of dependent variable	0.42	1.00	0.34	0.37	3.09	0.36
Observations	1576	1576	1576	1576	1576	1576

Notes: The table presents coefficients from separate OLS regressions. The dependent variables are: Female indicator; Normalized admission scores (mean zero, standard deviation one); Previous enrollment at USP dummy; Sao Paulo city of residence dummy; Student's age at admission; First admission list dummy. The explanatory variables are dummies for students' first names initials. Robust standard errors in parentheses. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.4 – Comparing course-sections 1 and 2

	Daytime stream			Evening stream		
	Section 1	Section 2	Diff.	Section 1	Section 2	Diff.
Female professor	0.22 (0.42)	0.23 (0.42)	-0.01	0.17 (0.38)	0.15 (0.36)	0.02
# students enrolled in course section	56.18 (13.21)	55.33 (14.74)	0.85	53.42 (14.05)	49.85 (13.19)	3.58*
Average grade	5.90 (0.94)	5.96 (0.86)	-0.06	5.31 (1.12)	5.51 (1.10)	-0.20
Average attendance rate	86.60 (6.96)	86.83 (6.75)	-0.24	82.97 (7.92)	84.49 (7.98)	-1.52
Average pass rate	0.82 (0.13)	0.83 (0.11)	-0.02	0.74 (0.17)	0.76 (0.15)	-0.03
Share of female students	0.25 (0.06)	0.28 (0.05)	-0.03***	0.19 (0.06)	0.21 (0.06)	-0.02**
Average admission score	660.39 (29.15)	658.85 (27.97)	1.54	632.38 (28.83)	635.48 (26.33)	-3.10
Share of students from daytime stream	0.91 (0.07)	0.93 (0.06)	-0.02**	0.09 (0.13)	0.08 (0.12)	0.01
Professor: Ph.D. in Economics	0.72 (0.45)	0.73 (0.45)	-0.01	0.70 (0.46)	0.67 (0.47)	0.03
Professor: Ph.D. in other math-intensive field	0.20 (0.40)	0.19 (0.39)	0.01	0.23 (0.42)	0.21 (0.41)	0.01
Professor: Ph.D. at USP	0.62 (0.49)	0.58 (0.50)	0.04	0.50 (0.50)	0.56 (0.50)	-0.06
Professor: Ph.D. abroad	0.33 (0.47)	0.37 (0.48)	-0.04	0.44 (0.50)	0.40 (0.49)	0.04
Professor's experience	18.37 (10.00)	16.21 (9.17)	2.16	16.45 (9.07)	14.52 (8.83)	1.93
Professor: # published papers	15.44 (13.44)	12.17 (10.95)	3.27*	13.48 (14.11)	12.08 (10.36)	1.39
Professor: % papers with <i>Qualis</i> classification	0.79 (0.28)	0.79 (0.28)	-0.00	0.71 (0.33)	0.69 (0.31)	0.02
Professor: % A papers	0.11 (0.14)	0.11 (0.18)	-0.01	0.12 (0.20)	0.09 (0.17)	0.02
N	148	144	292	125	125	250

Notes: The table presents descriptive statistics comparing sections 1 and 2 from the compulsory courses that our sample of students attend over the undergraduate studies, by the stream (time of day). We exclude unified course sections to compute these descriptive statistics (courses with a single section, where students from both sections 1 and 2 enroll). The column “Diff” reports the coefficient of a t-test of mean differences between section 1 and section 2. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 1.A.5 – Subsample of students without individual taxpayer number (CPF)
at USP data: Comparing students with or without CPF at our final
data

	CPF	No CPF	Diff.
Normalized admission scores	-0.38 (0.96)	-0.36 (0.98)	-0.02
Age	19.36 (3.03)	19.38 (2.68)	-0.02
Daytime classes	0.52 (0.50)	0.59 (0.50)	-0.07
Previous USP enrollment	0.09 (0.28)	0.15 (0.37)	-0.07
Sao Paulo city	0.85 (0.36)	0.95 (0.22)	-0.10
Sao Paulo state	0.99 (0.11)	0.97 (0.16)	0.01
First admission list	0.80 (0.40)	0.82 (0.39)	-0.02
Female	0.25 (0.43)	0.21 (0.41)	0.05
Observations	517	39	556

Notes: For the subsample of students that USP did not provide us CPF information, we compare students' characteristics with or without the individual taxpayer information (CPF). We could partially obtain the CPFs by merging USP data with the labor market database using the full name and birth date. The column "Diff." reports the coefficient of a t-test of mean differences between groups. Variables: Normalized admission scores (mean zero, standard deviation one), Student's age at admission, Daytime stream dummy, Previous USP enrollment, Sao Paulo city of residence dummy, Sao Paulo state of residence dummy, First USP admission list, gender dummy. Standard deviations in parentheses. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.6 – Brazilian Classification of Occupations and Industries

Panel A - Brazilian Classification of Occupations (CBO 2002)		
Occupation/Position	CBO	
Economist	Economist - 251205, Agroindustrial Economist - 251210, Financial Economist - 251215, Public Economics - 251225 , Environmental Economist - 251230, Urban Economics - 251235, Industrial Economist - 251220, Researcher in Economics - 203510, Professor in Economics - 234805	
Top Manager	Top manager from the public service (CBO family 114); Top Manager - General (CBO family 121); Top Manager - Production and Operation (CBO family 122); Top Manager - Support services (CBO family 123); Top Manager Service (Health, Education, Cultural, Social) (CBO family 131).	
Middle Manager	Middle Manager - Production and Operation (CBO family 141); Middle Manager - Support services (CBO family 142).	
Panel B - Brazilian Classification of Industries (CNAE - Classificacao Nacional de Atividades Economicas)		
Industry	CNAE 2.0	CNAE 1.0
Financial activities	Division 64 - Financial services, Division 65 - Insurance, Retirement plans, Health insurance, 66 - Subsidiary financial services	Division 65 - Financial services, Division 66 - Insurance, Retirement plans, Health insurance, 67 - Subsidiary financial services
Public sector	Division 84 - Public Administration, National Defense and Social Security	Division 75 - Public Administration, National Defense and Social Security
Consulting firms	Division 70 - Consulting firms	Subclass 74160 - Consulting firms
Research	Division 72 - Research and Development (R & D); Group 853 - Higher Education Institutions	Division 73 - Research and Development (R & D); Group 803 - Higher Education Institutions

Table 1.A.7 – Summary statistics: Long-term students' outcomes, years after expected graduation

	Full sample	Female	Male	Diff.
Panel A: One year after expected graduation				
Formal labor market	0.52 (0.50) [1576]	0.50 (0.50) [357]	0.52 (0.50) [1219]	-0.03
Firm partner	0.07 (0.26) [1576]	0.07 (0.26) [357]	0.07 (0.26) [1219]	0.00
Labor force participation	0.56 (0.50) [1576]	0.53 (0.50) [357]	0.57 (0.50) [1219]	-0.04
Normalized hourly real wages	0.00 (1.00) [787]	-0.14 (0.70) [170]	0.04 (1.07) [617]	-0.18*
Panel B: Two years after expected graduation				
Formal labor market	0.64 (0.48) [1576]	0.65 (0.48) [357]	0.64 (0.48) [1219]	0.01
Firm partner	0.09 (0.28) [1576]	0.09 (0.29) [357]	0.09 (0.28) [1219]	0.00
Labor force participation	0.69 (0.46) [1576]	0.70 (0.46) [357]	0.68 (0.47) [1219]	0.01
Normalized hourly real wages	-0.00 (1.00) [992]	-0.11 (0.86) [225]	0.03 (1.04) [767]	-0.14
Panel C: Five years after expected graduation				
Formal labor market	0.70 (0.46) [1576]	0.72 (0.45) [357]	0.69 (0.46) [1219]	0.02
Firm partner	0.16 (0.37) [1576]	0.14 (0.35) [357]	0.17 (0.38) [1219]	-0.03
Labor force participation	0.79 (0.41) [1576]	0.79 (0.41) [357]	0.79 (0.41) [1219]	0.00
Normalized hourly real wages	0.00 (1.00) [1082]	-0.02 (1.10) [252]	0.01 (0.97) [830]	-0.02

Notes: The table presents descriptive statistics of students' outcome variables one, two, or five years after expected graduation. The column "Diff." reports the coefficient of a t-test of mean differences between groups. Variables: Formal labor market: dummy for whether the student was at the Brazilian formal labor market dataset (RAIS); Firm partner: indicator variable equals one if a person is a firm partner/owner; Labor force participation: work at the formal labor market or own a firm; Normalized hourly real wages: average hourly real wages, 2002 prices in Brazilian *reais* (conditional on being at the formal labor market), normalized such as to have a mean of zero and standard deviation of one. Standard deviations in parentheses and the number of observations in square brackets. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 1.A.8 – Share of female professors, 2SLS Regressions: First stage estimations

	% female professors $\times$ female			Female professors		
Female $\times$ Predicted % female professors	0.780*** (0.049)	0.782*** (0.049)	0.781*** (0.050)	-0.019 (0.041)	-0.023 (0.041)	-0.025 (0.041)
Predicted % female professors	0.006 (0.012)	0.005 (0.012)	0.014 (0.030)	0.837*** (0.087)	0.838*** (0.085)	0.790*** (0.095)
Female	0.040*** (0.009)	0.040*** (0.009)	0.040*** (0.009)	0.001 (0.007)	0.003 (0.007)	0.003 (0.008)
Observations	1576	1576	1576	1576	1576	1576
F Statistics	242.91	243.73	286.31	35.17	36.92	201.96
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' covariates	No	Yes	Yes	No	Yes	Yes
Classes' covariates	No	No	Yes	No	No	Yes
Professors' covariates	No	No	Yes	No	No	Yes

Notes: The table presents the first stage of our 2SLS regressions. We estimate OLS regressions. The dependent variables are female professors' share in compulsory courses and its interaction term with the gender dummy. Predicted % of female professors is the share of female professors assigned to a class (section, stream, and admission year group) in compulsory courses. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.9 – IV estimates: Long-term effects of the predicted share of female professors on labor force participation

	Formal	Formal 1 +	Formal 2 +	Formal 5 +	Work	Work 1 +	Work 2 +	Work 5 +
Female × % female professors	-0.192 (0.222)	0.541 (0.464)	0.614* (0.340)	0.195 (0.297)	0.026 (0.199)	0.447 (0.560)	0.579 (0.471)	0.552* (0.280)
% female professors	0.144 (0.207)	-0.169 (0.415)	0.243 (0.379)	0.257 (0.422)	0.183 (0.156)	-0.336 (0.318)	0.129 (0.275)	0.157 (0.357)
Female	0.051 (0.037)	-0.084 (0.069)	-0.086 (0.076)	-0.010 (0.056)	0.004 (0.031)	-0.078 (0.083)	-0.074 (0.087)	-0.086 (0.056)
Observations	1576	1576	1576	1576	1576	1576	1576	1576
Mean of dependent variable	0.91	0.52	0.64	0.70	0.94	0.56	0.69	0.79
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students' covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Classes' covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Professors' covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: We run 2SLS regressions. The dependent variables are Formal: formal labor market participation dummy, and Work: labor force participation, also considering firm ownership. 1+, 2+, 5+ state that the variable refers to one, two, or five years after expected graduation. The regressions include the sample of students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (Admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 1.A.10 – IV estimates: Long-term effects of the predicted share of female professors on occupational and industry choice

	Occupation	Industry			
	Economist	Finance	Government	Consulting	Research
Female $\times$ % female professors	-0.121 (0.728)	0.052 (0.387)	0.219 (0.325)	1.101** (0.491)	-0.294 (0.211)
% female professors	-0.334 (0.426)	0.193 (0.155)	-0.212 (0.232)	0.091 (0.325)	0.258 (0.311)
Female	0.094 (0.144)	-0.051 (0.086)	-0.082 (0.058)	-0.155 (0.094)	0.048 (0.042)
Observations	1441	1441	1441	1441	1441
Mean of dependent variable	0.28	0.53	0.17	0.20	0.09
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes
Students' covariates	Yes	Yes	Yes	Yes	Yes
Classes' covariates	Yes	Yes	Yes	Yes	Yes
Professors' covariates	Yes	Yes	Yes	Yes	Yes

Notes: We run 2SLS regressions. The dependent variables are occupation dummies for economists, sector dummies for financial services, public sector (government), consulting firms, and research institutions. The regressions include the sample of students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (Admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.11 – IV estimates: Long-term effects of the predicted share of female professors on career progression

	Top Manager	Middle Manager
Female $\times$ % female professors	0.505** (0.197)	0.374 (0.374)
% female professors	-0.322 (0.306)	-0.254 (0.313)
Female	-0.093** (0.040)	-0.051 (0.075)
Observations	1441	1441
Mean of dependent variable	0.07	0.30
Cohort-stream fixed-effects	Yes	Yes
Students' covariates	Yes	Yes
Classes' covariates	Yes	Yes
Professors' covariates	Yes	Yes

Notes: We run 2SLS regressions. The dependent variables are occupation dummies for a top manager or middle manager. The regressions include the sample of students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (Admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.12 – IV estimates: Long-term effects of the predicted share of female professors on normalized real hourly wages (R\$ 2002 Brazilian *reais*)

	Norm.Wages+1	Norm.Wages+2	Norm.Wages+5
Female $\times$ % female professors	1.093 (1.212)	2.583** (0.899)	0.872 (0.794)
% female professors	-0.372 (1.019)	-1.391 (1.000)	-0.501 (0.654)
Female	-0.174 (0.186)	-0.434** (0.150)	-0.111 (0.140)
Observations	787	992	1082
Mean of dependent variable	0.00	-0.00	0.00
Cohort-stream fixed-effects	Yes	Yes	Yes
Students' covariates	Yes	Yes	Yes
Classes' covariates	Yes	Yes	Yes
Professors' covariates	Yes	Yes	Yes

Notes: We run 2SLS regressions. The dependent variable is the normalized average hourly wages, considering all labor contracts. We normalized each dependent variable to have a mean of zero and standard deviation of one. Wages' information is conditional on working in the formal labor market. 1+, 2+, 5+ state that the variable refers to one, two, or five years after expected graduation. The regressions include the sample of students admitted through USP's admission exam into the Economics undergraduate degree between 2000 and 2008. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (Admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 1.A.13 – Variation in the share of female classmates

	Mean	Standard deviation	Observations
Share of female classmates			
Raw variable	0.225	0.069	1576
Residuals after removing cohort-stream fixed-effects	-0.000	0.045	1576
Residuals students' covariates + cohort-stream FE	-0.000	0.045	1576
Residuals students' cov + classes' cov. + cohort-stream FE	-0.000	0.038	1576
Residuals students' cov + classes' cov. + professors' cov. + cohort-stream FE	-0.000	0.033	1576

Notes: In this table, we present the variation in the share of female classmates left after removing fixed effects and controlling for other explanatory variables, following [Bifulco, Fletcher e Ross \(2011\)](#) and [Lepine e Estevan \(2020\)](#). We gradually remove cohort-stream fixed-effects, students', classes', and professors' covariates.

Table 1.A.14 – Heckman Selection Two-Step Procedure: Long-term effects of the predicted share of female professors on the probability of working in a consulting firm

	Positive values			Input Zeros			Heckman		
Female $\times$ Predicted % female professors	0.880** (0.369)	0.891** (0.375)	0.862** (0.383)	0.779** (0.340)	0.812** (0.354)	0.794** (0.361)	0.886** (0.399)	0.815* (0.398)	0.746* (0.408)
Predicted % female professors	0.057 (0.320)	0.056 (0.324)	0.075 (0.289)	0.045 (0.294)	0.040 (0.296)	0.089 (0.258)	0.059 (0.318)	0.028 (0.324)	0.149 (0.293)
Female	-0.118 (0.077)	-0.127 (0.079)	-0.116 (0.081)	-0.101 (0.072)	-0.113 (0.075)	-0.103 (0.077)	-0.119 (0.085)	-0.105 (0.084)	-0.079 (0.089)
Mean of dependent variable	0.20			0.19			0.20		
Observations	1441	1441	1441	1576	1576	1576	1441	1441	1441
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students' covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Professors' covariates	No	No	Yes	No	No	Yes	No	No	Yes
Classes' covariates	No	No	Yes	No	No	Yes	No	No	Yes

Notes: In 'Positive values,' we run OLS without considering selection into the formal labor market. In 'Input zeros,' we run OLS following [Martorell e McFarlin Jr. \(2011\)](#) and inputting zeros for students that we do not find at the formal labor market (i.e., we consider that the student is not working in a consulting firm). The remaining column presents estimates of a Heckman Two-Step procedure. In the first stage, we estimate the selection equations using a Probit model, and in the second stage, we use OLS regressions. Instrument: % working at the formal labor market at SP city, by age, gender, and expected graduation year. The dependent variable is a Consulting firm industry dummy variable. Students' covariates: Normalized admission scores, Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.15 – Heckman Selection Two-Step Procedure: Long-term effects of high-performing female classmates on the probability of working as an economist

	Positive values			Input Zeros			Heckman		
% high-performing female classmates $\times$ female	0.638** (0.252)	0.674** (0.251)	0.712*** (0.233)	0.664** (0.232)	0.687*** (0.231)	0.726*** (0.212)	0.386 (0.264)	0.761*** (0.256)	0.840*** (0.233)
% high-performing female classmates	-0.101 (0.163)	-0.103 (0.160)	-0.135 (0.154)	-0.097 (0.152)	-0.100 (0.151)	-0.132 (0.149)	-0.074 (0.160)	-0.112 (0.158)	-0.160 (0.152)
Female	-0.045 (0.057)	-0.067 (0.057)	-0.074 (0.054)	-0.049 (0.054)	-0.064 (0.054)	-0.071 (0.051)	-0.014 (0.056)	-0.078 (0.053)	-0.090* (0.050)
Mean of dependent variable	0.28			0.26			0.28		
Observations	1441	1441	1441	1576	1576	1576	1441	1441	1441
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students' covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Classes' covariates	No	No	Yes	No	No	Yes	No	No	Yes
Professors' covariates	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Share of high-performing female classmates = # female high-performing among peers $\div$ # high performing among peers (leave-me out counts). High-performing: Top quartile (highest 25% admission scores). In 'Positive values,' we run OLS without considering selection into the formal labor market. In 'Input zeros,' we run OLS following [Martorell e McFarlin Jr. \(2011\)](#) and inputting zeros for students that we do not find at the formal labor market (i.e., we consider that the student is not working as an economist). The remaining column present estimates of a Heckman Two-Step procedure. In the first stage, we estimate the selection equations using a Probit model, and in the second stage, we use OLS regressions. Instrument: % working at the formal labor market at SP city, by age, gender, and expected year of graduation. The dependent variable is an economist occupation dummy variable. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.A.16 – Heckman Selection Two-Step Procedure: Long-term effects of predicted female professors' share in compulsory courses on the probability of working as a top manager

	Positive values			Input Zeros			Heckman		
Female × Predicted % female professors	0.382** (0.152)	0.386** (0.153)	0.400** (0.154)	0.328** (0.127)	0.329** (0.130)	0.337** (0.129)	0.372** (0.160)	0.398** (0.159)	0.411** (0.159)
Predicted % female professors	-0.031 (0.163)	-0.039 (0.156)	-0.251 (0.228)	-0.033 (0.134)	-0.040 (0.130)	-0.214 (0.195)	-0.035 (0.162)	-0.035 (0.154)	-0.258 (0.229)
Female	-0.071** (0.032)	-0.070* (0.033)	-0.075** (0.033)	-0.063** (0.028)	-0.061** (0.029)	-0.065** (0.029)	-0.068* (0.036)	-0.073* (0.035)	-0.079** (0.034)
Mean of dependent variable	0.07			0.06			0.07		
Observations	1441	1441	1441	1576	1576	1576	1441	1441	1441
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students' covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Professors' covariates	No	No	Yes	No	No	Yes	No	No	Yes
Classes' covariates	No	No	Yes	No	No	Yes	No	No	Yes

Notes: In 'Positive values,' we run OLS without considering selection into the formal labor market. In 'Input zeros,' we run OLS following [Martorell e McFarlin Jr. \(2011\)](#) and inputting zeros for students that we do not find at the formal labor market (i.e., we consider that the student is not working as a top manager). The remaining column presents estimates of a Heckman Two-Step procedure. In the first stage, we estimate the selection equations using a Probit model, and in the second stage, we use OLS regressions. Instrument: % working at the formal labor market at SP city, by age, gender, and expected year of graduation. The dependent variable is a Top manager occupation dummy variable. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 1.A.17 – Heckman Selection Two-Step Procedure: Long-term effects of the predicted share of female professors on normalized real hourly wages two years after graduation

	Positive values			Input Zeros			Heckman		
Female × Predicted % female professors	2.214*** (0.705)	2.126** (0.771)	2.126** (0.767)	2.000*** (0.593)	1.767*** (0.554)	1.764*** (0.555)	2.163** (0.924)	2.109* (1.000)	2.042** (0.955)
Predicted % female professors	-0.927 (0.911)	-1.101 (0.761)	-1.071 (0.694)	-0.475 (0.664)	-0.459 (0.624)	-0.329 (0.386)	-0.928 (0.916)	-1.101 (0.769)	-1.103 (0.708)
Female	-0.470*** (0.143)	0.363** (0.146)	-0.359** (0.133)	-0.397*** (0.133)	0.301** (0.123)	-0.299** (0.121)	-0.464** (0.171)	-0.360* (0.174)	-0.348** (0.159)
Mean of dependent variable	-0.00			-0.00			-0.00		
Observations	992	992	992	1554	1554	1554	992	992	992
Cohort-stream fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Students' covariates	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Classes' covariates	No	No	Yes	No	No	Yes	No	No	Yes
Professors' covariates	No	No	Yes	No	No	Yes	No	No	Yes

Notes: In 'Positive values,' we run OLS without considering selection into the formal labor market. In 'Input zeros,' we run OLS following [Martorell e McFarlin Jr. \(2011\)](#) and inputting zeros for students that we do not find at the formal labor market (i.e., we consider that the student is receiving a zero wage). The remaining column presents estimates of a Heckman Two-Step procedure. In the first stage, we estimate the selection equations using a Probit model, and in the second stage, we use OLS regressions. Instrument: Dummy equals one if students had information on individual taxpayer number (CPF). The dependent variable is the normalized average hourly wages, considering all labor contracts. We normalized each dependent variable to have a mean of zero and standard deviation of one. Students' covariates: Normalized admission scores (mean zero, standard deviation one), Previous USP enrollment, Student's age at admission, Sao Paulo city of residence dummy; Classes' covariates: Class size, Share of female classmates, Average peers' ability (admission scores), Share of peers with previous USP enrollment, Peers' average age. Professors' covariates: Share of female professors, Percentage with Ph.D. abroad, Experience, Share of A papers, Share of classified papers. Standard errors in parentheses, clustered at the cohort-stream level. P-values: * p<0.10, ** p<0.05, *** p<0.01.

2 Gender differences in reaction to high-stakes

With Fernanda Estevan and Louis-Philippe Morin ^{2.0.1}

2.1 Introduction

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

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

^{2.0.1}Financial support from the Social Sciences and Humanities Research Council of Canada (430-2013-1033), Sao Paulo Research Foundation (FAPESP - grant # 2015/21640-3 and 2017/50134-4), CALDO, British Academy and the Newton Fund (Newton Advanced Fellowship, AF140079), CNPq, FGV EESP, and University of Ottawa are gratefully acknowledged. Bruna received financial support from Sao Paulo Research Foundation (FAPESP - grant # 2017/23593-8) and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Finance Code 001). We are very grateful to COMVEST, UNICAMP's admission office, for providing the data and assistance during the project. This research has benefited from discussions with Ana Maria Fonseca de Almeida, Pierre Brochu, Abel Brodeur, Lena Edlund, Bruno Ferman, Jason Garred, Daniel Haanwinckel, Adam Lavecchia, Daniele Paserman, Aureo de Paula, and Christopher R. Taber and comments from seminar participants at PUC-Rio, University of Ottawa, FGV EESP, SBE Economia Aplicada and the 48th ANPEC National Meeting. We also thank Derek Rice, Duangsuda Sopchokchai, Bogdan Urban, and Tiago Ferraz for their excellent research assistance. All remaining errors are ours.

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 stakes are higher. In our setting, we observe female and male applicants taking the same exams but with varying stakes within the exam. More specifically, UNICAMP applicants have to write the admission exam in two phases, nearly two months apart. Both stages are composed of open-ended questions on typical high school subjects (e.g., biology, history, mathematics, etc.). In Phase 1, the final score is the unweighted average of all subjects, and therefore there is no advantage in doing exceptionally well on a specific subject. In Phase 2, students must again answer questions on the same subjects, but applicants' 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. 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 some obstacles that are pervasive 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 were unweighted). Furthermore, we can rule out the possibility that females or males do not provide significant effort (or slack off) in Phases 1 and 2. Indeed, both phases are competitive and count toward the final score, one must pass Phase 1 to write Phase 2, and relative (rather than absolute) performance determines admission. These features allow us to avoid the potential confounding factors one faces when students have an objective like being admitted or attaining a specific final grade and adjust their effort based on past performance (on midterms, for example).

We show that increased stakes affect the gender performance gap in a significant way. Moving from a non-priority to a priority subject reduces the female performance advantage by the equivalent of 9% of the within-student standard deviation. This effect is even larger for higher-ability applicants. These results

are not driven by one subject in particular. Doing the analysis separately by subject yields the same conclusions, although the change in the gender gap is more substantial for some subjects (e.g., history and physics).

Our environment is particularly suitable for investigating potential channels explaining our findings. Not only do we observe applicants' performance in each subject, but we also have information on their performance on each of the 12 questions per subject. We obtain results similar to [Schlosser, Neeman e Attali \(2019\)](#), that women tend to put a higher effort than men in questions with lower stakes. Female applicants have a higher number of 'perfect' (maximum) scores in non-priority subjects questions, suggesting that perfectionism or diligence could be behind our findings. Also, we find that females become relatively more likely to omit (not answer) questions when stakes increase, suggesting that gender differences in confidence play a role. We also find that reaction to difficult questions, time-management issues, and mental fatigue ([YOON et al., 2009](#)) cannot explain our findings as females actually perform better on more difficult questions (especially on priority subjects) and later questions. Finally, we can rule out gender-specific selection into priority subjects, gender differences in knowledge about UNICAMP rules, or women 'choking' under pressure as potential mechanisms underlying our results.

Our study closely relates to the literature stream that discusses gender differences in performance in high- versus low-stakes settings. Results suggest that women perform relatively worse than men when there is an increase in stakes. [Ors, Palomino e Peyrache \(2013\)](#) show that men outperform women at the highly selective HEC Paris admission exam (high-stakes) compared with both the end-of-high-school exam and first-year core courses at HEC, which have lower stakes. Using data from a high-achieving private school in Barcelona, [Azmat, Calsamiglia e Iriberri \(2016\)](#) show that men outperform women when stakes are higher by comparing their performance in different school exams with varying stakes (from 5% to 27% of the final grade) and an end-of-high-school exam. Also, by comparing results from a mock and the actual Chinese Gaokao exam from a province, [Cai et al. \(2019\)](#)'s results point to the same direction and have the advantage of being based on the same exam format. However, they note that they cannot disentangle

the effect of the exam ‘stakes’ and the degree of competition as the mock and actual examinations differ across both dimensions.

One mechanism investigated by [Cai et al. \(2019\)](#) is whether men do not take the mock exam seriously since it does not affect the likelihood of university admission. They run a survey on Gaokao preparation (in particular, study time) and do not find evidence of gender differences in study effort (for the mock exam or actual Gaokao) or increased effort as the exam approaches. However, based on the high-stakes actual GRE and a low-stakes voluntary experimental GRE section, [Schlosser, Neeman e Attali \(2019\)](#) argue that the primary mechanism to explain differences in performance between mock and real exams is that men react more to incentives. Moreover, women tend to put in some effort even when stakes are ‘nearly zero.’

We organize the rest of the paper as follows. The next section details UNICAMP’s admission exam process, and we describe the administrative data used in our analysis in Section 2.3. Section 2.4 presents our empirical strategy. Section 2.5 shows our main results, including estimates allowing for heterogeneity across subjects and applicant ability, and robustness checks. Finally, we investigate mechanisms that could potentially explain our results in Section 2.6 and provide a conclusion in Section 2.7.

2.2 UNICAMP admission exam

Each year, students applying to UNICAMP must write an admission exam (*vestibular*).^{2.2.1} When registering for the admission exam (about two months before writing the exam), candidates apply to up to three majors (ranked first, second, and third options), so admission is program-specific. Given that UNICAMP uses an admission allocation based on the Boston mechanism and admission is competitive, the vast majority of applicants who are denied admission to their first option are

^{2.2.1}This section describes the rules governing *vestibular* between 2000 and 2004, our period of analysis. While rules changed over time, few changes occurred over our period of analysis. We highlight these changes along with the discussion.

not admitted to any of their three options (ESTEVAN; GALL; MORIN, 2019).^{2.2.2}

Vestibular is composed of two sequential phases (hereafter referred to as P_1 and P_2 , respectively).^{2.2.3} P_1 is written in November, and P_2 in January (the academic year begins in late February). In both stages, all applicants write exactly the same examinations, regardless of their career choices. The order that the questions appear in the exams are the same for all applicants, and they can, within the same exam day, freely choose the order to answer the questions. Yet, we empirically observe that the mean scores decrease with the question's order, suggesting that they tend to answer the questions following the exam's layout.

In P_1 , all applicants answer the same twelve written-answer questions on six high school subjects (i.e., two questions per subject) irrespective of the major they apply to: mathematics, physics, chemistry, biology, history, and geography. These subjects carry the same weight in the grade calculation (i.e., the P_1 score) – each question is worth 2.5 points, totalizing a maximum grade of 30 points. Importantly, applicants do not have any incentive to perform particularly well on a specific high school subject in P_1 as they all carry the same weight—the stakes are identical across subjects. Applicants must also write an essay in P_1 , which is worth 30 points.^{2.2.4} UNICAMP computes the applicants' P_1 score using two different formulas and selects the most favorable one for each applicant. The first formula corresponds to the sum of the essay and the general questions to determine the applicant's P_1 score. In the second formula, the sum of the essay and general questions receives a weight of 80%, and ENEM, an end-of-high-school exam score calibrated to a maximum score of 60, gets a weight of 20%.^{2.2.5} Thus, under both

^{2.2.2} Around 10% of the UNICAMP applicants are admitted into the university. Among the admitted applicants, almost 90% are admitted in their first major choice. Most students (80%) indicate a second option for their career choices. For applicants that file a second option, 95% have the same priority subjects in the first and second choices. 58% of the candidates indicate a third career choice, and for them, 86% have the same priority discipline in the first and third options.

^{2.2.3} Additionally, some majors, like Dentistry and Performing Arts, require an aptitude test.

^{2.2.4} In 2004, the maximum Phase 1 score is 120 points. Each question is worth 5 points, totalizing 60 points per subject. The essay is worth 60 points. Since the relative weights remained unchanged in 2004, this difference does not affect our analysis.

^{2.2.5} The applicant must have taken ENEM in the two previous years and provided UNICAMP with her ENEM id. The second formula (that considers ENEM scores) is used for 90% of the applicants.

formulas, the maximum P_1 score is 60 points.

Applicants' P_1 score must be above a major-specific cutoff to qualify for P_2 . The baseline cutoff P_1 score is 50% (or 30 points). However, the major-specific cutoff score is adjusted upward or downward to guarantee that the number of applicants per major in P_2 is between three and eight per available slot.^{2.2.6}

If an applicant passes P_1 , she has to write P_2 . The P_2 exam covers the same subjects as P_1 plus Portuguese and a foreign language (English or French). There are twelve questions per subject in each of the eight P_2 subjects, each worth five points. Thus, the total number of points for each P_2 subject exam is 60 points. As in P_1 , P_2 exams are identical for all applicants and composed of written-answer questions (no multiple-choice questions). In contrast to P_1 , the subjects are weighted differently in P_2 . 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.

Appendix Table 2.A.1 presents the list of majors we consider in our sample along with their priority subjects, admission cutoff (a major competition measure), the proportion of females among their applicants, and the proportion of students applying to this major (all measured over the 2000-2004 period).^{2.2.7} Most majors have two priority subjects, and they vary significantly across majors. By far, the most popular (with 20 percent of applicants) and competitive major is medicine at UNICAMP. While there appear to be some clusters of priorities (e.g., biology and chemistry for life-science programs, mathematics and physics for engineering programs), there is still significant variation in priority subjects. Generally, these types of programs are more popular with a particular gender (e.g., life-science programs are usually more prevalent among females), but importantly there remains significant variation in priority subjects and the proportion of females associated with priority subjects. For example, chemical engineering, for which 46 percent of applicants are females, has chemistry and math as priority subjects. Econo-

^{2.2.6}In 2000, the number of applicants in P_2 could be up to 25 per slot in high-demand majors.

^{2.2.7}Programs offered during daytime and the ones offered in the evening (labeled with "(Eve.)" in Appendix Table 2.A.1) are treated as separate majors as they attract different types of potential students and are usually different in terms of admission competition (cutoffs).

mics has history and mathematics as priorities, and 41 percent of its applicants are females.^{2.2.8} Finally, three-quarters of food engineering applicants, who have mathematics and physics as priorities, are females.

The final admission score combines P_1 standardized score with a weight of two, P_2 priority disciplines standardized scores with a weight of two, and P_2 non-priority disciplines standardized scores with a weight of one.^{2.2.9} Thus, for a typical major with two priority subjects and no aptitude test, priority subjects count for one-third of applicants' final admission scores.

For admission, UNICAMP ranks applicants based on the (standardized) final scores and minimum grade requirements. The allocation mechanism is a version of the Boston mechanism, which initially considers applicants who chose the major as first choice before those who opted for it as second or third options.^{2.2.10} The major's cutoff is the lowest-ranked P_2 score that was admitted in the major.

2.3 Data

We work with data provided by *Comissão Permanente para os Vestibulares* (COMVEST), UNICAMP's admission office. The dataset is available for 2000-2008 and contains individual-level information on all applicants who wrote UNICAMP's admission exam, *vestibular*. In this paper, we limit our analysis to 2000-2004, since UNICAMP implemented its affirmative action policy in 2005 (ESTEVAN; GALL; MORIN, 2019). Focusing on the pre-affirmative action period allows us to leave any changes in the pool of applicants caused by the policy out of our analysis.^{2.3.1} We also concentrate on the six subjects, biology, chemistry, history, geography, mathematics, and physics, that are covered in both P_1 and P_2 (there

^{2.2.8}39% of the applicants are female in the Economics program offered in the evening.

^{2.2.9}Until 2003, the standardization of P_2 exams was done separately for applicants of majors within four defined areas. From 2004 on, the standardization considers the grades of P_2 exams of all students who participated in the exam. The standardization of the P_1 scores only considers the grades of candidates who passed P_1 . Aptitude tests carry a weight of two, except for Dentistry, which is one.

^{2.2.10}See Estevan, Gall e Morin (2019) for more details on the admission procedure.

^{2.3.1}The *vestibular* is also open to students who wish to practice before finishing high school. Since these *trainees* are not eligible for admission, we exclude them from our primary analysis.

are no questions specific to Portuguese or a foreign language in P_1). We do so to be able to control for applicants' subject-specific ability in our empirical approach.

Importantly, for P_1 survivors, we see the applicants' grades in each question of P_1 and P_2 in our dataset. More specifically, for each applicant, we observe twelve P_1 grades (two of each of the six subjects) and 96 P_2 scores (twelve of each of the eight subjects) covered in P_2 , and the essay score (Phase 1). For 2000-2002, we can also distinguish between a missing answer and an answer that received a zero grade. Since we have the applicants' major choices, we can easily detect the priority subjects (i.e., Phase 2 subjects that carry a weight of two instead of one in the final score).

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

Our initial sample contains the 58,547 applicants who attended both phases of the exam 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 percent), with age younger than 16 or older than 27 (2.5 percent), and with a missing ENEM score (5 percent). Our final sample contains 53,949 applicants for the 2000 to 2004 UNICAMP admission exams.

It is worth mentioning that we are looking at a selected pool of applicants, as we focus on applicants that attended both stages (Phase 1 and 2). UNICAMP is a highly competitive admission exam, and only 30% of the applicants pass Phase 1. Thus, we analyze gender gaps in performance for applicants that obtain better outcomes in a competitive environment.

Table 2.3.1 describes applicant-level information for our sample of interest. Male and female applicants differ in many meaningful dimensions, which will justify our empirical strategy in the next section. First, female applicants have, on average, significantly lower ENEM scores and apply to less competitive majors (based on the P_2 score cutoffs to be admitted). Both differences are substantial. Females' average ENEM score is 0.24 s.d. below the overall average, while males' is 0.19 s.d.

Table 2.3.1 – Descriptive statistics

	Full sample	Female	Male	Difference
Female	0.43 (0.50)			
Age	19.14 (1.60)	19.14 (1.53)	19.14 (1.64)	-0.00
Norm. ENEM scores	-0.00 (1.00)	-0.24 (1.05)	0.19 (0.92)	-0.43***
# priority subjects	1.79 (0.40)	1.72 (0.45)	1.85 (0.36)	-0.13***
Major cutoff (by year)	522.78 (78.91)	508.40 (84.24)	533.73 (72.72)	-25.33***
Biology is a priority subject	0.31 (0.46)	0.43 (0.50)	0.21 (0.41)	0.22***
Chemistry is a priority subject	0.26 (0.44)	0.32 (0.47)	0.21 (0.41)	0.11***
Geography is a priority subject	0.02 (0.12)	0.01 (0.12)	0.02 (0.13)	-0.00***
History is a priority subject	0.16 (0.37)	0.20 (0.40)	0.13 (0.34)	0.07***
Mathematics is a priority subject	0.56 (0.50)	0.38 (0.48)	0.69 (0.46)	-0.32***
Physics is a priority subject	0.40 (0.49)	0.23 (0.42)	0.53 (0.50)	-0.30***
Portuguese is a priority subject	0.09 (0.29)	0.15 (0.35)	0.05 (0.23)	0.09***
Normalized Phase 1 scores (average)	-0.00 (0.63)	-0.09 (0.66)	0.07 (0.59)	-0.16***
Normalized Phase 2 scores (average)	-0.00 (0.78)	-0.12 (0.80)	0.09 (0.75)	-0.21***
Normalized Phase 2 scores (weighted average)	0.07 (0.79)	-0.05 (0.81)	0.16 (0.77)	-0.21***
Phase 2 score standard deviation	7.24 (2.51)	7.11 (2.48)	7.33 (2.54)	-0.22***
Phase 2 score coefficient of variation	0.32 (0.17)	0.34 (0.19)	0.31 (0.16)	0.03***
Within-student stand.dev - Norm P2 Scores	0.62	0.60	0.64	
Within-student stand.dev - Norm P1 Scores	0.78	0.77	0.78	
# Applicants	53,949	23,303	30,646	53,949

Notes: ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. Phase 1 and 2 “scores (average)” are computed at the applicant level using equal weights. In contrast, “Phase 2 scores (weighted average)” are computed at the applicant level using a weight of 2 for priority subjects and 1 for non-priority subjects. * significant at 10%; ** significant at 5%; *** significant at 1%

above. The difference in the average major cutoff (about 25 points) represents 0.32 s.d. of the major-cutoff distribution. The gender difference in P_1 performance is not as large as the one observed in ENEM performance, suggesting that females prepare better for the admission exam.

Given females and males apply to different majors (see Table 2.A.1), it is not surprising that their priority subjects differ significantly. 43% of females have biology as a priority subject, a proportion twice as large as for males. Biology is a priority subject for biological sciences, dentistry, medicine, nursing, pharmacy, physical education, and phonology. A number of these majors have only one priority subject (e.g., biological sciences, dentistry, and nursing), which explains the difference in the number of priority subjects, lower for females. In contrast, males are significantly more likely to have mathematics (69% vs. 38%) or physics (53% vs. 23%) as priority subjects. Mathematics is a priority subject for mathematics, statistics, engineering programs, economics, information technology, construction technology, physics, computer science, media studies, environmental sanitation technology, and telecommunication technology.

Table 2.3.1 also shows gender differences in performance on Phase 2 of the exam, for the weighted (using different weights for priority subjects relative to non-priority subjects) and unweighted averages. Males do better than females, but this is not surprising given the differences in ENEM and P_1 performances. The full-sample weighted average above 0 indicates that applicants perform better in their priority subjects and, therefore, select into topics they are relatively better at. Finally, we present the within-applicant standard deviation for both P_1 and P_2 performance. Since our variation of interest occurs at the applicant level (priority vs. non-priority subjects), the within-applicant variation is more informative than the overall variation (which combines both the within- and across-applicant variation in performance) to gauge the magnitude of our effect. As expected, the within-applicant standard deviation (0.62) is significantly lower than the overall (normalized at 1.00).

In most cases, gender differences in major selection, and therefore priority-subject choice, would jeopardize the identification of our parameter of interest (i.e., the effect of increased stakes), especially if this selection is based on expected

performance in the priority subjects. The richness of our data—being able to observe applicants’ previous subject-specific performance and to include candidate fixed-effects (or an overall academic ability measure) given that we observe six outcomes per candidate—along with our empirical strategy will allow us to control for major self-selection based on subject-specific ability (or gains in performance), in a flexible way.

2.4 Empirical model

We now present an analytical framework to motivate our regression model and highlight its identification challenges. Imagine a student i writing an exam consisting of questions on a set of different subjects, s . Some subjects, called priority subjects, are weighted more heavily than others to determine the exam’s final score. Applicant i ’s performance on a specific subject s , y_i^s , can be expressed as:

$$y_i^s = \rho^s + \pi_i + \gamma_i^s + P_i^s \phi_i + P_i^s \omega_i^s + \varepsilon_i^s, \quad (2.4.1)$$

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. 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 student’s subject-specific additional performance change (over and above ϕ_i) when s is a priority. In this setup, we allow student ability and their reaction to a priority subject to differ across subjects. Finally, ε_i^s is a purely random performance shock. Note that none of the terms on the equation’s right-hand side (2.4.1) 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 (2.4.1) in terms of deviations from the group means:

$$y_i^s = \rho^s + \pi^g + \tilde{\pi}_i^g + \gamma^{s,g} + \tilde{\gamma}_i^{s,g} + P_i^s[\phi^g + \tilde{\phi}_i^g] + P_i^s[\omega^{s,g} + \tilde{\omega}_i^{s,g}] + \varepsilon_i^s, \quad (2.4.2)$$

where the parameters without the i subscripts represent group averages (e.g., $\pi^g \equiv E(\pi_i|g)$ and $\omega^{s,g} \equiv E(\omega_i^{s,g}|s,g)$), and g stands for gender. Parameters with tildes are the applicant's deviations from these group averages (e.g., $\tilde{\pi}_i^g \equiv \pi_i - \pi^g$). By construction, the expectations of tilde parameters are all equal to zero.^{2.4.1}

Equation (2.4.2) suggests the following regression equation:

$$\begin{aligned} y_i^s &= \pi^m + F_i \Delta \pi + P_i^s (\phi^m + \omega^{s,m}) + P_i^s F_i (\Delta \phi + \Delta \omega^s) + \rho^s + \tilde{\pi}_i^g + \\ &\quad \gamma^{s,g} + \tilde{\gamma}_i^{s,g} + P_i^s [\tilde{\phi}_i^g + \tilde{\omega}_i^{s,g}] + \varepsilon_i^s, \\ &= \beta_1 + F_i \beta_2 + P_i^s \beta_{3,s} + P_i^s F_i \beta_{4,s} + u_i^s, \end{aligned} \quad (2.4.3)$$

where 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,^{2.4.2} and:

$$u_i^s \equiv \rho^s + \tilde{\pi}_i^g + \gamma^{s,g} + \tilde{\gamma}_i^{s,g} + P_i^s [\tilde{\phi}_i^g + \tilde{\omega}_i^{s,g}] + \varepsilon_i^s. \quad (2.4.4)$$

Note that if we were to assume that applicants' 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 (2.4.3) become subject invariant (e.g., $\beta_{4,s} = \beta_4$, $\forall s$).^{2.4.3} In this case, our parameter of interest would be β_4 , the effect of increasing stakes 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' reaction to vary across subjects, then we will 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 (2.4.4) 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 $\tilde{\pi}_i^g$) are usually controlled for in the previous literature [see, for e.g., [Azmat, Calsamiglia e Iriberry \(2016\)](#), [Cai et al. \(2019\)](#)] using 'test-type'

^{2.4.1}Note that because all students write each and every subject, $E(\pi_i|s, g) = E(\pi_i|g)$, $\forall s$.

^{2.4.2}For example, π^m is the average general academic ability of male applicants and $\Delta \pi$ is the gender gap in general academic ability ($\pi^f - \pi^m$).

^{2.4.3}Furthermore, if we assume that applicants' 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 $\beta_{4,s}$ in equation (2.4.3) becomes subject invariant (e.g., $\beta_{4,s} = \beta_4$, $\forall s$), but $\beta_{3,s}$ would still vary across subjects.

and individual fixed-effects. In our case, one should control for each subject's difficulty, especially if harder subjects are more likely to be a priority for a specific gender. We do this by including subject fixed-effects in our regressions. Second, if the general-academic distributions differ across gender and females and males apply to different programs, then we would expect some correlation between $P_i^s F_i$ and applicant relative ability ($\tilde{\pi}_i^g$). Hence, equation (2.4.4) suggests controlling for student ability when estimating equation (2.4.3)—this is done using individual fixed-effects in [Azmat, Calsamiglia e Iriberry \(2016\)](#) and [Cai et al. \(2019\)](#).

So far, our identification challenges and their solutions seem similar to the ones in [Azmat, Calsamiglia e Iriberry \(2016\)](#) and [Cai et al. \(2019\)](#). An additional concern for us is that applicants are tested on different subjects, some of which may favour female or male applicants (captured by $\gamma^{s,g}$ in equation (2.4.4)).^{2.4.4} 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 stakes. This potential issue motivates the use of gender-specific subject fixed-effects in our regression.

Finally, the presence of $\tilde{\gamma}_i^{s,g} + P_i^s[\tilde{\phi}_i^g + \tilde{\omega}_i^{s,g}]$ highlights a common issue when trying to estimate a model with random coefficients. One would wish that these terms are uncorrelated with our regressor of interest ($P_i^s F_i$), but given that applicants choose their major, and therefore their priority subjects, such assumption is unlikely to hold. It is quite plausible that applicants choose their major based on their relative overall academic ability ($\tilde{\phi}_i^g$) or their comparative advantages (e.g., $\tilde{\gamma}_i^{s,g}$ or $\tilde{\omega}_i^{s,g}$) 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 e Vytlačil \(1998\)](#), [Wooldridge \(2005\)](#)]. Usually, estimating a treatment effect in the presence of correlated coefficients is challenging and requires the use of control functions, instrumental variables or selection models ([DAHL, 2002](#); [HECKMAN](#); [URZUA; VYTLACIL, 2006](#))

^{2.4.4}See, for example, [Ellison e Swanson \(2010\)](#), [Niederle e Vesterlund \(2010\)](#), and [Ellison e Swanson \(2018\)](#) for discussions on gender differences in math ability.

Fortunately, the richness of our data allows us to control for such selection. Each applicant has to answer questions on the same subjects during the first phase of the admission exam as they do in the second phase. Importantly, in Phase 1, each subject is equally weighted to determine who will move on to the second phase and overall performance in Phase 1 has a small impact on the likelihood to be admitted (it weights 2, equivalent to one priority subject), conditional on making it to the second stage, as no weight is attached to P_1 performance in the calculation of the final P_2 exam grade.^{2.4.5} There is no incentive to do well in particular subjects. Hence, we can use applicants' performance on each Phase 1 subject to control for their relative subject-specific ability ($\tilde{\gamma}_i^{s,g}$), the same performance interacted with the priority dummy variable to control for their potential relative subject-specific performance gain when a subject is a priority ($\tilde{\omega}_i^{s,g}$). Additionally, we can use applicants' relative ENEM scores (a measure of overall applicant ability)^{2.4.6} interacted with the priority dummy variable to control for $\tilde{\phi}_i^g$.

Equations (2.4.3) and (2.4.4), along with the available applicant information, motivate our main regression equation:

$$y_i^s = \beta_{3,s}P_i^s + \beta_{4,s}P_i^sF_i + G(\widetilde{P1}_i^{s,g}) + H(P_i^s \times \widetilde{ENEM}_i^g) + J(P_i^s \times \widetilde{P1}_i^{s,g}) + \rho^s + \pi_i + \gamma^{s,g} + v_i^s, \quad (2.4.5)$$

where $\widetilde{P1}_i^{s,g}$ is the applicant's relative performance in subject s on Phase 1 and

^{2.4.5}The fact that P_1 performance has only a weight of 2 in the calculation of the final admission grade (and that they do not know their relative performance—only the P_1 raw score) is another important advantage over some of the previous studies on the subject [e.g., [Azmat, Calsamiglia e Iriberry \(2016\)](#)]. 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 to it by providing more or less effort on P_2 . For example, some applicants could slack off on P_2 (or a final exam), knowing that their P_1 (or midterm) relative performance allows them to do so. If the gender differential in reaction to P_1 achievement varied across priority and non-priority subjects, such differential would lead to a biased estimation of β_4 .

^{2.4.6}To test if ENEM scores capture overall students' ability, we have checked if the normalized ENEM scores correlate with the standardized Phase 1 final score (general questions scores plus essay score). We observe a high correlation (55%) that holds both for female (60%) and male applicants (51%). When we regress the Phase 1 score on the ENEM score allowing for different correlations by UNICAMP admission year, we find a highly statistically significant positive association (t-statistic 229). Although this relation varies by year, it is always positive and significant at a 1% level.

$\widetilde{ENEM}_i^g$ is the applicant's relative performance on the ENEM exam (performance minus group g average, as defined above). $G(\cdot)$, $H(\cdot)$ and $J(\cdot)$ are flexible functions meant to capture $\tilde{\gamma}_i^{s,g}$, $P_i^s \tilde{\phi}_i^g$ and $P_i^s \tilde{\omega}_i^{s,g}$, respectively.^{2.4.7} Subject and individual fixed-effects will capture ρ^s and π_i , while gender-subjects interactions will capture $\gamma^{s,g}$. Note that the constant term (β_1), the gender gap in overall performance (β_2) and the student relative ability ($\tilde{\pi}_i^g$) will be absorbed by the individual fixed-effects (π_i). Our performance measure y_i^s is the applicant's Phase 2 score in subject s , normalized by subject and admission year.^{2.4.8}

There are two main necessary assumptions for $P_i^s \times \widetilde{ENEM}_i^g$ and $P_i^s \times \widetilde{P1}_i^{s,g}$ to be valid proxy variables for $P_i^s \tilde{\phi}_i^g$ and $P_i^s \tilde{\omega}_i^{s,g}$. First, our proxy variables should be redundant in equation (2.4.3). That is, the proxy variables should be irrelevant in (2.4.3) if we could control directly for general and subject-specific (interacted with the priority dummy), and the other included regressors. In our case, this assumption seems reasonable as what should determine an applicant's performance is her ability, not its measure. Second, our interacted abilities, $P_i^s \tilde{\phi}_i^g$ and $P_i^s \tilde{\omega}_i^{s,g}$, should be uncorrelated with each regressor once we partial out their proxy variables. In particular for us, if we could regress our interacted abilities on their proxies, the error terms from these regressions should not be correlated with $P_i^s F_i$. Essentially, this second assumption rules out the possibility that our proxy variables mismeasure ability systematically differently across gender. We now investigate the potential validity of this last assumption in detail by investigating the role of gender differences in taste for majors.

2.4.1 Validity

Our results could be driven by gender-specific selection into priority subjects. Suppose men consider their comparative advantage when selecting their major, and women do not consider or consider to a lesser extent. In that case, the gender gap

^{2.4.7}In our main specification, these functions have a quartic form. For example, $G(\widetilde{P1}_i^{s,g}) \equiv \sum_{j=1}^4 \alpha_j (\widetilde{P1}_i^{s,g})^j$.

^{2.4.8}As a robustness test, we use raw Phase 2 scores as our dependent variable, controlling for year fixed-effects. Results are qualitatively the same. We report the estimates in Table 2.A.2.

could derive from gender differences in selection, not gender differences in reaction to stakes. As we have discussed above, our empirical strategy deals with gender differences in comparative advantage, and we do not expect that this endogeneity issue affects our estimates. Moreover, the results presented in Table 2.4.1 support our identification strategy's validity.

We investigate whether we observe gender differences in 'future' increased stakes in Phase 1 of the admission exam (remember that in P_1 all subjects are equally weighted). If gender differences in comparative advantage explained our findings, we would find lower female Phase 1 performance in subjects that will be a priority in the second stage. Besides, if females were to concentrate in P_1 on disciplines that will be priorities in P_2 while males do not, then one could argue that females cannot improve much more (at least not as much as males) in priority subjects between P_1 and P_2 . Such a situation would have consequences for interpreting our results in our preferred specifications in Table 2.5.1, where we control for P_1 subject-specific performance. Female specialization in P_1 and P_2 combined with a lack of reaction in P_1 and specialization in P_2 from males could lead to a negative estimate for our parameter of interest (i.e., a negative bias), despite females being the ones reacting more to the increased stakes.

Table 2.4.1 presents the results from estimating regressions where applicants' P_1 subject-specific scores are regressed on a dummy variable, Priority, equal to 1 if the subject will be a priority in P_2 , zero otherwise and Female $\times$ Priority interaction term (and similar regressors as in our main specification). Two findings come out of Table 2.4.1. First, applicants do better in subjects that will be a priority in P_2 , suggesting that students apply to majors in which they have a comparative advantage. This finding suggests that we should control for P_1 scores in P_2 to control for such selection. Second, females do not seem to concentrate more or underperform in subjects that will be priorities in P_2 . In fact, the estimate is close to zero. For example, the Female $\times$ Priority parameter estimate is a non-statistically significant -0.01 in our preferred specification, and it is fairly small compared to the within-applicant standard deviation for P_1 scores (0.78). These findings suggest that applicants' potential reaction to future priority subjects in P_1 is not a confounding factor in our results.

Table 2.4.1 – Priority and P_1 subject-specific performance

	(1)	(2)	(3)	(4)
<i>Dependent variable: Phase 1 normalized subject-specific scores</i>				
Female	0.025*** (0.005)	0.038*** (0.008)		
Priority	0.344*** (0.005)	0.331*** (0.005)	0.302*** (0.005)	0.306*** (0.005)
Female $\times$ Priority	-0.014** (0.007)	0.008 (0.007)	-0.008 (0.007)	-0.010 (0.007)
Norm. ENEM scores	0.531*** (0.003)	0.531*** (0.003)		
Number of observations	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes
Individual fixed-effects	No	No	Yes	Yes
Relative perf. ENEM $\times$ Priority	No	No	No	Yes

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

Furthermore, we analyzed if P_1 performance in a subject correlates with the likelihood of having this subject as a priority in Phase 2 (see Table 2.A.7). Our response variable is a priority subject binary variable that equals one if a subject was a priority for the applicant in Phase 2 and zeroes otherwise. We regress the priority subject dummy on normalized Phase 1 score, a female dummy, and the interaction of the gender dummy and P_1 scores. In column (2), we include subject fixed-effects, in column (3), we add subject-gender fixed-effects, and in column (4), we control for the normalized ENEM scores. Column (5) includes individual fixed-effects, and column (6) controls simultaneously for the subject, subject-gender, and individual fixed-effects. Overall, we note that a higher score in Phase 1 is positively related to the priority subjects' choice. After the inclusion of the individual fixed-effects (column (5)), we observe that a one-standard-deviation higher P_1 score is associated with an increase in the probability that the applicant selects the

priority subject by nine percentage points for men and eight percentage points for women. Once we control for subject fixed-effects, the results are the opposite: an increase by eight percentage points in the likelihood for male applicants and nine percentage points for females. If anything, females seem slightly more inclined to base their comparative advantage when choosing their major, not less. Overall, we observe that while a comparative advantage, measured by higher subject-specific scores, correlates with the priority subjects' choice, the gender differences in the correlations are small and do not seem to explain our main findings.

2.5 Main results

Table 2.5.1 presents the results from estimating different variations of equation (2.4.5), imposing a common β_4 across subjects. In all specifications, we cluster our standard errors at the student level. In column (1), we look at the overall gender difference in performance, controlling for student ability. The results suggest that female applicants, once we control for ability using ENEM, perform significantly better than males (by 0.066 s.d.) in P_2 subjects that are non-priority. This advantage disappears completely for priority subjects—the gender performance gap in these subjects is 0.000 (0.066 – 0.066).^{2.5.1} The estimate on the ‘Female $\times$ Priority’ is sizeable as it represents about 11 percent of the within-applicant standard deviation (0.62) — the within-applicant standard deviation captures the variation in performance across all subjects within applicants.

Our measure of ability, ENEM, looks like a good predictor of applicants' performance. A one-standard-deviation increase in ENEM is associated with a 0.54 s.d. increase in performance (with a t-statistic over 200). Looking at the subject fixed-effects, students tend to perform on average best in geography and worst in mathematics.

The specification in column (1) of Table 2.5.1 does not attempt to control for students selection into majors, other than through a linear control for overall

^{2.5.1}Testing that the sum of the Female and Female $\times$ Priority parameters is equal to 0 results in a p-value of 0.96.

Table 2.5.1 – Priority subjects and gender performance gap

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Phase 2 normalized subject-specific scores</i>						
Female	0.066*** (0.005)	0.163*** (0.007)	0.133*** (0.006)			
Priority	0.580*** (0.004)	0.560*** (0.004)	0.436*** (0.004)	0.461*** (0.004)	0.471*** (0.004)	0.490*** (0.005)
Female × Priority	-0.066*** (0.006)	-0.049*** (0.006)	-0.049*** (0.006)	-0.061*** (0.006)	-0.055*** (0.006)	-0.055*** (0.006)
Norm. ENEM scores	0.543*** (0.003)	0.543*** (0.003)	0.407*** (0.002)			
Chemistry	0.024*** (0.003)	0.101*** (0.004)	0.103*** (0.004)	0.103*** (0.004)	0.103*** (0.004)	0.101*** (0.004)
Geography	0.156*** (0.004)	0.184*** (0.006)	0.157*** (0.006)	0.164*** (0.005)	0.165*** (0.005)	0.162*** (0.005)
History	0.076*** (0.004)	0.090*** (0.005)	0.077*** (0.005)	0.081*** (0.005)	0.082*** (0.005)	0.079*** (0.005)
Mathematics	-0.147*** (0.004)	-0.098*** (0.005)	-0.046*** (0.005)	-0.056*** (0.005)	-0.054*** (0.005)	-0.059*** (0.005)
Physics	-0.062*** (0.003)	0.040*** (0.005)	0.071*** (0.005)	0.066*** (0.005)	0.066*** (0.005)	0.062*** (0.005)
Female × Chemistry		-0.178*** (0.007)	-0.193*** (0.007)	-0.193*** (0.006)	-0.194*** (0.006)	-0.193*** (0.006)
Female × Geography		-0.072*** (0.009)	-0.100*** (0.009)	-0.099*** (0.008)	-0.099*** (0.008)	-0.099*** (0.008)
Female × History		-0.035*** (0.008)	-0.055*** (0.008)	-0.054*** (0.008)	-0.053*** (0.008)	-0.053*** (0.008)
Female × Mathematics		-0.101*** (0.008)	-0.165*** (0.007)	-0.153*** (0.007)	-0.153*** (0.007)	-0.150*** (0.007)
Female × Physics		-0.227*** (0.007)	-0.289*** (0.007)	-0.279*** (0.006)	-0.277*** (0.006)	-0.276*** (0.006)
Number of observations	323694	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949	53949
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	No	Yes

Notes: Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for P_1 relative performance and for 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%.

ability. The goal of column (1) is simply to present a baseline performance gender gap between females and males that can serve as a comparison when interpreting the results from other specifications. As Table 2.3.1 shows visible gender differences in major selection, it is unlikely that the β_4 estimate captures causality unless our measure of ability completely captures this selection. The following specifications sequentially attempt to capture more complex gender/individual differences in major selection.

Column (2) includes female-subject interaction terms, which allow females to perform on average better (or worse) in different subjects. In this case, the parameter for ‘Female’ represents the performance gender gap in biology. Allowing for such gender differences reduces the β_4 estimate, but it remains highly significant and represents 8 percent of the within-applicant standard deviation. The female-subject interaction terms suggest that females perform on average better than males (controlling for ENEM) in biology, geography, history, and mathematics, but worse in physics and slightly worse in chemistry. In column (3), we introduce a flexible (quartic) function of the applicant’s subject-specific performance in P_1 . These additional covariates allow us to go a step further in controlling for potential selection based on comparative advantage. Under this specification, the applicant’s overall ability will be captured by ENEM, while her (additional) subject-specific ability will be captured by the quartic function in P_1 scores. While the main parameter estimate does not change significantly, the parameter estimate for ENEM drops significantly, which is not surprising: subject-specific ability matters.

Columns (4) to (6) introduce a more flexible way to capture overall ability by dropping ENEM and replacing it with individual fixed-effects.^{2.5.2} The main impact of using fixed-effects instead of ENEM (column (4) versus (3)) is to increase the magnitude of our parameter estimate of interest to 10 percent of a within-applicant standard deviation. Finally, in columns (5) and (6), we add more covariates to control for major self-selection by interacting both the applicants’ relative overall and subject-specific performances with the priority-subject indicator variable. The introduction of these covariates is meant to capture major selection based on individual absolute and comparative advantages ($P_i^s[\tilde{\phi}_i^g + \tilde{\omega}_i^{s,g}]$ in equation (2.4.4)).

^{2.5.2}Doing so, the fixed-effects absorb the Female dummy.

Despite controlling for overall and subject-specific ability in fairly flexible ways, we still find that applicants' performance increases significantly when facing priority subjects, but females do not react as much as males when facing these increased stakes.

2.5.1 Heterogeneity

The imposition of a common effect of increased stakes across subjects in Table 2.5.1 may seem quite 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. We take two different approaches to investigate the potential heterogeneity in the priority-subject effects across subjects. First, we estimate equation (2.4.5) allowing β_3 and β_4 to vary across subjects. The advantage of using this approach is that we can continue using our applicant fixed-effects. Second, we estimate equation (2.4.5) separately by subject. When estimating the regressions separately by subject, we can no longer use applicant fixed-effects since we have one observation per applicant. However, we can relax the link between ENEM (and P_1) and the subject-specific performance, as we run one regression per subject and do not need to impose common coefficients.

Table 2.5.2 presents the results when we estimate equation (2.4.5) without imposing common β_3 and β_4 . We focus our discussion on specification (6), in which we have our full set of controls. We do not find any gender differences in reaction when stakes are larger in biology or chemistry, but observe large reactions (around -0.11, or 18 percent of a within-applicant standard deviation) when stakes increase in any other subjects. Looking back at our descriptive statistics (Table 2.3.1), we see that the two subjects for which we do not find gender differences in reaction to stakes (biology and chemistry) are also the two subjects where the gender difference in priority-subject proportions are the largest in favor of females.^{2.5.3} However, the

^{2.5.3}Biology and chemistry are the priority subjects of the medicine majors. Medicine is the most popular and competitive major at UNICAMP. We test if our results are robust to the exclusion of medicine majors (Medicine UNICAMP and Famerp). The gender gap in priority subjects is larger when we restrict the sample: female applicants' performance is 0.092 s.d lower (15% of the within-applicant standard deviation). Looking at the heterogeneous impacts by subject, we still

Table 2.5.2 – Phase 2 normalized subject-specific scores: Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Phase 2 normalized subject-specific scores</i>						
Female	0.119*** (0.008)	0.102*** (0.007)			0.096*** (0.007)	
Priority	0.753*** (0.010)	0.546*** (0.009)	0.312*** (0.008)	0.306*** (0.009)	0.532*** (0.010)	0.329*** (0.009)
Female × Priority	-0.052*** (0.014)	-0.036*** (0.013)	0.009 (0.011)	0.014 (0.011)	-0.047*** (0.013)	0.009 (0.011)
Female × Priority × Chemistry	0.073*** (0.018)	0.038** (0.017)	0.019 (0.015)	0.013 (0.015)	0.032* (0.017)	0.015 (0.015)
Female × Priority × Geography	-0.033 (0.061)	-0.022 (0.059)	-0.096* (0.056)	-0.117** (0.056)	-0.019 (0.059)	-0.112** (0.056)
Female × Priority × History	-0.095*** (0.025)	-0.084*** (0.023)	-0.119*** (0.021)	-0.111*** (0.021)	-0.078*** (0.023)	-0.108*** (0.021)
Female × Priority × Mathematics	0.007 (0.022)	-0.012 (0.020)	-0.103*** (0.019)	-0.100*** (0.019)	-0.005 (0.020)	-0.093*** (0.019)
Female × Priority × Physics	-0.110*** (0.022)	-0.106*** (0.019)	-0.137*** (0.018)	-0.134*** (0.018)	-0.093*** (0.019)	-0.126*** (0.018)
Priority × Chemistry	-0.154*** (0.013)	-0.110*** (0.012)	-0.154*** (0.011)	-0.155*** (0.011)	-0.114*** (0.012)	-0.157*** (0.011)
Priority × Geography	-0.106*** (0.040)	0.005 (0.039)	0.515*** (0.037)	0.561*** (0.037)	0.032 (0.038)	0.553*** (0.037)
Priority × History	-0.098*** (0.018)	0.008 (0.016)	0.522*** (0.015)	0.537*** (0.015)	0.026 (0.016)	0.531*** (0.015)
Priority × Mathematics	-0.371*** (0.016)	-0.222*** (0.015)	0.207*** (0.014)	0.219*** (0.014)	-0.221*** (0.015)	0.208*** (0.014)
Priority × Physics	-0.244*** (0.015)	-0.147*** (0.013)	0.201*** (0.013)	0.211*** (0.013)	-0.142*** (0.013)	0.200*** (0.013)
Norm. ENEM scores	0.538*** (0.003)	0.405*** (0.002)			0.391*** (0.002)	
Number of observations	323694	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	Yes	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	Yes	Yes	No	Yes
Relative perf. ENEM × Priority	No	No	No	Yes	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	Yes	Yes

Notes: Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for P_1 relative performance and for 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%.

link between gender differences in subject-specific performance and the gender gap in reaction to increased stakes does not follow a systematic pattern. When looking back at column (2) of Table 2.5.1, indeed, females perform better in biology (conditional on ENEM), but this is not the case for chemistry (where we do not find statistical differences). Relative to males, females perform better in geography, history, and mathematics. Conversely, the largest gender difference in reaction to higher stakes is in physics, where we also observe the largest gender gap in performance and second to the largest difference in priority-subject proportion.

Table 2.5.3 presents the results when we estimate the effect of increased stakes separately for each subject.^{2.5.4} The estimates tell a slightly different story than the ones observed in Table 2.5.2. All estimated impacts of increasing stakes are negative, in line with our results from Table 2.5.1, and they are all statistically significant, except for geography.^{2.5.5} The non-statistically significant result for geography is not surprising as only 2 percent of our 53,949 applicants have it as a priority subject, reflecting in the standard errors. As in the previous table, the point estimates for Female $\times$ Priority do not reveal a systematic pattern — larger estimates are not associated with subjects less popular among females or in which they tend to perform worse relative to males. Overall, the findings from Tables 2.5.2 and 2.5.3 suggest that the main results are not driven by specific subjects. Also, our results are not necessarily driven by subjects for which boys are known to exhibit overconfidence (BORDALO et al., 2019).

Next, we consider whether the effect of increasing stakes on the performance gender gap varies with student ability. There are a few ways to approach this

observe no gender differences in reaction to stakes in biology and a marginally significant relative better female performance in chemistry when the discipline is a priority (by 0.04 s.d.).

^{2.5.4}Note that when we estimate our regression by subject, we can no longer use individuals fixed-effects and therefore rely on ENEM and P_1 subject-specific performance to control for applicant subject-specific ability. We present the results where control for the interaction terms ‘Relative perf. ENEM $\times$ Priority’ and ‘Relative P_1 scores $\times$ Priority’ to be as close as possible to column (6) in Table 2.5.1, but without the individual fixed-effects.

^{2.5.5}The difference seems to be mainly explained by the use of participant fixed-effects in the specification (6) of Table 2.5.2. Imposing common quartic functions in ENEM, P_1 , and their interaction with Priority across subjects (in Table 2.5.2) does not seem to play a major role in explaining the difference in estimates. We can see the role of imposing common quartic functions by comparing the results from Table 2.5.3 and the ones in the specification (5) in Table 2.5.2. The estimates of interest in Table 2.5.3 and specification (5) in Table 2.5.2 are fairly similar.

Table 2.5.3 – Phase 2 normalized subject-specific scores: Regression results by subject

	Biology	Chemistry	Geography	History	Mathematics	Physics	Portuguese
Female	0.079*** (0.008)	-0.104*** (0.007)	0.035*** (0.008)	0.075*** (0.008)	-0.008 (0.010)	-0.197*** (0.007)	0.549*** (0.008)
Priority	0.513*** (0.013)	0.359*** (0.014)	0.477*** (0.052)	0.573*** (0.017)	0.333*** (0.012)	0.392*** (0.010)	0.241*** (0.024)
Female $\times$ Priority	-0.035*** (0.013)	-0.025* (0.013)	-0.023 (0.057)	-0.132*** (0.018)	-0.051*** (0.014)	-0.133*** (0.012)	-0.101*** (0.026)
Norm. ENEM scores	0.356*** (0.004)	0.341*** (0.004)	0.413*** (0.004)	0.382*** (0.004)	0.531*** (0.005)	0.324*** (0.004)	0.437*** (0.004)
Number of applicants	53949	53949	53949	53949	53949	53949	53949
Relative performance P1 scores	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	No	No	No	No
Relative perf. ENEM $\times$ Priority	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Relative perf. P1 scores $\times$ Priority	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Heteroskedasticity-robust standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

question, given our setup. We present the results where we interact Female $\times$ Priority with the ENEM relative performance measure and use a linear function of ENEM relative performance $\times$ Priority instead of a quartic function. We make this last change to see more easily if, on average, the effect of increased stakes on the gender performance gap increases or decreases with ENEM.^{2.5.6} Table 2.5.4 suggests that the impact of higher stakes on the gender performance gap increases in magnitude with applicants' ENEM performance. At the mean ENEM performance, the effect of increased stakes on the gender gap is very close to what we found in Table 2.5.1 (-.58 versus -.55). The difference in performance between priority and non-priority subjects increases for males with higher ENEM scores. All else equal, a male student with an ENEM score one standard deviation above the mean would have improved his performance on a priority subject by 0.055 s.d. more than a male student with an ENEM score at the mean. For females, this difference

^{2.5.6}Using a quartic function instead of a linear function has little impact on our 'Female $\times$ Priority' and 'Female $\times$ Priority $\times$ Rel. Performance ENEM' parameter estimates. These results are available upon request.

would be 0.019 (0.055-0.036).^{2.5.7} So, as we compare females and males with higher ENEM scores, the performance gender gap becomes more and more in favor of male students. For applicants one standard deviation above the mean, increased stakes will change the gender performance gap by 0.094 (15% of the within-applicant standard deviation). Doing a similar exercise, but looking at the link between P_1 subject-specific performance (instead of ENEM) and the effect of increased stakes on the gender gap yields similar results. Given that admission to UNICAMP is very competitive, our findings suggest that the gender difference in reaction to increased stakes could affect the gender representation of admitted students.^{2.5.8}

2.5.2 Robustness checks

As a first robustness check, we ran the same regressions, including applicants' performance in Portuguese. In this specification, we use the essay score as a Portuguese score in Phase 1. The results presented in Appendix Table 2.A.3 are almost identical once we include our full set of controls.

Our analytical framework suggests that we use equation (2.4.5) as our main econometric model and control for relative performance in Phase 1 scores in our regressions. Nevertheless, as a robustness test, we used the difference between Phase 2 and Phase 1 normalized scores as the dependent variable. This specification would deal with potential measurement issues related to the use of Phase 1 scores as proxies of subject-specific ability. We present the results using this alternative response variable in Table 2.5.5. We observe that female applicants have a large improvement in Phase 2 non-priority subjects than males (4% of a standard deviation in column (1)). The increase in scores from Phase 1 to Phase is larger in the priority subjects (around 0.2 s.d). As in our main regressions, we find that women react less to an increase in stakes: their change in performance is 4% of a standard deviation lower.

^{2.5.7}This difference is statistically significant at a 1% level.

^{2.5.8}We have run the regressions separately for female and male applicants to analyze the pattern of high-achieving students in priority subjects by gender. Men and women perform better in priority subjects, and the reaction to stakes is higher for the 'high-ability' students in both genders. However, the coefficients for males are larger both for the average and high-performing applicants. We can interpret this result as further evidence that women react less to an increase in stakes.

Table 2.5.4 – Interaction term: Relative performance ENEM

	(1)	(2)	(3)	(4)	(5)
<i>Dependent variable: Phase 2 normalized subject-specific scores</i>					
Female	-0.167*** (0.005)	-0.070*** (0.007)	-0.041*** (0.006)		
Priority	0.579*** (0.004)	0.558*** (0.004)	0.432*** (0.004)	0.458*** (0.004)	0.480*** (0.005)
Female × Priority	-0.067*** (0.006)	-0.050*** (0.006)	-0.047*** (0.006)	-0.058*** (0.006)	-0.058*** (0.006)
Female × Priority × Relative Performance ENEM	-0.017*** (0.006)	-0.016*** (0.006)	-0.028*** (0.005)	-0.039*** (0.005)	-0.036*** (0.005)
Female × Relative Performance ENEM	0.023*** (0.006)	0.022*** (0.006)	0.010** (0.005)		
Priority × Relative Performance ENEM	0.046*** (0.004)	0.046*** (0.004)	0.052*** (0.004)	0.045*** (0.004)	0.055*** (0.004)
Relative Performance ENEM	0.522*** (0.004)	0.523*** (0.004)	0.392*** (0.004)		
Number of observations	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	Yes

Notes: Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for relative P_1 performance and for 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%.

So far, we analyzed only priority subjects associated with the applicant's first major choice. One could argue that if women are more risk-averse than men, they could consider all their (up to three) major choices in their strategies during the admission exam instead of focusing on their first major option. This mechanism is unlikely to explain our results, as most admitted applicants obtain their first option (90%). Moreover, for most contestants, the priority subjects are the same in the second (95%) or third (86%) choices. Therefore, we do not expect our findings to vary substantially when we use priority subjects' from the following career choices. Still, in Table 2.A.4, we checked if our results remain similar if we take into account the priority subjects associated with the applicant's second (panel A) and third

Table 2.5.5 – Alternative dependent variable: Normalized Phase 2 - Phase 1 scores

	(1)	(2)	(3)	(4)
<i>Dependent variable: Normalized Phase 2 scores - Phase 1 scores</i>				
Female	0.040*** (0.004)	0.117*** (0.008)		
Priority	0.212*** (0.005)	0.208*** (0.005)	0.204*** (0.006)	0.204*** (0.006)
Female × Priority	-0.039*** (0.007)	-0.047*** (0.008)	-0.046*** (0.008)	-0.041*** (0.008)
Norm. ENEM scores	0.156*** (0.002)	0.157*** (0.002)		
Number of observations	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes
Individual fixed-effects	No	No	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	Yes

Notes: Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control 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%.

(panel B) career choices. Not all applicants indicate more than one career option. Thus, we restrict our sample in this exercise to the candidates that have registered other major choices: 44,246 applicants signal a second option and 22,858, a third option. When we use the following options' priority disciplines, we note that the gender gaps in priority subjects are more sensitive to covariates' inclusion. However, our preferred specification, which includes the complete set of control variables, presents a similar but slightly smaller gender difference in reaction to the increased stakes: women have a 0.05 s.d. lower performance in the priority subjects.

2.6 Potential channels

This section inquires into potential channels that could explain the results presented in Table 2.5.1. We start by analyzing gender differences in exam strategy

and understanding in which types of questions women underperform relative to men. Then we discuss potential mechanisms that may be related to these findings.

2.6.1 Gender differences in exam strategy

While the average performance can say something about the effect of increasing stakes, our data allow us to analyze applicants' performance in more detail, which could inform us on gender differences in strategies adopted during the exam.

Compared to males, female applicants show less variation in their performance across subjects. In Table 2.6.1, we explore the correlation between the applicant's gender and the Phase 2 score coefficient of variation (across subjects). Conditional on ENEM performance, women seem to equalize their effort among different subjects, incurring a lower Phase 2 score coefficient of variation. The correlation is stronger (more negative) for high-achieving female applicants.

Table 2.6.1 – Phase 2 score coefficient of variation

	(1)	(2)	(3)
<i>Dependent variable: Phase 2 score coefficient of variation</i>			
Female	0.029** (0.007)	-0.008* (0.003)	-0.008* (0.003)
Norm. ENEM scores		-0.085*** (0.008)	-0.078*** (0.008)
Female × Norm. ENEM scores			-0.014*** (0.001)
Mean of dependent variable	0.33		
Stand.dev of dependent var	0.16		
Number of applicants	53949	53949	53949
Year dummies	Yes	Yes	Yes

Notes: To calculate the coefficient of variation across subject scores, we include Foreign Language and Portuguese scores.

Our results in Table 2.6.2 suggest that females put more effort into non-priority subjects while males focus on correctly answering the higher stakes questions.

Indeed, women have a higher number of ‘perfect scores’ in non-priority disciplines and fewer maximum scores in priority subjects than men.^{2.6.1}

Table 2.6.2 – Priority subjects and number of ‘perfect scores’

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: # perfect scores in Phase 2 questions per subject</i>						
Female	0.039*** (0.006)	0.260*** (0.010)	0.229*** (0.010)			
Priority	0.723*** (0.009)	0.610*** (0.010)	0.488*** (0.009)	0.489*** (0.009)	0.448*** (0.010)	0.431*** (0.012)
Female × Priority	-0.280*** (0.013)	-0.156*** (0.013)	-0.159*** (0.013)	-0.162*** (0.013)	-0.158*** (0.013)	-0.158*** (0.013)
Norm. ENEM scores	0.553*** (0.004)	0.554*** (0.004)	0.420*** (0.003)			
Mean Y	1.33					
Stand.dev Y	1.71					
Number of observations	323694	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the number of ‘perfect scores’ obtained by an applicant in a given P_2 subject. ‘Perfect score’ is a question for which the applicant obtained the maximum score. The ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants’ P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants’ subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for relative P_1 performance and for 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%.

Surprisingly, the effect of increasing stakes is smaller for women when they face difficult questions. We identify difficult questions following the strategy used by Iriberry e Rey-Biel (2019). We begin by computing the average performance on each question (or each subject). Then, we classify a question in a specific subject as *difficult* if the average performance on this question is below the median of question

^{2.6.1}Our definition of ‘perfect scores’ is the number of questions for which the applicant obtained the maximum score (each question is worth 5 points). As there are 12 questions per subject in Phase 2, the response variable will vary between 0 and 12 ‘perfect scores.’

average scores for that subject.^{2.6.2} We interact this dummy variable with Female $\times$ Priority to investigate whether the questions' difficulty could exacerbate females' reaction to priority subjects. Table 2.6.3 suggests the opposite. This is true for all four outcomes considered (question score, perfect score, omitted question, and zero score).

We use the question's difficulty combined with its order in the exam to evaluate if female and male applicants react differently to 'negative shocks.'^{2.6.3} We define, by subject, three different events as 'negative shocks.' The first event equals one if the first question of a Phase 2 exam^{2.6.4} is difficult. Second, we specify a 'negative shock' if the exam's most challenging question^{2.6.5} was among the first four questions. Lastly, the other event considered a 'negative shock' is to have a very difficult question^{2.6.6} among the exam's early questions (1 to 4). We present the results from this exercise in Table 2.A.6. While males have worse performance when exposed to adverse shocks, the opposite occurs for women.

More generally, women seem to perform equally well at the beginning and the end of the exam. Each subject exam consists of 12 questions. We separate each subject exam into three parts: *early* questions (question 1 to 4), *mid-exam* questions (question 5 to 8), and *late* questions (question 9 to 12). Although we cannot identify the exact order in which applicants answer the questions, the proportions of zeros and omitted questions are significantly larger for questions 9 to 12 than for questions 1 to 4, suggesting that applicants answer questions in the order presented to them. Table 2.6.4 suggests that while males' performance worsens as we move to questions toward the end of the exam (their performance on early questions is better than on late questions) in priority subjects, females'

^{2.6.2}As robustness checks, we have also looked at different definitions for the level of difficulty. In particular, we classify a question as *very difficult* if the average score on this question is below the bottom quartile. Finally, we have also defined a question as *most difficult* if it is the question with the lowest average performance (within the subject). Our results remain qualitatively similar and are available upon request.

^{2.6.3}Cai et al. (2019) find for Chinese students that females are more responsive to adverse performance shocks than males.

^{2.6.4}We define an exam as the set of questions from the same phase, subject, and admission year.

^{2.6.5}The most difficult question is the item, per subject and year, that presented the lowest average score.

^{2.6.6}'Very difficult' questions are the items from an exam with the lowest 25% average scores.

Table 2.6.3 – Priority subjects, question's difficulty and performance

	Score	Perfect Score	Omission	Zero score
Priority	0.604*** (0.006)	0.114*** (0.001)	-0.042*** (0.001)	-0.047*** (0.001)
Female $\times$ Priority	-0.185*** (0.007)	-0.053*** (0.002)	0.016*** (0.001)	0.004*** (0.001)
Difficult question	-1.194*** (0.002)	-0.105*** (0.001)	0.045*** (0.001)	0.142*** (0.001)
Female $\times$ Difficult question	-0.054*** (0.004)	-0.007*** (0.001)	0.000 (0.001)	0.024*** (0.001)
Priority $\times$ Difficult question	-0.384*** (0.005)	-0.157*** (0.001)	0.007*** (0.001)	0.004*** (0.001)
Female $\times$ Priority $\times$ Difficult question	0.237*** (0.007)	0.080*** (0.002)	-0.019*** (0.001)	-0.022*** (0.002)
Mean Y	2.14	0.11	0.06	0.17
Std.dev Y	1.65	0.31	0.24	0.37
# observations	3884328	3884328	2275344	2275344
# applicants	53949	53949	31602	31602
Subject FE	Yes	Yes	Yes	Yes
Subject-gender FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Relative performance P1 scores	Yes	Yes	Yes	Yes
Relative perf. ENEM $\times$ Priority	Yes	Yes	Yes	Yes
Relative perf. P1 scores $\times$ Priority	Yes	Yes	Yes	Yes

Notes: Score is the question raw score. Perfect Score, Omission, and Zero Score are dummy variables. For the Omission and Zero score variables, we restrict the sample to the years 2000-2002. Priority is a dummy variable equal to 1 if the subject is a priority in P_2 . Biology is the baseline subject. Phase 1 subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use a quartic function to control for relative ENEM performance interacted with 'Priority.' Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

relative performance remains unchanged.^{2.6.7}

Table 2.6.4 – Priority subjects and within-exam performance (Early vs. Late questions)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Average score first 4 questions - Average score last 4 questions</i>						
Female	-0.030*** (0.005)	0.007 (0.010)	0.004 (0.010)			
Priority	0.004 (0.006)	-0.014** (0.006)	-0.020*** (0.006)	-0.023*** (0.006)	0.011 (0.007)	0.017** (0.008)
Female × Priority	-0.027*** (0.008)	-0.016* (0.009)	-0.012 (0.009)	-0.028*** (0.009)	-0.019** (0.009)	-0.018* (0.009)
Norm. ENEM scores	0.049*** (0.002)	0.049*** (0.002)	0.042*** (0.002)			
Mean Y	0.62					
Stand.dev Y	1.27					
Number of observations	323694	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	No	Yes	Yes
Relative perf. P1 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. Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for relative P_1 performance and for 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%.

Even if women focus more on non-priority subjects and difficult questions, we find no evidence that skipping questions impair women's performance in priority subjects. For 2000-2002, we can separate omitted questions from attempted questions with a score of zero.^{2.6.8} Table 2.6.5 suggests that females leave significantly more questions blank when stakes increase (especially when we compare the

^{2.6.7}The sum of the 'Priority' and 'Female × Priority' parameter estimates is very close to zero and statistically insignificant.

^{2.6.8}Since there are 12 questions per subject, the number of omissions and zeros varies between 0 and 12.

estimate to the average number of omitted questions).^{2.6.9} The results are reversed, in Table 2.6.6, as increased stakes lower the number of zeros of females relative to males (basically, zeros conditional on trying to answer). Treating omitted questions as zeros results in a negligible effect of increased stakes on the gender gap in zeros (results upon request). These results suggest that women and men use different strategies when faced with questions they do not know the answer to. If females had tried (e.g., guessed an answer) some of the questions they omitted and some of their answers translated in non-zero scores (without affecting their performance on other questions), they could have improved their performance relative to males. However, if all their ‘guessed’ answers resulted in zero scores or affected their performance on other questions (e.g., by reducing the amount of time for the other questions), these differences in strategies would not explain the gender gap in priority subjects. Given that the exam is based on short-answer questions instead of multiple-choice questions, it is not clear that inducing females to guess more on questions they struggle with would improve their performance.

2.6.2 Discussion

Women overperform men in difficult questions in priority subjects, ruling out gender differences in applicants’ ability levels to explain our main results. Indeed, the gender gap in priority subjects’ performance arises in easy questions, i.e., those that most applicants answer correctly.

Instead, our evidence suggests that women’s perfectionism or diligence may be potential explanations for our findings. Indeed, a striking pattern is that, relative to men, women obtain more perfect scores in non-priority subjects, which receive less weight in the final score. Also, the coefficient of variation suggests that female applicants, particularly the high-achieving ones, equalize more their effort in different subjects in general.

We find no evidence that gender differences in mental fatigue, time-management issues, or negative shocks are behind our results. While applicants seem to answer

^{2.6.9}Early or late questions do not drive these results, as the estimates are similar in relative terms (compared to the average). These results are available upon request.

Table 2.6.5 – Priority subjects and number of omitted questions (2000-2002)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: # omitted questions in Phase 2 per subject</i>						
Female	-0.168*** (0.013)	-0.129*** (0.014)	-0.096*** (0.014)			
Priority	-0.508*** (0.011)	-0.533*** (0.013)	-0.428*** (0.012)	-0.397*** (0.011)	-0.447*** (0.012)	-0.470*** (0.013)
Female × Priority	0.080*** (0.014)	0.112*** (0.018)	0.102*** (0.017)	0.101*** (0.014)	0.081*** (0.015)	0.083*** (0.015)
Norm. ENEM scores	-0.309*** (0.007)	-0.309*** (0.007)	-0.177*** (0.007)			
Mean Y	0.74					
Stand.dev Y	1.52					
Number of observations	189612	189612	189612	189612	189612	189612
Number of applicants	31602	31602	31602	31602	31602	31602
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the number of questions omitted by an applicant in a given P_2 subject. The sample is restricted to 2000-2002. Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for relative P_1 performance and for 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%.

questions in the order they are presented, gender differences in performance are not more likely in the beginning or at the end of the exam as stakes rise.^{2.6.10} Moreover, 'negative shocks' related to difficult questions in the early questions do not impair women's performance. If anything, men seem to be more negatively affected by these shocks than women.

^{2.6.10} As further evidence that mental fatigue does not explain our results, we split our sample by exam days (Day 1: Portuguese and Biology; Day 2: Chemistry and History; Day 3: Physics and Geography; Day 4: Mathematics and English). If mental fatigue played a role, we would expect the gender gap in reaction to priority subjects to increase over the exam days. We do not observe this pattern. The Female × Priority coefficient is negative for all exam days and is statistically significant for all days, except for Exam Day 1. The highest impact size is the third exam day, and the lowest is the fourth. The results are available upon request.

Table 2.6.6 – Priority subjects and number of zeros, excluding omitted questions (2000-2002)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: # zeros in Phase 2 questions per subject</i>						
Female	0.037*** (0.011)	-0.158*** (0.016)	-0.117*** (0.016)			
Priority	-0.673*** (0.011)	-0.571*** (0.012)	-0.441*** (0.011)	-0.455*** (0.011)	-0.475*** (0.012)	-0.537*** (0.015)
Female × Priority	0.026 (0.016)	-0.097*** (0.018)	-0.101*** (0.017)	-0.070*** (0.016)	-0.084*** (0.016)	-0.083*** (0.016)
Norm. ENEM scores	-0.823*** (0.006)	-0.823*** (0.006)	-0.664*** (0.006)			
Mean Y	1.98					
Stand.dev Y	2.02					
Number of observations	189612	189612	189612	189612	189612	189612
Number of applicants	31602	31602	31602	31602	31602	31602
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	No	Yes

Notes: The dependent variable is the number of zeros obtained by an applicant in a given P_2 subject. The sample is restricted to 2000-2002. Zeros exclude omitted questions and include questions the applicant answered and received a score of zero. Biology is the baseline subject. ENEM scores are normalized such that the mean is 0 and standard deviation 1 for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1 for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. We use quartic functions to control for relative P_1 performance and for 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%.

Our results show that female applicants omit more questions when the stakes increase but have fewer zero scores in attempted questions from priority subjects. The observed pattern could indicate gender differences in self-confidence levels. It is possible that female applicants do not write items that they are not sure about the correct answer. Likewise, males can be more overconfident, assuming that they know the question's topic more than they actually do. As the UNICAMP admission exam consists of open-ended questions, if an applicant does not know how to answer a question, guessing could lead to a waste of time (AKYOL; KRISHNA; WANG, 2019). Thus, conditional on self-confidence levels, it may be 'optimal' to

skip the item.

We also exploit other potential channels that our evidence rules out. Gender differences in information about the UNICAMP admission exam's rule could drive our impacts. Suppose male applicants have better access to information than females. In that case, women do not focus on the priority subjects because they are not aware of the higher weight, not because they react differently to increased stakes. We estimate the main regression for different subsamples that potentially have distinct information levels about the admission exam to investigate this hypothesis. First, we split the sample into applicants taking the UNICAMP exam for the first-time and students who have taken the exam in previous years. Candidates that have applied to UNICAMP before probably know how the university calculates the final admission score. Next, we examine if the results depend on whether the applicant did a preparatory course. Applicants can enroll in preparatory classes during or after the high-school, aiming to better prepare for university admission exams. Also, UNICAMP is located in Campinas city, and nearby schools may specialize in the UNICAMP admission exam. Thus, we compare the effects for students who attended schools in the Campinas metropolitan region with students from other cities. We investigate heterogeneous impacts by school type (private or public) and parental educational level (university-educated parents vs. lower degrees). In all subsamples, we observe the same pattern of reduced women's performance in priority subjects, and the coefficients' sizes are also similar.^{2.6.11}

Finally, gender differences in reaction to increased stakes do not seem to arise because women 'choke' under pressure. To examine this mechanism, we analyze if the gender gaps in priority subjects differ in subgroups for which the stakes are higher. We compare the estimates of applicants who applied only to UNICAMP to the results of the contestants who applied to other universities (Table 2.A.8). We assume that the stakes are higher for students that do not have outside options. Results are qualitatively the same for the subgroups. As in our primary analysis, female performance reduces by nearly 0.056 s.d. in priority subjects for the applicants that applied to other universities (95%). Female underperformance in priority subjects is lower in magnitude for contestants who applied only to

^{2.6.11}Results are available upon request.

UNICAMP (0.047 s.d. reduction). The evidence suggests that a lower female performance under pressure does not seem to explain our findings.

2.7 Conclusion

This paper provides evidence that females and males react differently to increased stakes in a natural environment. In our setting, some tasks are more rewarding than others, but no-one assigns the task (the individual is free to put more effort on the more rewarding task than on the less rewarding), and the agent's response to the higher-reward request is private. Therefore, there is no apparent external pressure to accept working on the less rewarding task, unlike in [Babcock, Recalde e Vesterlund \(2017\)](#), [Babcock et al. \(2017\)](#).

The richness of our data allows us to flexibly control for applicants' major-choice self-selection issues (through the use of applicant fixed-effects combined with multiple subject-specific ability measures) – something that data rarely permit.

Our findings suggest that increasing stakes decreases females' performance relative to males, and this decrease is more pronounced for higher-ability applicants. We do not find systematic heterogeneity in the effect of increasing stakes across subjects. The gender gap analysis for different subgroups enables us to eliminate reaction to pressure and knowledge about the exam rules as potential channels. After investigating applicants' performance in more detail (at the exam question level), we rule out gender differences in performance on difficult questions, mental fatigue, and time-management issues as potential mechanisms. If anything, as stakes increase, female performance increases more relative to males on difficult questions and questions that appear late in the exam.

One intriguing finding that deserves more attention is that females become significantly more likely to omit (not answer) questions when stakes increase. Moreover, female applicants have a higher number of maximum scores in non-priority subjects, parts of the exam that count less toward the final score. Taken together, the pieces of evidence suggest that gender differences in perfectionism and self-confidence can explain our findings.

Appendix – Figures and Tables

Tables

Table 2.A.1 – Priority subjects by undergraduate program

Major	Port.	Bio.	Chem.	Hist.	Geo.	Math.	Phys.	Cutoff	Prop. Female	Prop. App.
Medicine UNICAMP		✓	✓					616	0.614	0.204
Medicine Famerp		✓	✓					600	0.606	0.059
Computer Engineering						✓	✓	584	0.116	0.054
Electrical Engineering						✓	✓	582	0.105	0.032
Computer Science (Eve.)						✓	✓	578	0.176	0.037
Control and Automation Engineering (Eve.)						✓	✓	575	0.066	0.032
Economics				✓		✓		575	0.412	0.032
Electrical Engineering (Eve.)						✓	✓	570	0.076	0.013
Social Communication (Media Studies)				✓		✓		565	0.578	0.007
Pharmacy		✓	✓					560	0.787	0.008
Chemical Engineering			✓			✓		557	0.456	0.018
Economics (Eve.)				✓		✓		556	0.388	0.017
Food Engineering						✓	✓	552	0.781	0.027
Biological Sciences		✓						549	0.676	0.045
Mechanical Engineering						✓	✓	545	0.063	0.029
Chemical Engineering (Eve.)			✓			✓		539	0.356	0.009
Food Engineering (Eve.)						✓	✓	518	0.734	0.010
Social Sciences	✓			✓				507	0.590	0.015
History	✓			✓				503	0.523	0.016
Language Studies (TB)	✓			✓				501	0.818	0.011
Civil Engineering						✓	✓	493	0.270	0.018
Chemistry Technology (Eve.)			✓					487	0.504	0.008
Social Sciences (Eve.)	✓			✓				484	0.521	0.015
Chemistry			✓					480	0.611	0.015
Biological Sciences (T) (Eve.)		✓						480	0.675	0.020
Geology and Geography					✓			461	0.411	0.008
Physics (Eve.)						✓	✓	460	0.163	0.007
Physics, Math, Applied Math and Comp						✓	✓	455	0.323	0.022
Language Studies (T) (Eve.)	✓			✓				452	0.779	0.009
Philosophy	✓							447	0.430	0.006
Chemistry and Physics (T) (Eve.)			✓				✓	445	0.384	0.005
Linguistics	✓			✓				445	0.767	0.004
Geography (Eve.)					✓			441	0.369	0.007
Dentistry		✓						440	0.717	0.033
Pedagogy (T)	✓			✓				422	0.941	0.009
Phonology	✓	✓						418	0.969	0.006
Information Technology (Eve.)						✓		416	0.261	0.010
Pedagogy (T) (Eve.)	✓			✓				415	0.920	0.011
Physical Education		✓		✓				405	0.488	0.018
Agricultural Engineering						✓	✓	405	0.294	0.010
Telecommunications Technology						✓		403	0.201	0.001
Nursing UNICAMP		✓						402	0.915	0.018
Mathematics (T) (Eve.)						✓	✓	395	0.419	0.008
Physical Education (Eve.)		✓		✓				392	0.440	0.015
Statistics						✓	✓	382	0.471	0.009
Information Technology						✓		360	0.333	0.004
Nursing Famerp		✓						357	0.921	0.012
Environmental Sanitation Technology (Eve.)						✓		328	0.536	0.010
Construction Technology (Eve.)						✓		324	0.396	0.006
Environmental Sanitation Technology						✓		310	0.594	0.001

Note: Majors with (Eve.) are programs offered in the evening. (T) and (TB) refer to teaching majors (“licenciaturas”) and teaching/bachelor majors, respectively. Prop. Female is the proportion of applicants to the major that are females, and Prop. App. is the proportion of students applying to the major (both measured over the 2000-2004 period).

Table 2.A.2 – Priority subjects and gender performance gap (Raw Phase 2 scores)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Phase 2 subject-specific raw scores</i>						
Female	0.564*** (0.049)	1.713*** (0.074)	1.379*** (0.067)			
Priority	6.288*** (0.047)	5.876*** (0.051)	4.529*** (0.047)	4.670*** (0.045)	4.727*** (0.050)	4.984*** (0.058)
Female $\times$ Priority	-1.198*** (0.065)	-0.783*** (0.070)	-0.744*** (0.065)	-0.843*** (0.062)	-0.763*** (0.063)	-0.730*** (0.064)
ENEM scores	0.428*** (0.002)	0.428*** (0.002)	0.312*** (0.002)			
Mean dependent variable	25.63					
Stand dev dependent variable	10.54					
Number of observations	323694	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949	53949
Year Fixed-Effects	Yes	Yes	Yes	No	No	No
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM $\times$ Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores $\times$ Priority	No	No	No	No	No	Yes

Notes: We use raw ENEM, Phase 1 and Phase 2 scores in the regressions. All specifications controls for year fixed-effects (individual fixed-effects absorb the year fixed-effects). P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A.3 – Priority subjects and gender performance gap (Including Portuguese)

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Phase 2 normalized subject-specific scores</i>						
Female	0.153*** (0.005)	0.170*** (0.007)	0.143*** (0.006)			
Priority	0.608*** (0.004)	0.550*** (0.004)	0.430*** (0.004)	0.465*** (0.004)	0.472*** (0.004)	0.487*** (0.005)
Female $\times$ Priority	-0.141*** (0.006)	-0.067*** (0.006)	-0.062*** (0.006)	-0.063*** (0.006)	-0.057*** (0.006)	-0.057*** (0.006)
Norm. ENEM scores	0.536*** (0.003)	0.537*** (0.003)	0.414*** (0.002)			
Number of observations	377643	377643	377643	377643	377643	377643
Number of applicants	53949	53949	53949	53949	53949	53949
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM $\times$ Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores $\times$ Priority	No	No	No	No	No	Yes

Notes: ENEM scores are normalized such that the mean is 0 and standard deviation 1, for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1, for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A.4 – Priority subjects of third and second major choices

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Priority subjects 2nd choice</i>						
Female	0.069*** (0.006)	0.169*** (0.008)				
Priority 2 nd Choice	0.566*** (0.005)	0.537*** (0.005)	0.405*** (0.005)	0.442*** (0.004)	0.450*** (0.005)	0.470*** (0.006)
Female $\times$ Priority 2 nd Choice	-0.068*** (0.006)	-0.035*** (0.007)	-0.008 (0.006)	-0.054*** (0.006)	-0.048*** (0.006)	-0.047*** (0.006)
Norm. ENEM scores	0.554*** (0.003)	0.555*** (0.003)	0.410*** (0.003)			
Number of observations	265476	265476	265476	265476	265476	265476
Number of applicants	44246	44246	44246	44246	44246	44246
<i>Panel B: Priority subjects 3rd choice</i>						
Female	0.009 (0.008)	0.100*** (0.011)				
Priority 3 rd Choice	0.676*** (0.008)	0.626*** (0.011)	0.477*** (0.009)	0.508*** (0.009)	0.513*** (0.010)	0.527*** (0.011)
Female $\times$ Priority 3 rd Choice	-0.138*** (0.010)	-0.075*** (0.014)	-0.063*** (0.012)	-0.057*** (0.012)	-0.051*** (0.012)	-0.051*** (0.012)
Norm. ENEM scores	0.492*** (0.004)	0.493*** (0.004)	0.359*** (0.003)			
Number of observations	137148	137148	137148	137148	137148	137148
Number of applicants	22858	22858	22858	22858	22858	22858
Subject fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject-gender fixed-effects	No	Yes	Yes	Yes	Yes	Yes
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes
Individual fixed-effects	No	No	No	Yes	Yes	Yes
Relative perf. ENEM $\times$ Priority	No	No	No	No	Yes	Yes
Relative perf. P1 scores $\times$ Priority	No	No	No	No	No	Yes

Notes: ENEM scores are normalized such that the mean is 0 and standard deviation 1, for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1, for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A.5 – Priority, very difficult questions and P_2 performance

	Score	Perfect Score	Omission	Zero score
Priority	0.507*** (0.005)	0.069*** (0.001)	-0.043*** (0.001)	-0.051*** (0.001)
Female $\times$ Priority	-0.128*** (0.006)	-0.031*** (0.001)	0.015*** (0.001)	0.001 (0.001)
Very difficult question	-1.175*** (0.002)	-0.088*** (0.000)	0.063*** (0.001)	0.161*** (0.001)
Female $\times$ Very difficult question	-0.028*** (0.004)	-0.003*** (0.001)	0.002** (0.001)	0.029*** (0.002)
Priority $\times$ Very difficult question	-0.377*** (0.005)	-0.133*** (0.001)	0.015*** (0.001)	0.027*** (0.002)
Female $\times$ Priority $\times$ Very difficult question	0.247*** (0.007)	0.071*** (0.001)	-0.031*** (0.002)	-0.031*** (0.003)
Mean Y	2.14	0.11	0.06	0.17
Std.dev Y	1.65	0.31	0.24	0.37
# observations	3884328	3884328	2275344	2275344
# applicants	53949	53949	31602	31602
Subject FE	Yes	Yes	Yes	Yes
Subject-gender FE	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes
Relative performance P1 scores	Yes	Yes	Yes	Yes
Relative perf. ENEM $\times$ Priority	Yes	Yes	Yes	Yes
Relative perf. P1 scores $\times$ Priority	Yes	Yes	Yes	Yes

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

Table 2.A.6 – Gender differences in Phase 2 normalized scores

	(1)	(2)	(3)
<i>Dependent variable: Phase 2 normalized subject-specific scores</i>			
First question in subject is difficult	0.006 (0.006)		
Female $\times$ First question is difficult	0.006 (0.008)		
Most difficult question is among the first 4		-0.009** (0.004)	
Female $\times$ Most difficult in first 4		0.017*** (0.006)	
Very difficult question among the first 4			-0.018*** (0.005)
Female $\times$ very difficult in first 4			0.030*** (0.007)
# observations	323694	323694	323694
# applicants	53949	53949	53949
Subject FE	Yes	Yes	Yes
Subject-gender FE	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes
Relative performance P1 scores	Yes	Yes	Yes
Relative perf. ENEM $\times$ Priority	Yes	Yes	Yes
Relative perf. P1 scores $\times$ Priority	Yes	Yes	Yes

Notes: Subject-specific scores are normalized such that the mean is 0 and standard deviation 1, for each subject-year. We classify a question in a specific subject as *difficult* if the average performance on this question is below the median of question average scores for that subject, and a very difficult question if the average score is among the lowest 25% in the exam (subject-year). Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2.A.7 – Probability of choosing priority subjects

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Priority subject dummy</i>						
P1 normalized subject-specific scores	0.073*** (0.001)	0.067*** (0.001)	0.064*** (0.001)	0.065*** (0.001)	0.093*** (0.001)	0.078*** (0.001)
Female × P1 normalized scores	-0.004*** (0.001)	0.009*** (0.001)	0.007*** (0.001)	0.007*** (0.001)	-0.010*** (0.002)	0.007*** (0.002)
Female	-0.026*** (0.001)	-0.025*** (0.001)	0.228*** (0.004)	0.226*** (0.004)		
Norm. ENEM scores				-0.003*** (0.000)		
Number of observations	323694	323694	323694	323694	323694	323694
Number of applicants	53949	53949	53949	53949	53949	53949
Subject fixed-effects	No	Yes	Yes	Yes	No	Yes
Subject-gender fixed-effects	No	No	Yes	Yes	No	Yes
Individual fixed-effects	No	No	No	No	Yes	Yes

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

Table 2.A.8 – Only UNICAMP × Other admission processes

	Only UNICAMP						Other admission processes					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Dependent variable: Phase 2 normalized subject-specific scores</i>												
Female	0.006 (0.023)	0.105*** (0.032)	0.094*** (0.027)				0.067*** (0.005)	0.165*** (0.007)	0.135*** (0.007)			
Priority	0.665*** (0.020)	0.645*** (0.021)	0.463*** (0.020)	0.508*** (0.020)	0.524*** (0.022)	0.533*** (0.024)	0.575*** (0.004)	0.554*** (0.004)	0.434*** (0.004)	0.458*** (0.004)	0.468*** (0.005)	0.488*** (0.005)
Female × Priority	-0.074*** (0.026)	-0.054* (0.028)	-0.040 (0.026)	-0.053** (0.026)	-0.046* (0.026)	-0.047* (0.026)	-0.067*** (0.006)	-0.049*** (0.006)	-0.050*** (0.006)	-0.062*** (0.006)	-0.056*** (0.006)	-0.056*** (0.006)
Norm. ENEM scores	0.419*** (0.010)	0.419*** (0.010)	0.295*** (0.008)				0.546*** (0.003)	0.546*** (0.003)	0.412*** (0.002)			
# observations	17148	17148	17148	17148	17148	17148	306042	306042	306042	306042	306042	306042
# applicants	2858	2858	2858	2858	2858	2858	51007	51007	51007	51007	51007	51007
Subject FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	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
Relative performance P1 scores	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Relative perf. ENEM × Priority	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Relative perf. P1 scores × Priority	No	No	No	No	No	Yes	No	No	No	No	No	Yes

Notes: ENEM scores are normalized such that the mean is 0 and standard deviation 1, for each year. Subject-specific scores are normalized such that the mean is 0 and standard deviation 1, for each subject-year. The P_1 subject relative performance score is the applicants' P_1 subject score minus her/his gender-group average the year of application. The ENEM relative performance is the applicants' subject score minus her/his gender-group normalized ENEM average the year of application. Cluster-robust standard errors (at the applicant level) are shown in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

3 Gender peer effects on major choice

3.0.1

3.1 Introduction

Even though women have surpassed men in many educational outcomes, female students remain underrepresented in math-intensive and competitive fields, hampering women's career prospects and widening the gender wage gap (BROWN; CORCORAN, 1997; MACHIN; PUHANI, 2003; KIRKEBOEN; LEUVEN; MOGSTAD, 2016). A vast literature has shown that peers can have meaningful impacts on students' behaviors and decisions (SACERDOTE, 2014). In Brazil, students choose their major at the end of high-school at the university application processes. Thus, in our context, high-school peers may be especially relevant in their career choice.

This paper examines if high-school peers' gender composition influences students' major choice, especially for women. I ask the following questions: Are female students exposed to higher shares of female peers more likely to select math-intensive fields of study? Does the gender composition of high-school peers impact the likelihood of choosing female or male-dominated fields? Are women with higher shares of female schoolmates more likely to choose competitive majors, i.e., majors with higher admission cutoffs? I use admission data from a selective university, UNICAMP, linked with public administrative data containing Brazilian high schools' universe to answer these questions. UNICAMP is a large and highly prestigious public university located in Sao Paulo state, the country's wealthiest state. The empirical strategy to identify causal peer effects is to exploit the idiosyncratic variation in students' gender composition within schools across cohorts.

^{3.0.1}Financial support from Sao Paulo Research Foundation (FAPESP - grant # 2017/23593-8) and from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES - Finance Code 001) are gratefully acknowledged. I am grateful to COMVEST, UNICAMP's admission office, for providing the data and assistance during the project. I thank Leonardo Rosa for his help in the collection of Brazilian School Census Data.

The gender composition of peers in high-school could influence the choice of the university major through different channels. First, the gender composition in high-school can affect students' achievement, which could alter career decisions, as students who have higher grades are more likely to be admitted into more competitive and monetary rewarding majors. The previous literature shows that a higher proportion of female peers improve the learning environment, resulting in a better performance for both women and men (HU, 2015; LAVY; SCHLOSSER, 2011; WHITMORE, 2005; HOXBY, 2000). Specifically, in Brazil, Firpo, Jales e Pinto (2015) estimate gender peer effects in elementary schools and verify that the percentage of female students in a classroom increases mathematics performance, although females underperform in the subject.

It is also worth noting that gender peer impacts on performance can differ by gender, and some papers do find evidence of heterogeneous effects. Black, Devereux e Salvanes (2013) find heterogeneous gender peer impacts for female and male students from Norway high-schools. Women have better educational outcomes when female peers' share increases, but they find the opposite effect for men. Hill (2015) finds a similar pattern in the United States. The paper verifies that the proportion of friends from the opposite gender reduces students' performance during high-school. Therefore, a rise in female schoolmates' percentage would decrease achievement for men and increase for women.

The gender composition of high-school peers could also impact perceptions and aspirations by influencing the perceived gender stereotypes. Dasgupta, Mcmanus e Hunsinger (2015) analyze if higher shares of female peers in a male-dominated environment, the Engineering major, impact women's outcomes. The study finds that the group's gender composition affects women's behavior in different aspects. Women assigned to higher shares of female peers had lower anxiety levels, participated more in verbal activities, and had less gender-stereotyped views about Engineering. Besides, females assigned to the male-majority groups presented lower self-confidence levels and less ambitious career goals.

Another possibility is that female peers in high-school could impact women's major choices through changes in the reaction to competition. Even when there are no gender differences in performance under competition, women are less likely to

choose to perform in a competitive setting (NIEDERLE; VESTERLUND, 2007). The gender composition of peers can influence sex differences in competitiveness. Gneezy, Niederle e Rustichini (2003) run experiments in non-competitive environments and find no gender gaps in performance. However, under competition, male performance increases, but female participants do not change their achievement. The gender composition of groups influences the gender differences in performance in competitive environments: in single-sex settings, women perform better under competition than in the non-competitive experiment, reducing the gender gap in achievement.

Moreover, the gender composition of schoolmates could influence gender differences in risk aversion. Booth, Cardona-Sosa e Nolen (2014) randomly assigns first-year university students to three different treatment groups: all-female, all-male, and coeducational classes. The paper aims to test if classmates' gender composition influences female risk-taking behavior, measured by lotteries in a controlled experiment. Initially, women are more risk-averse than men in all treatment groups. However, eight weeks after the initial assignment, the female students from single-sex classes are more likely to make risky choices than females assigned to the mixed-gender groups. The authors do not find impacts for men assigned to single-sex groups.

Gender differences in psychological attributes, such as competitiveness and risk aversion, could relate to gender differences in career choices. As women are less competitive and more risk-averse, they may decide to play it safe and choose majors with lower admission scores or applicants per slot. The literature finds that experimental measures of willingness to compete are positively correlated with students' choices of math and science-intensive fields and predict the choice of more prestigious fields of study. Gender differences in competitiveness partially account for the observed gender gaps in career choice (BUSER; NIEDERLE; OOSTERBEEK, 2014; BUSER; PETER; WOLTER, 2017; NIEDERLE, 2017).

In the existing literature, some papers evaluated if the gender composition of peers influences career choice.^{3.1.1} The articles examine different traits of the

^{3.1.1}In Brazil, we have a coeducational system and single-sex schools are extremely rare. Jackson (2012) finds in Trinidad and Tobago that female students that attend single-sex schools are more

major choice, and the most commonly analyzed features are math-intensity and the gender composition of the field. The results found so far are mixed.^{3.1.2}

Schøne, Simson e Strøm (2020) analyze gender peer effects on students' educational choices in secondary schools in Norway. The study verifies that students exposed to higher shares of female peers are more likely to select STEM subjects instead of language disciplines. When investigating potential mechanisms that could explain these findings, the paper finds that, for female students, a higher women representation increases performance in Mathematics and leads to lower gender discrimination. In turn, for men, the percentage of female peers increases their willingness to compete.

At the high-school level, Goulas, Megalokonomou e Zhang (2018) investigate gender peer effects at Greek public schools on major choice. The paper estimates the impacts of both female classmates and neighbors. Women with a higher share of female peers during high-school are more likely to select STEM fields and enroll in majors with higher expected earnings. The effects hold both when they define peers as neighbors or classmates. However, for the same educational level, Brenøe e Zölitz (2020) and Mouganie e Wang (2020) find opposite gender peer impacts. Brenøe e Zölitz (2020) examine if female high-school peers' share impacts the likelihood of selecting a STEM major using data from Danish schools. The results show that higher shares of female peers decrease the probability of choosing STEM fields for female students, and it increases the probability for males. Looking at long-term labor market impacts, female peers' share leads to higher occupational segregation among genders, resulting in lower female future earnings. Mouganie e Wang (2020) study for Chinese high-schools if the gender composition of peers and high-performing peers (measured by mathematical ability) affects STEM choices. They find that higher shares of high-achieving female peers increase female students' likelihood of choosing science tracks and decrease this probability for men. For women, the impacts persist when looking at university outcomes, as they are more likely to choose a STEM major. Nevertheless, the impact of the gender composition

likely to take male-dominated science courses.

^{3.1.2}Some studies examine if older siblings, and their gender composition, impact career choices of the younger children (JOENSEN; NIELSEN, 2018; ALTMEJD et al., 2019).

of overall high-school peers is quite different. While there are no gender peer effects for female students, males exposed to higher shares of female peers are more likely to pursue a STEM career.

Other papers evaluate if the gender composition of peers affects the probability of selecting a female or male-dominated field. [Schneeweis e Zweimüller \(2012\)](#) examine the impacts of female peers' share for students of secondary schools in Austria. The study assesses if the gender composition of peers in grades 5 to 8 influence the choice of male-dominated school type in the ninth grade. They find that female students are less likely to choose female-dominated schools when exposed to higher proportions of female peers.

At the higher education level, [Zölitz e Feld \(2020\)](#) examine if the percentage of female peers in course sections of undergraduate programs from a Dutch business school influence students' major choices and labor market outcomes, and if the effects vary by gender. They use the random assignment of students to sections in compulsory courses as a source of exogenous variation in the gender of female peers. Female students' exposure to higher percentages of female classmates reduces the likelihood of choosing male-dominated majors and increases female-dominated career choices. For men, the opposite is true, as, for them, higher shares of female classmates lead to a greater likelihood of male-dominated fields and reduce their probability of female-dominated majors. The impacts on major choice result in lower female earnings in the labor market, while there are no effects for men. The share of female peers also reduces working hours for females and increase their job satisfaction.

Closely related to this study, [Anelli e Peri \(2019\)](#) examine in the city of Milan, Italy, the impact of the gender composition of high-school classmates on major choice. The paper assesses if a higher percentage of female peers influence the choice of female or male-dominated fields and if the differences in career choices affect long-term labor market outcomes. An advantage of this paper is that Italian high-schools randomly assign students to classes, allowing the authors to estimate the causal impacts of female classmates. In general, the article does not find gender peer effects on major choices. However, male students exposed to extremely high shares of female peers (above 80%) are more likely to select a male-dominated major.

Low-achieving male students drive this impact. Due to their higher major attrition rates, this effect does not lead to differences in the probability of graduating from a male-dominated field or in labor market outcomes. There are no short or long-term impacts of the share of female high-school peers on career selection for women.

This paper contributes to the literature in different ways. First, most articles focus on high-school gender peer effects for developed nations, while in this paper, I focus on the Brazilian context, a developing country. Another advantage of the Brazilian educational system is that, like in some European countries, in Brazil, the students will select their major at the end of high-school, before university entrance, when they apply for the university admission exams. Therefore, in this setting, the impact of high-school peers can be especially relevant. Moreover, UNICAMP is a large university that offers slots for several different majors. Thus, we have a comprehensive sample of applicants (almost 140 thousand participants) and research fields. Additionally, as I have historical UNICAMP information on how difficult the admission in a specific major is (average final admission cutoff) and the number of applicants per slot, I can construct competitiveness measures for the fields of study, response variables neglected by the previous literature. Finally, as students who register at the UNICAMP admission exam have to answer a detailed questionnaire, I have data on a host of socioeconomic characteristics and reasons to choose the university and major, enabling me to exploit potential mechanisms driving the results.

The results show that the percentage of female high-school peers impact the major choice, and the gender peer effects differ for men and women. First, I find that a larger female peers' share reduces the candidates' probability of applying for a math-intensive or science field for both female and male applicants. Regarding the gender composition of the major choice, I verify that higher percentages of own-gender peers increase the probability of selecting a male-dominated field and reduces the likelihood of choosing a gender-balanced field. Finally, only for women, the proportion of female schoolmates increases the degree of competitiveness of the major choice. Female applicants exposed to higher female peers' shares in high school select majors with higher admission cutoffs.

Although we find that the gender composition of peer influences major

choice, the impact sizes are modest when I linearly include female peers' share in the regressions. When I estimate the impacts of having a female majority cohort (above 50%), the effects are qualitatively similar, and impact sizes are slightly larger, suggesting non-linear gender effects. Another interesting finding when evaluating non-linear impacts is that female applicants in schools with extremely low percentages of female peers drive most of the results. Women who studied in high-schools with less than one-third of female colleagues are more likely to select gender-balanced fields and less likely to choose male-dominated majors at UNICAMP. Also, female applicants exposed to low proportions of female peers in high-school ($\leq 33\%$) choose less competitive majors, with lower UNICAMP's admission cutoffs.

Gender peer impacts are more substantial for students with less-educated mothers, suggesting that female role models at home could partially explain our results. To shed light on other underlying mechanisms, I investigate whether high-school gender composition affects performance and the reasons why the applicants have chosen UNICAMP university and the major. As female peers' share increases overall performance, I would expect applicants to choose more competitive majors, but I only verify impacts on women's choice of more competitive fields. A possible feature that could explain the gender differences is that females are more responsive to performance changes than men ([RASK; TIEFENTHALER, 2008](#); [KUGLER; TINSLEY; UKHANEVA, 2017](#)). I do not observe compelling effects on reasons to choose UNICAMP or the major. Nevertheless, female peers' share increases women's degree of certainty in their career choice, suggesting that exposure to a higher female representation raises women's confidence levels.

The paper is organized as follows. Section 3.2 briefly describes the Brazilian institutional background and UNICAMP admission exam. In the following section, I present the data sets used in the analysis and report some descriptive statistics. In Section 3.4, I present the empirical strategy and provide evidence of the validity of the identification strategy. In Section 3.5, I report the main results and robustness checks. Section 3.6 present heterogeneous gender peer impacts, considering students' and schools' characteristics. In Section 3.7, I discuss potential channels that could explain the results. Lastly, Section 3.8 concludes.

3.2 Institutional background

In my study, I evaluate gender peer effects in Brazilian high-schools. In Brazil, students begin basic education when they are six or seven years old, and this educational stage lasts for nine years (five years of elementary school and four years of middle school). If there is no grade retention or dropout, the student will be between 15 and 16 years old when she starts the high-school. Upper-secondary education has a duration of three years. The Brazilian federal legislation requires that individuals attend schools until they are 17 years old. Therefore, high-school attendance is, in theory, compulsory. Nevertheless, the federal government does not enforce the law (ROSA, 2019).

High-school students can attend public (state, federal, or city) or private high-schools. While public schools are tuition-free, a private school can charge tuition and other school fees. Private education is not heavily regulated in Brazil, although schools must meet minimal requirements to operate (ROSA, 2019). Selection into public schools generally depends on slot availability and distance to the student's residence.^{3.2.1} Families can freely choose schools based on schools' characteristics, such as peers' composition.^{3.2.2} Typically, private high-schools prepare students better for university admission exams than public high-schools.

In Brazil, most students attend public high-schools, and only 14.4% enroll in private institutions (INEP/MEC, 1998–2003).^{3.2.3} However, since UNICAMP is a selective university, 69% of the students attended private high-schools in our sample. Thus, our empirical strategy should deal with the self-selection of the students into schools.

To obtain information on students' major choices after high-school graduation, I use admission exam data from a prestigious Brazilian university, UNICAMP. UNICAMP is a large public university located in Sao Paulo state, the wealthiest

^{3.2.1}A few technical state or federal schools use admission exams to select the highest-ranked students (in terms of admission scores).

^{3.2.2}Some prestigious private institutions can also have admission exams to select high-achieving students.

^{3.2.3}The share is similar when I restrict the analysis to Sao Paulo state: 14% of high-school enrollments were in private institutions.

state in the country. Like most public universities in Brazil, UNICAMP is tuition-free. Besides being tuition-free, UNICAMP has a good reputation and, therefore, is a highly selective university. Students interested in attending UNICAMP must pass a competitive admission process.

The admission process for individuals interested in attending an undergraduate program at UNICAMP consists of a university-specific entrance exam, known as *vestibular*, held once a year. In Brazil, students usually decide their major before university entrance. At UNICAMP, the contestants will choose the major when they register for the application process and are allowed to apply to up to three majors (ranked by individual preferences).

UNICAMP admission exam is composed of two different stages, Phase 1 and Phase 2. All UNICAMP applicants answer precisely the same exams in both stages, regardless of the major choice. In Phase 1, applicants write an essay and answer twelve questions (two per subject) on six different subjects (Biology, Chemistry, Geography, History, Mathematics, and Physics). Only applicants who pass Phase 1 (major-specific cutoff) will attend the second stage. Phase 2 covers the same subjects, in addition to tests on Portuguese and a Foreign Language (English or French). The Phase 2 exam consists of twelve questions per subject. The major choice will define one or two ‘priority subjects’ in the second stage with a higher weight (double) when calculating the final admission score. The formula to calculate the final UNICAMP admission score adds standardized Phase 1 and 2 scores: Phase 1 final score with a weight of 2, Phase 2 priority subjects with a weight of 2, and non-priority Phase 2 subjects weight 1. Finally, UNICAMP ranks the applicants in decreasing admission score order and uses a version of the Boston Mechanism to assign students to the major-specific available slots. For detailed information on the UNICAMP admission exam, see [Estevan, Gall e Morin \(2019\)](#).

3.3 Data

The primary data source is the UNICAMP admission exam data. UNICAMP data is gathered by Comissão Permanente para os Vestibulares (COMVEST), the

university admission office. UNICAMP data provides detailed Phase 1 and Phase 2 performance, major choices, end-of-high-school test scores (ENEM scores),^{3.3.1} and a host of socioeconomic characteristics.

My sample comprises individuals that applied to UNICAMP between 2000 and 2004, focusing on years previous to UNICAMP's affirmative action policies. I impose some data restrictions to create the final sample. First, I exclude trainee students, i.e., students that have done the admission exam as a practice test. Also, I need information on the year of high-school graduation and school identification to merge UNICAMP data with Brazilian School Census data. Therefore, I excluded applicants that have missing values on any of these variables (10% of applicants do not have information on graduation year or school name).

Additionally, in the Brazilian School Censuses previous to the 1998 year, there is no information on high-school students' gender composition. Thus, I have to restrict the sample to students that graduated from high-school after 1997 (the restriction excludes 11% of the candidates). Finally, I dropped observations with missing values on crucial applicants' characteristics (gender, ENEM scores, and age), excluding 8% of the initial sample. The restricted UNICAMP sample contains 156,855 applicants.

I link UNICAMP data with Brazilian high school data using school name and graduation year. I gathered high-school information from the Brazilian School Census (*Censo Escolar*) (INEP/MEC, 1998–2003), a public dataset collected by the Ministry of Education that contains the universe of Brazilian schools. I focus on school years between 1998 and 2003.^{3.3.2} *Censo Escolar* provides school information, such as infrastructure, whether the school is public or private, the number of employees, location, etc., and data on a few students' characteristics by grade and

^{3.3.1}ENEM is a national standardized test that students take in their last year of high-school. Although the ENEM exam is not compulsory for high-school students, it is a high-stake exam, as students can use their grades in many different university admission processes. At UNICAMP admission exam, ENEM can be used in the composition of Phase 1 final score, accounting for up to 20% of the P1 score.

^{3.3.2}For previous high-school cohorts, I cannot assess the share of female peers in the last year of high-school. Moreover, as I am focusing on the 2000 to 2004 UNICAMP admission years, and as high-school students must graduate from high-school to enter a university, our sample of applicants have graduated before 2004.

school years, such as gender and age composition.

I have successfully merged UNICAMP and Census data for 95% of the UNICAMP restricted sample (148,849 applicants). I have excluded from the main sample inconsistencies between the two data sources, resulting in an additional 5% reduction in the number of observations.^{3.3.3} Subsequently, I dropped students that were not in the regular educational system: indigenous education and high-schools with no grade separation (401 applicants). Also, I excluded candidates that attended single-sex schools during the high-school (19 students). Lastly, I dropped applicants who did not have any schoolmates applying to UNICAMP (915 observations). For them, I cannot include school fixed-effects in the main regressions. The final sample contains 139,896 applicants, 3,643 schools, and 13,595 school-cohort observations.

3.3.1 Descriptive statistics

Table 3.A.1 presents descriptive statistics for the final sample of UNICAMP applicants. The table reports the statistics for the full sample and separately by gender. The column ‘Diff.’ shows the differences in group averages (Female - Male). We note that UNICAMP is a gender-balanced admission exam, 51% of the candidates are women. Female applicants have lower socioeconomic status than men: they come from less affluent families (lower family income and parental education) and are more likely to attend public elementary and middle schools. Although male candidates come from higher-income households, they are more likely to work during the high-school (15% of women work and 19% of men) and have higher grade retention rates in high-school.

Table 3.A.2 describes the high-school data (*Censo Escolar*) for our sample of applicants. As in Table 3.A.1, I present the averages for the whole sample and separate means for men and women, and the last column reports differences in group averages. Female applicants are exposed to higher shares of female schoolmates (leave-me out mean) in their last year of high-school: 53% of their peers are women,

^{3.3.3}The first data restriction is to exclude students for which information diverged in the datasets, such as location (state), public/private schools, etc. I have also excluded schools that had not reported any high-school enrollments in the respective school year.

while for men, this share is 51%. The vast majority of students attended high-schools located in Sao Paulo state (versus schools in other Brazilian states). This percentage is higher for women (93% for women and 91% for men). Although most students attend private high-schools (around 70%), there are some gender differences. Women are more likely to attend state schools and less likely to attend private or federal schools than men.

In Table 3.3.1, I report the means and standard deviations for the outcome variables used in the analysis. The paper's primary goal is to examine if higher percentages of female peers during high-school influence female students' major choices. First, I look at measures of math-intensity in a major. I create two different dummy variables that indicate a math-intensive field: a dummy that equals one if the major has Mathematics or Physics as priority subjects, and zero otherwise;^{3.3.4} and a binary variable that equals one if the applicant has chosen a Science, Technology, Engineering, and Mathematics (STEM) major, and zero for other fields of study. I use the International Standard Classification of Education (ISCED 2011) to define STEM majors (UNESCO, 2011). The fields of education considered as STEM are Life Sciences, Physical Sciences, Engineering, Computing, Mathematics and Statistics.^{3.3.5} Female applicants are less likely to apply for math-intensive fields (26% for women versus 60% for men) and STEM majors (34% of female applicants and 60% of the male applicants).

Second, I assess female peers' impact on the likelihood of choosing a female

^{3.3.4}The following majors have Mathematics or Physics as priority disciplines: Agricultural Engineering, Architecture and Urban Planning, Chemical Engineering, Civil Engineering, Computer Engineering, Computer Science, Construction Technology, Control and Automation Engineering, Economics, Electrical Engineering, Environmental Sanitation Technology, Food Engineering, Information Technology, Mathematics, Mechanical Engineering, Physics, Applied Mathematics, Social Communication (Media Studies), Statistics, and Telecommunications Technology. The Philosophy major had Math and Portuguese as priority subjects in 2000 and 2001, but from 2002 on its only priority was Portuguese.

^{3.3.5}I have included the following UNICAMP majors in the STEM classification: Applied Mathematics, Architecture and Urban Planning, Agricultural Engineering, Biological Sciences, Civil Engineering, Chemical Engineering, Chemistry, Chemistry Technology, Computer Engineering, Computer Science, Control and Automation Engineering, Construction Technology, Electrical Engineering, Environmental Sanitation Technology, Food Engineering, Geology and Geography, Information Technology, Mathematics, Mechanical Engineering, Physics, Statistics, Telecommunications Technology.

or male-dominated field. I calculate female applicants' share by major choice, considering the universe of UNICAMP applicants for all available years (2000 - 2008). Using the female shares by major, I constructed four different measures of the gender composition of a field: the average percentage of female applicants, and dummy variables for female-dominated, male-dominated, and gender-balanced fields. I rank majors in decreasing order of female candidates' share and consider the career as female (male) dominated if the major is amongst the 25% higher (lower) female percentages. The remaining majors are considered gender-balanced fields. As expected, women are more concentrated in female-dominated areas, while men are more likely to apply for male-dominated fields. It is interesting to note that women are also more prone to choose gender-balanced fields than men.

Finally, I also investigate the impacts of female peers on the probability of choosing competitive fields of study. To measure the competitiveness of a major, I create two variables using UNICAMP admission data and calculating the averages for the admission years between 2000 and 2008 by career choice (for the universe of applicants): Total number of applicants divided by the total number of available slots; and Final UNICAMP admission cutoff, the final minimum score that an applicant needs to obtain for UNICAMP admission (major-specific). Female candidates choose fields with a higher number of applicants per slot, but the final admission cutoff of their major choices is 7 points lower.

I also examine potential mechanisms that could be driving our results. I present the summary statistics for these variables in Tables 3.A.3, 3.A.4, 3.A.5. One potential channel that could lead to gender differences in the major choice is the gender gap in performances. Female applicants have lower ENEM scores (0.4 standard deviations lower) and underperform in Stage 1. The lower Phase 1 performance appears in all subjects. The only exception is the essay score, where women outperform men. In Tables 3.A.4 and 3.A.5, I present the distribution for women and men of their declared reasons to choose UNICAMP and the major. Women are more prone to apply to UNICAMP because the university is tuition-free and has a good reputation, and less likely to choose because UNICAMP is the best university for their career choice. In terms of career decisions, men are more likely to declare their choice based on comparative advantage (personal ability). In

contrast, women declare more often that their choice is based on preferences and desire to contribute to society.

Table 3.3.1 – Outcome variables

	All	Female	Male	Diff.
Math or Physics are priority subjects	0.42 (0.49)	0.26 (0.44)	0.60 (0.49)	-0.33***
STEM Major	0.47 (0.50)	0.34 (0.47)	0.60 (0.49)	-0.27***
Male-dominated field	0.26 (0.44)	0.08 (0.26)	0.45 (0.50)	-0.38***
Female-dominated field	0.16 (0.37)	0.26 (0.44)	0.06 (0.23)	0.21***
Gender-balanced field	0.58 (0.49)	0.66 (0.47)	0.49 (0.50)	0.17***
Average share of female applicants in career (2000-2008)	0.50 (0.24)	0.62 (0.18)	0.38 (0.24)	0.24***
Average # applicants per slot (2000-2008)	33.45 (25.94)	36.49 (27.70)	30.25 (23.53)	6.24***
Average admission scores cutoff (2000-2008)	525.66 (72.80)	522.14 (78.60)	529.38 (65.93)	-7.24***
Observations	139,896	71,742	68,154	139,896

Notes: The table presents the descriptive statistics of the dependent variables by gender. The column “Difference” reports the coefficient of a t-test of mean differences between groups. Standard deviations in parentheses. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.4 Empirical strategy

The main empirical challenge to estimate causal peer effects is that students usually sort into groups. The self-selection of students into schools and classrooms, defined by [Manski \(1993\)](#) as ‘correlated peer effects’, hinders the identification of the peer impacts that emanate directly from schoolmates’ characteristics.

In this paper, I use a methodology proposed by [Hoxby \(2000\)](#) to deal with the selection problem. I exploit idiosyncratic cohort variation in female students’ proportion within high-schools across cohorts, controlling for school-fixed effects and school-specific linear trends. Many studies that examine gender peer effects on major choice rely on a similar identification strategy ([SCHNEEWEIS; ZWEIMÜLLER, 2012](#); [GOULAS; MEGALOKONOMOU; ZHANG, 2018](#); [BRENØE; ZÖLITZ, 2020](#); [SCHØNE; SIMSON; STRØM, 2020](#)). Following [Lavy e Schlosser \(2011\)](#), I also focus

on the gender composition of the school's entire grade cohort (i.e., all classrooms) when estimating female peers' impacts, as students could potentially self-select into classrooms.

The hypothesis behind this strategy is that while parents and students select into schools, they cannot correctly predict the cohort-to-cohort variations in the students' gender composition, and those changes follow random processes. The inclusion of school fixed-effects controls for time-constant unobserved school characteristics that could relate both to the gender composition and female students' major choices. I also add a linear time-trend for each school to control for time-varying circumstances that could confound the estimated peer impacts.

To investigate gender peer effects on students' major choices, I estimate the following econometric model, where i indexes student, s schools, c cohort and t UNICAMP admission exam year:

$$y_{isct} = \gamma_0 + \beta_1 female_i + \beta_2 Share_HSFem_{isc} + \beta_3 female_i \times Share_HSFem_{isc} + \alpha_s + \alpha_t + X_i\omega + Z_{sc}\delta + \gamma_{st} + \epsilon_{isct} \quad (3.4.1)$$

y_{isct} are the dependent variables. Variable $female_i$ is a gender indicator, and $Share_HSFem_{isc}$ is the share of female peers in school s in the high-school graduation year c , excluding the student i (leave-me out proportion of female students in the third year of high-school). α_s are school fixed-effects, and α_t are year fixed-effects (UNICAMP admission exam year).^{3.4.1} X_i is a vector of student-level characteristics (normalized ENEM scores, age,^{3.4.2} and an indicator variable for work during the high-school). Z_{sc} represent school-cohort variables: school-grade size,^{3.4.3} the share of students that attend daytime classes, peers' average age. γ_{st} is a school-specific linear time-trend. In all the estimates, I cluster the standard errors at the school level.

^{3.4.1}The admission exam year is strongly correlated to the students' high-school graduation year (77% correlation). I control for year fixed-effects in the main specification, but results remain similar if I replace them with cohort fixed-effects. Results are available upon request.

^{3.4.2}I control for both the student's age at the admission exam year and its square.

^{3.4.3}I include a third-order polynomial of the school-grade size (total enrollments in the third-year of high-school).

The key coefficient of interest is β_3 . β_3 captures the differential impact for women of having a higher share of female peers. β_2 represents gender peer effects that hold both for men and women. β_1 controls for gender differences in the outcomes variables.

As previously described, this paper aims to examine if exposure to higher shares of female peers during high-school affects major choices. The first set of outcome variables (y_{isct}) measures the math-intensity of the field of study: a dummy variable for Math or Physics as priority subjects, and a STEM field indicator variable. Then, I assess female peer effects on the gender composition of the chosen major: dummy variables for female-dominated, male-dominated, and gender-balanced fields, and the average share of female applicants in the career. Finally, I investigate the impact on competitive majors' choice, measured by the average number of applicants per slot and final admission cutoff.

As a robustness check, I follow [Brenøe e Zölitz \(2020\)](#) and include school-cohort fixed-effects in the regressions. The idea is to control for unobservables at the level in which selection potentially occurs. However, in this specification, I can only investigate the impact of high-school female peers on the gender gaps in major choices. I use the equation (3.4.2) below in this analysis. y_{isct} , $female_i$, $Share_HSFem_{isc}$, X_i and γ_{sc} have the same definition as in equation (3.4.1). α_{sc} are school-cohort dummies. The coefficient δ_2 represents the impact of higher shares of female schoolmates on the gender difference in outcomes.

$$y_{isct} = \delta_0 + \delta_1 female_i + \delta_2 female_i \times Share_HSFem_{isc} + \alpha_{sc} + X_i \omega + \gamma_{sc} + \epsilon_{isct} \quad (3.4.2)$$

3.4.1 Evidence on identification strategy

To estimate gender peer effects, it is necessary to have sufficient variation in the within-school proportion of female students. Although in Brazil we have a gender-balanced coeducational system (53% of the high-school students are women in our sample of schools), we still observe for all cohorts that there is substantial variation in the percentage of female students across schools (Figure

3.A.1). However, as the empirical strategy uses the within-school variation in the gender composition, the essential is to have sufficient variation after controlling for school fixed-effects. Figure 3.4.1 reports the histograms of the overall and within-school proportion of female students. Although the overall variation in the women's percentage is wider, there is substantial within-school variation.

Figure 3.4.1 – Histogram of the proportion of female students in high-school (Year 3)

Note: The first histogram presents the distribution of the overall share of female students, while the second histogram presents the distribution of the proportion of female students within schools. I report the percentages using Brazilian School Census data for all schools from our sample. Within-school values: $p_{sc} - \bar{p}_s + \bar{p}$, where p_{sc} is the share of female students in the third year of high-school for school s and cohort c , $\bar{p}_s$ is the school's average share, and $\bar{p}$ is the overall mean.

Following [Bifulco, Fletcher e Ross \(2011\)](#), in Table 3.4.1, I examine the extent of variation in gender composition left after removing school-fixed effects and controlling for other covariates. Removing school-fixed effects reduces total variation by 41%. The inclusion of year fixed-effects, students' and schools' characteristics

only slightly alters the remaining variation. When I add school-specific time-trends, the variation is reduced by 48%. Comparing with the results found in the previous literature, it seems that there is enough within-school variation in our sample to estimate the gender peer effects.

Table 3.4.1 – Variation in the gender composition after removing school fixed-effects and linear trends

	Mean	Standard deviation	Observations
Share of female students			
Raw variable	0.523	0.104	139896
Residuals after removing school fixed-effects	0.000	0.061	139896
Residuals after removing school and time fixed-effects	0.000	0.061	139896
Residuals school FE + time FE + students' covariates	-0.000	0.061	139118
Residuals school FE + time FE + students' and schools' covariates	-0.000	0.061	138372
Resid. school and time FE + students' and schools' cov. + school linear trend	0.000	0.054	138372

Notes: In this table, I present the variation in the share of female students left after removing school fixed effects and controlling for other explanatory variables, following Bifulco, Fletcher e Ross (2011). Time-fixed effects are admission year fixed effects. Students' covariates: gender dummy, age, normalized ENEM scores, a dummy for work during high-school. School-cohort covariates: peers' age, the share of students in daytime classes, cohort size (third-order polynomial of the number of enrollments in the third year of high-school).

To further show the within-school variation in the percentage of female students, in Figure 3.4.2, I plot the density for changes in female students' proportion between consecutive years within schools. We note that the distribution of changes in female students' shares is similar across cohorts. Female percentage variation between consecutive years follows a normal distribution, and most of the changes are below 20 percentage points (from -20 to +20). Most importantly, I observe substantial within-school variation in students' gender composition from one year to the next.

The other hypothesis that needs to hold for the causal estimation of gender peer effects is that the within-school variation in female students' share is as good as random. To test this assumption's validity, I run balancing tests to check if within-school changes in female shares relate to students' covariates. In Table 3.4.2, I report the balancing tests. The dependent variables are the students' characteristics, and the key explanatory variable is the share of female peers in the third year of high-school. In column (1), I control only for school and year fixed-effects. Column

(2) adds peers' covariates and school-grade size (polynomial). In column (3), I include school and time fixed-effects, plus a school-specific linear trend. Column (4) includes peers' and schools' covariates to the model presented in column (3). In general, we observe no correlation between students' background characteristics and the share of female schoolmates.^{3.4.4}

Figure 3.4.2 – Density for changes in the proportion of females between consecutive years

Note: Change in female composition between consecutive years, within schools, in percentage points.

In Figure 3.4.3, I plot the distribution of within-school variation in the share

^{3.4.4}The construction of the female peers' share (leave-me out mean) leads to a negative mechanical correlation between the gender dummy and the share of female peers. [Guryan, Kroft e Notowidigdo \(2009\)](#) proposes a correction to the randomization check test. I have run the balancing tests using this correction, but the balancing test remains similar for the female dummy. As in Table 3.4.2, the female coefficient is small and statistically insignificant when controlling for school and time fixed-effects. However, the coefficient increases with the inclusion of school linear trends, and I observe a negative correlation that is statistically significant at a 1% level. The tests of the validity of the empirical strategy suggests that my preferred specification do not include school-specific trends.

Table 3.4.2 – Balancing test - Correlation between the share of female peers and students' characteristics

	(1)	(2)	(3)	(4)
<i>Dependent variables:</i>				
Female	-0.024 (0.039)	-0.034 (0.038)	-0.160*** (0.042)	-0.165*** (0.042)
Age	0.662** (0.308)	1.151*** (0.218)	0.900** (0.364)	1.391*** (0.271)
Household Size	-0.016 (0.050)	-0.010 (0.051)	0.026 (0.057)	0.027 (0.058)
Grade retention	-0.006 (0.012)	-0.001 (0.010)	-0.006 (0.013)	-0.003 (0.012)
Admission exam fee exemption	0.008 (0.010)	0.008 (0.010)	0.007 (0.009)	0.007 (0.009)
Family income - Less than 5 minimum wages	0.011 (0.017)	0.008 (0.017)	-0.006 (0.018)	-0.001 (0.018)
Family income - 5 to 10 minimum wages	0.015 (0.020)	0.023 (0.020)	0.012 (0.023)	0.017 (0.023)
Family income - 10 to 15 minimum wages	-0.010 (0.018)	-0.009 (0.018)	-0.002 (0.020)	-0.005 (0.020)
Family income - 15 to 20 minimum wages	-0.047*** (0.017)	-0.043** (0.017)	-0.035* (0.019)	-0.034* (0.020)
Family income - 20 to 30 minimum wages	0.008 (0.018)	0.007 (0.018)	0.024 (0.020)	0.017 (0.020)
Family income - More than 30 minimum wages	0.023 (0.019)	0.015 (0.018)	0.007 (0.019)	0.006 (0.019)
Only public high school	-0.012* (0.006)	-0.008 (0.006)	-0.013* (0.007)	-0.009 (0.007)
Only private elementary school	0.008 (0.024)	0.003 (0.024)	0.033 (0.027)	0.026 (0.026)
Only public elementary school	-0.028 (0.025)	-0.019 (0.025)	-0.043* (0.026)	-0.033 (0.025)
Mother - Higher Education	-0.019 (0.026)	-0.032 (0.026)	0.001 (0.030)	-0.013 (0.030)
Mother - Upper Secondary School	0.005 (0.027)	-0.011 (0.026)	-0.022 (0.030)	-0.033 (0.030)
Mother - Lower Secondary School	0.014 (0.016)	0.023 (0.016)	0.020 (0.018)	0.028 (0.018)
Mother - Primary School	-0.001 (0.017)	0.012 (0.015)	-0.004 (0.018)	0.006 (0.017)
Mother - No Primary School	0.001 (0.010)	0.008 (0.009)	0.005 (0.011)	0.012 (0.010)
Father - Higher Education	0.035 (0.025)	0.021 (0.024)	0.042 (0.028)	0.026 (0.028)
Father - Upper Secondary School	-0.073*** (0.024)	-0.077*** (0.024)	-0.080*** (0.028)	-0.083*** (0.028)
Father - Lower Secondary School	0.007 (0.014)	0.008 (0.014)	0.005 (0.016)	0.007 (0.016)
Father - Primary School	0.022 (0.016)	0.032** (0.015)	0.024 (0.018)	0.034** (0.018)
Father - No Primary School	0.010 (0.010)	0.016* (0.010)	0.009 (0.012)	0.015 (0.012)
Mother - High-Earning Occupation	-0.000 (0.004)	0.000 (0.004)	0.002 (0.004)	0.002 (0.004)
Mother - White-Collar Worker	0.005 (0.024)	-0.004 (0.024)	0.004 (0.028)	-0.005 (0.028)
Mother - Grey-Collar Worker	-0.007 (0.017)	-0.013 (0.017)	-0.015 (0.018)	-0.021 (0.019)
Mother - Blue-Collar Worker	-0.016 (0.012)	-0.010 (0.011)	-0.028* (0.014)	-0.022 (0.014)
Mother - House Wife	0.019 (0.025)	0.027 (0.026)	0.037 (0.028)	0.044 (0.029)
Mother - Other occupations	-0.001 (0.005)	-0.000 (0.005)	-0.000 (0.006)	0.001 (0.006)
Father - High-Earning Occupation	-0.003 (0.006)	-0.005 (0.006)	-0.004 (0.007)	-0.004 (0.007)
Father - White-Collar Worker	0.028 (0.023)	0.019 (0.023)	0.044 (0.028)	0.033 (0.026)
Father - Grey-Collar Worker	-0.024 (0.016)	-0.023 (0.016)	-0.032* (0.018)	-0.031* (0.018)
Father - Blue-Collar Worker	-0.006 (0.018)	0.004 (0.016)	-0.010 (0.021)	-0.002 (0.018)
Father - Other occupations	0.001 (0.006)	0.000 (0.006)	-0.002 (0.006)	-0.001 (0.006)

Key explanatory variable: Percentage of female peers in Year 3 of High-School. Standard-errors clustered at the school-level. Column (1) controls for school and time fixed-effects, column (2) includes peers' and schools' covariates as control variables. In column (3), I include both school and year fixed-effects, plus a school-specific linear-trend. Column (4) adds peers' and schools' covariates to the column (3) model.

of female students. If the empirical strategy is valid, we would expect that the histogram would resemble a natural random process. To calculate the within-school variation, I have regressed the female share on school and cohort fixed-effects and plotted the residuals, obtaining the within-school across cohorts changes. For comparison purposes, I also included a graph of a normal distribution. Graphically, the variation in the percentage of female students within school seems like random fluctuations, and the distribution closely follows the gaussian curve. Figure 3.A.2 presents the same graph, residualized for school and cohort fixed-effects, plus a school-specific trend. When I include the linear trends, the variation is highly concentrated around zero and visually does not look like a normal distribution. Therefore, the preferred specification controls only for year and school-fixed effects.

Figure 3.4.3 – Year-to-year variation in the proportion of female high school students

Note: This graph shows year-to-year variation in the percentage of female students in the third year of high-school school peers, plotted along with a normal distribution. To calculate the within-school variation, I have regressed female share on school and cohort fixed-effects and plotted the residuals. Each high school-cohort represents one observation.

Finally, I test whether there is school autocorrelation in the percentage of female students. Following the strategy proposed by [Brenøe e Zölitz \(2020\)](#), I run one regression for each school from our sample (3,573 schools)^{3.4.5} to test if the share of female enrollments for a cohort (t) predicts the proportion of female students in the next cohort ($t+1$). According to [Brenøe e Zölitz \(2020\)](#), although the serial correlation would not directly jeopardize identification, it could suggest the existence of confounding time-varying factors. Table 3.4.3 reports the percentage of schools for which the female share in a cohort predicts the women's share in the following year. Columns (1) and (2) use school data from 1998 to 2007, all the years for which I collected information. Columns (3) and (4) restrict the analysis to the Census years used in the main regressions (1998 to 2003). In columns (1) and (3), I regress the share in $t+1$ on the share in t , and in columns (2) and (4), I include linear trends. In all columns and significance levels, the pattern suggests no serial correlation in female students' percentage. At the 1% significance level, we note that the proportion of school coefficients that are statistically significant varies between 0.9% and 0.5%. Besides, in all columns, less than 5% of the school coefficients are statistically significant at a 5% level (maximum 4.6% and minimum 2.6%). Finally, the proportion of schools for which the coefficients are significant at a 10% level ranges from 8.7% to 6.9%. Therefore, there is no evidence of school autocorrelation in gender composition.

3.5 Results

Table 3.5.1 reports the impacts of the proportion of female peers in the third year of high-school on the probability of choosing a math-intensive field. Each panel presents estimates for a different dependent variable. The first panel shows the impacts on the probability of having Mathematics or Physics as priority subjects, and the second panel presents the impacts on the likelihood of applying for a STEM major. In column (1), I control for school and year fixed-effects, and in the next columns, I gradually include explanatory variables. Column (2) adds students'

^{3.4.5}In the sample, I have 3,643 schools, and for 3,573 (98%) I have Census information for at least two different cohorts.

Table 3.4.3 – School-level autocorrelation in the percentage of female students

	Census Years 1998-2007		Census Years 1998-2003	
	(1)	(2)	(3)	(4)
<i>Proportion of School Coefficients</i>				
Significant at a 1% level	0.008	0.009	0.007	0.005
Significant at a 5% level	0.046	0.043	0.033	0.026
Significant at a 10% level	0.087	0.083	0.074	0.069
# Schools	3573	3573	3430	3430
School-specific linear trend	No	Yes	No	Yes

Note: Columns (1) and (2) use school data from 1998 to 2007. Columns (3) and (4) restricts the analysis to the Census years that I use in the main regressions (1998 to 2003). In columns (1) and (3), I just regress the share in $t+1$ on the share in t , and in columns (2) and (4) I include linear trends. I run one regression per school.

covariates, column (3) controls for the size of high-school grade, and in column (4) I add peers' characteristics. In column (5), I control for school and time dummies plus a school-specific linear trend. Lastly, column (6) includes students' covariates in the model presented in column (5). In all regressions, I cluster standard-errors at the school level. In general, the results are not sensitive to covariates' inclusion, and estimates are similar across columns. The preferred specification is column (4), where I include all covariates, but do not add the linear trend, because the school trend reduces the variation in schoolmates' gender composition. Thus, when interpreting the point estimates, I will focus on column (4).

To sum up, Table 3.5.1 suggests that the proportion of female peers reduces the candidates' probability of applying for a math or science intensive field, although those impacts have modest sizes. In general, there are no gender differences in the estimated effects. Women are less prone to choose majors with Mathematics or Physics as priority disciplines (around 34 p.p. lower likelihood). Female peers' proportion reduces the probability of selecting a math-intensive area both for men and women, and there are no gender differences in the estimated impacts. If we increase by ten percentage points the female peers' share, this likelihood decreases by 1.4 percentage points. The effects represent less than a 2.5% reduction for men

and a 5.6% decrease for women. The last panel shows gender peer effects on STEM major choice. A lower percentage of female candidates apply for STEM majors (28 p.p. lower). The proportion of female schoolmates reduces the probability that the student applies for a STEM field at UNICAMP. There is no differential impact by the candidates' gender. If there is a one-standard-deviation higher share of female peers, students are 1.4 percentage points less likely to choose a science career. This impact represents a 2.4% reduction for male applicants and a 4.3% reduction for females. Interestingly, when I split the sample into competitive (highest 25% admission cutoffs) and non-competitive majors (the remaining fields of study), I observe a different pattern by gender. For female applicants, a ten percentage points larger share of female schoolmates decreases by one percentage point the likelihood of selecting a non-competitive STEM major, while the probability of choosing a competitive STEM major slightly increases (0.3 p.p.). In contrast, for men, the same increment in the share of female peers leads to a 2.4 percentage points lower likelihood of selecting a competitive STEM major and a one percentage point larger probability of choosing a non-competitive STEM field of study.^{3.5.1}

Next, I look at how gender peer impacts influence the choice of female or male-dominated majors. The outcomes of interest are binary variables that indicate if a field is female-dominated, male-dominated, or gender-balanced. Additionally, I assess the impacts on the average share of female applicants in the chosen career. Table 3.5.2 summarizes the impacts for these response variables. Columns (1) to (6) are analogous to the columns from the previous table. I observe that the estimates almost do not change when I gradually include covariates. This table's interpretation also focuses on column (4), the preferred model.

As we would expect, female candidates are more likely to apply for female-dominated fields. They have a 20 percentage points higher probability than men. There are no gender peer effects on the likelihood of choosing a female-dominated area. Female applicants have, on average, a much lower probability of selecting male-dominated fields. The gender difference is 47 percentage points, *ceteris paribus*. For men, the share of female peers decreases the likelihood of choosing a male-dominated

^{3.5.1}In future versions of this working paper, I intend to expand the analysis of competitive and non-competitive STEM majors.

Table 3.5.1 – Impact of high-school female peers on the likelihood of choosing a math-intensive field

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Math or Physics are priority subjects</i>						
Female $\times$ % female peers	0.034 (0.053)	0.064 (0.052)	0.061 (0.054)	0.062 (0.054)	0.032 (0.056)	0.062 (0.055)
% female peers - Year 3 High school	-0.155*** (0.038)	-0.144*** (0.037)	-0.136*** (0.039)	-0.141*** (0.039)	-0.149*** (0.044)	-0.129*** (0.041)
Female	-0.334*** (0.029)	-0.345*** (0.029)	-0.343*** (0.030)	-0.344*** (0.030)	-0.332*** (0.031)	-0.343*** (0.031)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.425	0.424	0.424	0.424	0.425	0.424
<i>Dependent variable: STEM Major</i>						
Female $\times$ % female peers	0.032 (0.047)	0.059 (0.047)	0.056 (0.048)	0.057 (0.048)	0.029 (0.050)	0.056 (0.050)
% female peers - Year 3 High school	-0.150*** (0.038)	-0.139*** (0.037)	-0.130*** (0.038)	-0.138*** (0.038)	-0.143*** (0.046)	-0.122*** (0.043)
Female	-0.270*** (0.026)	-0.279*** (0.026)	-0.277*** (0.026)	-0.278*** (0.027)	-0.268*** (0.028)	-0.276*** (0.028)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.465	0.465	0.465	0.465	0.465	0.465
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

field; for women, the opposite is true. If there is an increase by ten percentage points in female schoolmates' proportion, the male probability of choosing a male-dominated field reduces by two percentage points, while female likelihood increases by 0.4 percentage points.

Interestingly, women apply for gender-balanced fields at a higher rate than men: female probability is 26 p.p. larger. The percentage of female peers in high-school raises the male likelihood of selecting gender-balanced careers and reduces female probability. A one-standard-deviation larger share of female peers increases by 1.6 p.p. the likelihood for men and decreases by 0.1 p.p. for women.

When exposed to larger shares of female peers during the high-school, female applicants tend to choose a major with a lower average female candidates percentage. Simultaneously, for male applicants, the share of female peers increases the proportion of female contestants of the chosen undergraduate program. A 10 percentage points higher proportion of female schoolmates increases by one percentage point the average career share of female candidates for men and reduces by 0.1 p.p. for women.

In conclusion, female peers' share has opposite impacts on career selection for female and male applicants. For women, there is a 5% increase in the probability of selecting a male-dominated field, a 0.4% decrease in gender-balanced career choice, and the average share of female applicants in the major choice reduces by 0.2%. Male applicants become less likely to choose male-dominated fields (4% decrease) and more likely to select gender-balanced areas (3%), resulting in major choices with 3% higher female shares.

In Table 3.5.3, I present the gender peer impacts on the degree of competitiveness of the applicant's major choice. I use two variables to measure how competitive a major is: average number of applicants per slot and final admission scores cutoff. The columns follow the same pattern as in Tables 3.5.1 and 3.5.2. It is interesting to observe that female applicants choose undergraduate programs with a higher number of applicants per slot. However, their chosen fields of study have lower admission cutoffs, i.e., the major-specific minimum score needed to UNICAMP's admission.

Table 3.5.2 – Impact of high-school female peers on the probability of selecting a male-dominated major

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Female-dominated field</i>						
Female × % female peers	-0.034 (0.059)	-0.042 (0.060)	-0.041 (0.060)	-0.043 (0.061)	-0.035 (0.062)	-0.041 (0.063)
% female peers - Year 3 High school	0.006 (0.033)	0.021 (0.033)	0.020 (0.034)	0.018 (0.034)	-0.004 (0.035)	0.013 (0.036)
Female	0.215*** (0.034)	0.202*** (0.034)	0.201*** (0.034)	0.202*** (0.034)	0.215*** (0.035)	0.201*** (0.035)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.164	0.164	0.164	0.164	0.164	0.164
<i>Dependent variable: Male-dominated field</i>						
Female × % female peers	0.190*** (0.067)	0.215*** (0.067)	0.213*** (0.068)	0.212*** (0.068)	0.191*** (0.070)	0.216*** (0.071)
% female peers - Year 3 High school	-0.189*** (0.047)	-0.180*** (0.045)	-0.175*** (0.046)	-0.176*** (0.046)	-0.183*** (0.051)	-0.167*** (0.048)
Female	-0.459*** (0.038)	-0.467*** (0.038)	-0.466*** (0.039)	-0.465*** (0.039)	-0.459*** (0.040)	-0.467*** (0.040)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.259	0.259	0.259	0.259	0.259	0.259
<i>Dependent variable: Gender-balanced field</i>						
Female × % female peers	-0.156*** (0.041)	-0.174*** (0.041)	-0.172*** (0.040)	-0.169*** (0.041)	-0.156*** (0.042)	-0.174*** (0.041)
% female peers - Year 3 High school	0.183*** (0.037)	0.159*** (0.036)	0.155*** (0.035)	0.158*** (0.036)	0.187*** (0.040)	0.154*** (0.038)
Female	0.244*** (0.022)	0.266*** (0.022)	0.265*** (0.022)	0.263*** (0.022)	0.244*** (0.023)	0.266*** (0.023)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.577	0.577	0.577	0.578	0.577	0.577
<i>Dependent variable: Average % of female applicants in career (2000-2008)</i>						
Female × % female peers	-0.101** (0.045)	-0.114** (0.045)	-0.113** (0.045)	-0.113** (0.045)	-0.102** (0.047)	-0.115** (0.047)
% female peers - Year 3 High school	0.099*** (0.027)	0.097*** (0.026)	0.094*** (0.027)	0.094*** (0.027)	0.095*** (0.030)	0.089*** (0.028)
Female	0.284*** (0.026)	0.285*** (0.026)	0.284*** (0.026)	0.284*** (0.026)	0.285*** (0.027)	0.285*** (0.027)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.499	0.499	0.499	0.499	0.499	0.499
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

The first panel reports the impacts on the major's mean number of applicants per slot. The coefficient of the share of female peers is positive and marginally statistically significant in most specifications. The evidence suggests that female schoolmates increase the competitiveness measured by the number of candidates per available place for both genders. Still, results are not robust across the models, and the point estimates are volatile. The second panel shows the effects of high-school peers' gender composition on the choice of competitive majors, measured by admission cutoff. I observe lower cutoff scores for women: in column (4), I find that female candidates apply, on average, to majors that have a 10.7 lower admission cutoff. For comparison purposes, the average admission score cutoff in our sample is 526 points. The percentage of female high-school peers increases the competitiveness of women's career choices, while there are no gender peer impacts for men. An increase by ten percentage points will raise by 2 points female admission cutoff. The increase represents a 0.3% higher women's average cutoff. The inclusion of students' characteristics in the regressions increases the interaction term coefficients.^{3.5.2}

3.5.1 Robustness checks

This section demonstrates the robustness of the main findings to different specifications and sample choices. First, as previously mentioned in Section 3.4, I estimate the regressions using equation (3.4.2), controlling for school-cohort fixed-effects. This specification enables me to control for potential selection issues occurring over the cohorts within the school that could impact the estimates. However, in this model, I can only analyze female peers' impact on gender gaps in major choices.^{3.5.3} Tables 3.5.4, 3.5.5 and 3.5.6 report these results. Estimates corroborate the results found in the main regressions.

^{3.5.2}I have also run an additional regression replacing the average admission cutoff with a competitive dummy. The dummy equals one if the major has the top 25% average cutoff and zero otherwise. In the preferred specification (column (4)), I find that if the female percentage of high-school peers increases by ten percentage points, the probability of selecting a competitive major increases by one percentage point for female applicants and decreases 1.5 p.p. for males.

^{3.5.3}Equation (3.4.1) allows us to estimate gender peer effects on the major choice directly. For instance, we can examine the impact of the share of female peers on the likelihood of selecting a math-intensive field, as well as the differential impact by gender.

Table 3.5.3 – Impact of high-school female peers on the choice of competitive majors

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Average # applicants per slot (2000-2008)</i>						
Female × % female peers	0.977 (1.963)	1.182 (1.912)	1.347 (1.920)	1.409 (1.928)	1.102 (2.029)	1.240 (1.964)
% female peers - Year 3 High school	5.210*** (1.701)	3.317** (1.606)	2.859* (1.625)	2.997* (1.650)	4.715** (1.922)	2.565 (1.779)
Female	6.034*** (1.059)	6.809*** (1.041)	6.724*** (1.046)	6.686*** (1.051)	5.936*** (1.096)	6.767*** (1.070)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	33.447	33.467	33.467	33.473	33.447	33.467
<i>Dependent variable: Average final admission scores cutoff (2000-2008)</i>						
Female × % female peers	13.911** (5.910)	17.609*** (6.555)	18.095*** (6.523)	18.262*** (6.546)	13.896** (6.051)	17.283** (6.744)
% female peers - Year 3 High school	-0.147 (4.862)	-6.483 (4.947)	-7.781 (5.013)	-7.494 (5.084)	-0.570 (5.226)	-7.070 (5.236)
Female	-12.183*** (3.286)	-10.334*** (3.655)	-10.585*** (3.638)	-10.683*** (3.650)	-12.254*** (3.364)	-10.156*** (3.760)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	525.665	525.739	525.739	525.755	525.665	525.739
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.5.4 presents peer impacts on the majors' math-intensity. In all columns (1 to 4), I include school-cohort fixed effects, in columns (2) and (4) I add the students' covariates, and columns (3) and (4) incorporate school-specific time trends. Women are less likely to apply to a STEM field or majors with Mathematics or Physics as priority subjects. The gender composition of high-school peers does not influence gender differences in the likelihood of selecting a STEM major or a career that has Math or Physics as a priority discipline in Stage 2.

Next, Table 3.5.5 presents the differential impacts on the average percentage of women in the chosen field of study. Female applicants are more likely to apply for female-dominated and gender-balanced majors, while males are more prone to select male-dominated fields. The proportion of female schoolmates reduces gender

gaps. A higher female peers' share increases both the female relative likelihood of applying for a male-dominated and the male relative probability of selecting a gender-balanced major. It also decreases the gender differences in the average share of females in the major choice.

Table 3.5.6 reports impacts on the choice of competitive majors. As I have previously noted, female candidates select majors with a higher number of applicants per available slot, but with lower final admission cutoffs. The percentage of female peers in high-school reduces the gender gap in the career's degree of competitiveness. If women representation rises by ten percentage points, the gender difference in the admission cutoff falls by two percentage points.

Table 3.5.4 – School-Cohort FE: Impact of high-school female peers on the probability of choosing a math-intensive field

	(1)	(2)	(3)	(4)
<i>Dependent variable: Math or Physics are priority subjects</i>				
Female × % female peers	0.061 (0.059)	0.073 (0.059)	0.061 (0.059)	0.073 (0.059)
Female	-0.342*** (0.032)	-0.343*** (0.032)	-0.342*** (0.033)	-0.343*** (0.032)
Observations	136569	135786	136569	135786
Mean dependent variable	0.424	0.424	0.424	0.424
<i>Dependent variable: STEM Major</i>				
Female × % female peers	0.052 (0.051)	0.063 (0.051)	0.052 (0.052)	0.063 (0.052)
Female	-0.275*** (0.028)	-0.276*** (0.028)	-0.275*** (0.028)	-0.276*** (0.028)
Observations	136569	135786	136569	135786
Mean dependent variable	0.465	0.465	0.465	0.465
School-Cohort Fixed-Effects	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	No	Yes
School linear trend	No	No	Yes	Yes

Notes: Key explanatory variable: percentage of female peers (leave-me out share) in the third year of high-school interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * p<0.10, ** p<0.05, *** p<0.01.

I also test if the estimated impacts are robust to the exclusion of specific majors. I run 59 regressions (one for each major), each estimation excluding

Table 3.5.5 – School-Cohort FE: Impact of high-school female peers on the likelihood of selecting a male-dominated major

	(1)	(2)	(3)	(4)
<i>Dependent variable: Female-dominated field</i>				
Female × % female peers	-0.048 (0.068)	-0.059 (0.069)	-0.048 (0.068)	-0.059 (0.069)
Female	0.222*** (0.038)	0.211*** (0.038)	0.222*** (0.038)	0.211*** (0.039)
Observations	136569	135786	136569	135786
Mean dependent variable	0.164	0.164	0.164	0.164
<i>Dependent variable: Male-dominated field</i>				
Female × % female peers	0.221*** (0.075)	0.232*** (0.076)	0.221*** (0.076)	0.232*** (0.076)
Female	-0.471*** (0.042)	-0.472*** (0.043)	-0.471*** (0.043)	-0.472*** (0.043)
Observations	136569	135786	136569	135786
Mean dependent variable	0.258	0.258	0.258	0.258
<i>Dependent variable: Gender-balanced field</i>				
Female × % female peers	-0.173*** (0.044)	-0.173*** (0.044)	-0.173*** (0.045)	-0.173*** (0.044)
Female	0.249*** (0.024)	0.261*** (0.024)	0.249*** (0.024)	0.261*** (0.024)
Observations	136569	135786	136569	135786
Mean dependent variable	0.578	0.579	0.578	0.579
<i>Dependent variable: Average % of female applicants in career (2000-2008)</i>				
Female × % female peers	-0.120** (0.051)	-0.127** (0.051)	-0.120** (0.051)	-0.127** (0.051)
Female	0.292*** (0.029)	0.290*** (0.029)	0.292*** (0.029)	0.290*** (0.029)
Observations	136569	135786	136569	135786
Mean dependent variable	0.500	0.500	0.500	0.500
School-Cohort Fixed-Effects	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	No	Yes
School linear trend	No	No	Yes	Yes

Notes: Key explanatory variable: percentage of female peers (leave-me out share) in the third year of high-school interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.5.6 – School-Cohort FE: Impact of high-school female peers on the choice of competitive majors

	(1)	(2)	(3)	(4)
<i>Dependent variable: Average # applicants per slot (2000-2008)</i>				
Female × % female peers	0.574 (2.059)	1.278 (2.012)	0.574 (2.084)	1.278 (2.036)
Female	6.032*** (1.099)	6.518*** (1.078)	6.032*** (1.113)	6.518*** (1.091)
Observations	136569	135786	136569	135786
Mean dependent variable	33.534	33.554	33.534	33.554
<i>Dependent variable: Average final admission scores cutoff (2000-2008)</i>				
Female × % female peers	16.025** (6.272)	19.829*** (7.040)	16.025** (6.348)	19.829*** (7.126)
Female	-13.770*** (3.447)	-12.029*** (3.877)	-13.770*** (3.489)	-12.029*** (3.925)
Observations	136569	135786	136569	135786
Mean dependent variable	525.983	526.055	525.983	526.055
School-Cohort Fixed-Effects	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	No	Yes
School linear trend	No	No	Yes	Yes

Notes: Key explanatory variable: percentage of female peers (leave-me out share) in the third year of high-school interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * p<0.10, ** p<0.05, *** p<0.01.

one field of study – all its applicants. I plot the coefficients obtained in these regressions, together with horizontal lines that indicate the point estimate for the full sample (solid red line) and its confidence intervals (dashed black lines). Also, to turn visualization easier, I highlight the ‘zero impact’ by including an additional horizontal line (dashed orange line) in the graphs where the coefficients equal zero. All coefficients that appear above the orange line represent positive point estimates, and below the line are negative coefficients. Figures 3.A.3 to 3.A.10 present the results for this robustness check. For all response variables, the gender peer effects are robust to the exclusion of any major. The point estimates lie within the confidence interval from the estimation using the full sample, and the coefficients are close to the impacts obtained in our main regressions.

Next, I have tested if the impacts on the likelihood of selecting female-

dominated, male-dominated, or gender-balanced majors remain similar if I use alternative measures of the gender composition of the field of study. Instead of using UNICAMP admission data, I construct analogous dependent variables using the Brazilian Higher Education Census (INEP/MEC, 2000–2004). The Brazilian Higher Education Census (*Censo da Educação Superior*) contains information for all higher education institutions that offer undergraduate programs in Brazil. I use data from 2000 to 2004 and calculate females' share among all undergraduate students enrolled in the Brazilian institutions by major choice.^{3.5.4} I report the results in Table 3.A.6. Female applicants are 25 percentage points more likely to choose a female-dominated major, 46 p.p. less likely to select male-dominated fields, and 21 p.p. more prone to select a gender-balanced field of study. We find results similar to Table 3.5.2. If female peers' share in high school increases, male applicants are less likely to select male-dominated majors and more likely to choose gender-balanced fields. For females, the effect sizes are much smaller, and the impacts are the opposite: an increase in male-dominated and a decrease in gender-balanced major choices.

As a further robustness check, I drop schools that experienced a change higher than 50% in the gender composition of the third year of high school, compared to the school's average share of female students.^{3.5.5} Results are qualitatively the same and quantitative similar once I drop these applicants. Another restriction was to exclude outlier schools. The main concern is that a few schools with a large (or small) proportion of female students drive the impacts. To deal with this possibility, I have excluded from my sample the school-cohorts at the top 1% and bottom 1% in the share of female students (1st and 100th percentiles). Results are robust to the sample restrictions.^{3.5.6}

As previously mentioned, some technical high-schools select their students by admission exams. The students take an exam before the beginning of high-school,

^{3.5.4}Both UNICAMP and Higher Education Census data do not use standard international classifications to define a major. Therefore, I manually matched the study fields according to their name.

^{3.5.5}School's average percentage of female students is calculated using School Censuses from 1998 to 2007, and the year 1996. For the year 1997, there is no information on the gender composition of students.

^{3.5.6}Results are available upon request.

and the best-ranked students are admitted into the school. The most prestigious technical schools are public (state or federal) and prepare well students for the university admission exams. As these schools may use different selection processes, I run an additional robustness check excluding applicants that attended technical schools (13% of the primary sample). In general, the results obtained after the exclusion are qualitatively the same. A higher share of female peers reduces the likelihood of choosing math-intensive or STEM majors. The proportion of same-gender peers increases the probability of selecting male-dominated fields while decreasing the probability of choosing gender-balanced majors. However, when I exclude technical schools, I find that larger percentages of female peers lead to more competitive major choices for both men and women, in contrast to the main results that only affect females. The female peers' share increases the average number of applicants per slot and admission cutoff of the major choice.

Lastly, I check if results change if I drop small school-cohorts that do not have more than 30 students in the third year of high-school. This restriction excluded around 9,000 applicants (6%).^{3.5.7} The idea behind it is that small changes in the number of female students in smaller schools could lead to a wider percentage change, influencing the estimated effects. The exclusion of smaller cohorts does not considerably alter the estimated impacts.

3.5.2 Non-linear effects

In the previous estimates, I have assumed that gender peer impacts are linear. In this subsection, I test alternative functional forms, allowing for the possibility of non-linear effects. First, I assess the impacts of having a female majority in the third year of high-school. [Lavy e Schlosser \(2011\)](#) find that gender peer effects are stronger for high percentages of female students, especially when women represent the majority in a cohort, and I will test if the same pattern holds in our setting. In this exercise, the independent variable is a dummy variable indicating if the applicant's peers are at least 50% women. Thus, I replace the

^{3.5.7}Results are available upon request.

proportion of female peers $Share_HSFem_{isc}$ in equation (3.4.1) with the binary variable $1(\%femalepeers \geq 50)_{isc}$.

Table 3.5.7 displays the impact on the likelihood of selecting a math-intensive field. Column (1) to (6) are analogous to the previous tables (see description of Table 3.5.1). As before, this subsection comments on the coefficients presented in column (4), the specification that controls for the school and year dummies, students' covariates, cohort size, and peers' characteristics. I verify that having a women's predominance during high-school reduces the probability of choosing a math-intensive field by 1.6 percentage points. There is no differential impact on females outcomes. The impact of female prevalence on the likelihood of selecting a STEM major is only marginally significant.

Table 3.5.8 shows impacts on the gender composition of the chosen major. For men, a female majority among their high-school peers leads to a two percentage point lower probability of choosing a male-dominated field and to a two percentage points higher likelihood of a gender-balanced major, resulting in major choices with 1.2 p.p. higher female shares. For the female applicants, women's predominance impacts are the opposite: they are 1.6 p.p. more likely to select a male-dominated field and one percentage point less prone to choose a gender-balanced field of study, reducing by around one percentage point the share of female applicants in their major choice.

Subsequently, I assess the effects on the selected major's competitiveness degree (Table 3.5.9). The female majority influences the choice of competitive majors only for women. For female candidates, having more female than male peers increases by 2 points the admission cutoff score of the chosen major. For men, a female majority in high-school reduces the admission cutoff by 1.6 points, but the coefficient is not statistically significant at a 5% level.

As an additional specification for non-linear impacts, I split the sample of school-cohorts into three different categories based on the female peers' percentage:

Table 3.5.7 – Impact of high-school female majority on the probability of choosing a math-intensive field

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Math or Physics are priority subjects</i>						
Female $\times$ % female peers ≥ 50	0.013 (0.010)	0.015 (0.010)	0.015 (0.010)	0.016 (0.010)	0.013 (0.010)	0.016 (0.010)
% female peers ≥ 50	-0.018** (0.008)	-0.016** (0.007)	-0.016** (0.007)	-0.016** (0.007)	-0.018** (0.008)	-0.016** (0.008)
Female	-0.325*** (0.009)	-0.322*** (0.009)	-0.322*** (0.009)	-0.322*** (0.009)	-0.324*** (0.009)	-0.321*** (0.009)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.425	0.424	0.424	0.424	0.425	0.424
<i>Dependent variable: STEM Major</i>						
Female $\times$ % female peers ≥ 50	0.007 (0.009)	0.009 (0.009)	0.009 (0.009)	0.009 (0.009)	0.007 (0.010)	0.010 (0.009)
% female peers ≥ 50	-0.014* (0.007)	-0.013* (0.007)	-0.012* (0.007)	-0.013* (0.007)	-0.013* (0.008)	-0.012 (0.008)
Female	-0.258*** (0.008)	-0.255*** (0.008)	-0.254*** (0.008)	-0.254*** (0.008)	-0.257*** (0.009)	-0.254*** (0.009)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.465	0.465	0.465	0.465	0.465	0.465
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: indicator variable for percentages of female peers higher than 50%, and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5.8 – Impact of high-school female majority on the likelihood of selecting a male-dominated major

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Female-dominated field</i>						
Female $\times$ % female peers ≥ 50	-0.006 (0.010)	-0.007 (0.010)	-0.007 (0.010)	-0.007 (0.010)	-0.005 (0.011)	-0.006 (0.011)
% female peers ≥ 50	0.000 (0.006)	0.002 (0.006)	0.002 (0.006)	0.002 (0.006)	0.001 (0.006)	0.003 (0.006)
Female	0.201*** (0.010)	0.184*** (0.010)	0.184*** (0.010)	0.185*** (0.010)	0.201*** (0.010)	0.184*** (0.010)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.164	0.164	0.164	0.164	0.164	0.164
<i>Dependent variable: Male-dominated field</i>						
Female $\times$ % female peers ≥ 50	0.035*** (0.012)	0.037*** (0.013)	0.037*** (0.013)	0.037*** (0.013)	0.035*** (0.013)	0.038*** (0.013)
% female peers ≥ 50	-0.023** (0.010)	-0.022** (0.009)	-0.021** (0.009)	-0.021** (0.009)	-0.026*** (0.010)	-0.024*** (0.009)
Female	-0.383*** (0.013)	-0.380*** (0.013)	-0.380*** (0.013)	-0.379*** (0.013)	-0.383*** (0.013)	-0.380*** (0.013)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.259	0.259	0.259	0.259	0.259	0.259
<i>Dependent variable: Gender-balanced field</i>						
Female $\times$ % female peers ≥ 50	-0.029*** (0.009)	-0.031*** (0.009)	-0.030*** (0.009)	-0.030*** (0.009)	-0.030*** (0.010)	-0.031*** (0.009)
% female peers ≥ 50	0.023*** (0.009)	0.020** (0.008)	0.019** (0.008)	0.019** (0.008)	0.025*** (0.009)	0.021*** (0.008)
Female	0.182*** (0.009)	0.196*** (0.009)	0.195*** (0.008)	0.195*** (0.009)	0.182*** (0.009)	0.196*** (0.009)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.577	0.577	0.577	0.578	0.577	0.577
<i>Dependent variable: Average % of female applicants in career (2000-2008)</i>						
Female $\times$ % female peers ≥ 50	-0.020** (0.008)	-0.021*** (0.008)	-0.021*** (0.008)	-0.021*** (0.008)	-0.020** (0.008)	-0.021** (0.008)
% female peers ≥ 50	0.013*** (0.005)	0.013*** (0.005)	0.012** (0.005)	0.012** (0.005)	0.014*** (0.005)	0.014*** (0.005)
Female	0.244*** (0.008)	0.239*** (0.008)	0.239*** (0.008)	0.239*** (0.008)	0.244*** (0.008)	0.239*** (0.008)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.499	0.499	0.499	0.499	0.499	0.499
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: indicator variable for percentages of female peers higher than 50%, and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5.9 – Impact of high-school female majority on the choice of competitive majors

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Average # applicants per slot (2000-2008)</i>						
Female $\times$ % female peers ≥ 50	0.204 (0.440)	0.318 (0.425)	0.335 (0.427)	0.358 (0.428)	0.170 (0.454)	0.273 (0.439)
% female peers ≥ 50	0.368 (0.373)	0.105 (0.352)	0.065 (0.351)	0.071 (0.357)	0.314 (0.413)	0.034 (0.392)
Female	6.405*** (0.369)	7.212*** (0.368)	7.202*** (0.369)	7.182*** (0.371)	6.389*** (0.381)	7.226*** (0.380)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	33.447	33.467	33.467	33.473	33.447	33.467
<i>Dependent variable: Average final admission scores cutoff (2000-2008)</i>						
Female $\times$ % female peers ≥ 50	2.590** (1.249)	3.354** (1.307)	3.398*** (1.306)	3.489*** (1.305)	2.561** (1.286)	3.275** (1.348)
% female peers ≥ 50	-0.648 (0.935)	-1.510 (0.929)	-1.612* (0.927)	-1.620* (0.933)	-0.888 (1.029)	-1.775* (1.023)
Female	-6.655*** (1.129)	-3.384*** (1.200)	-3.410*** (1.200)	-3.483*** (1.198)	-6.725*** (1.161)	-3.329*** (1.236)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	525.665	525.739	525.739	525.755	525.665	525.739
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: indicator variable for percentages of female peers higher than 50%, and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

(0 to 33%), ($> 33\%$ and $\leq 66\%$), and ($> 66\%$).^{3.5.8} I estimate the following regression:

$$\begin{aligned}
 y_{isct} = & \gamma_0 + \beta_1 female_i + \beta_2 1(\%femalepeers \leq 33)_{isc} + \beta_3 1(\%femalepeers > 66)_{isc} \\
 & + \beta_4 female_i \times 1(\%femalepeers \leq 33)_{isc} + \beta_5 female_i \times 1(\%femalepeers > 66)_{isc} \\
 & + \alpha_s + \alpha_t + X_i\omega + Z_{sc}\delta + \gamma_{st} + \epsilon_{isct}
 \end{aligned}
 \tag{3.5.1}$$

The definition of the variables are the same as in equation (3.4.1), but in equation (3.5.1) instead of linearly including the share of female peers, I use the categories of female percentage, also interacting them with the gender dummy. The baseline category is school-cohorts that have between 33% and 66% of women. Tables 3.5.10, 3.5.11, and 3.5.12 report the results for this exercise.

I verify that having a proportion of female high-school peers greater than 66% does not influence the major choice, neither for female or male applicants. However, having an extremely low female representation in high-school affects outcomes, especially for female candidates. Female students that studied in grades with at most 33% female peers are relatively less likely to apply for a male-dominated field (around 8 percentage points below) and more likely to apply for gender-balanced majors (5.5 p.p. greater). Besides, while men exposed to lower women representation during high-school choose more competitive courses (admission cutoff approximately 5 points larger), female students select less competitive majors when exposed to low shares of female peers in high school (reduces by 3 points the admission cutoff).

3.6 Heterogeneous impacts

To better comprehend the gender peer effects, I explore heterogeneous impacts of the percentage of female high-school peers across different dimensions. First, I examine if the effect varies according to the maternal educational level. Applicants that have parents with higher academic levels may have access to

^{3.5.8}I have run a similar regression, but using other categories: top 25% share of female peers, second and third quartiles, and bottom 25% female peers' proportion. Results are similar when I use this specification.

Table 3.5.10 – Female peers > 66% and ≤ 33% dummies: Impact on the probability of choosing a math-intensive field

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Math or Physics are priority subjects</i>						
Female × Female peers ≤ 33%	0.002 (0.028)	-0.007 (0.028)	-0.006 (0.029)	-0.006 (0.029)	0.003 (0.028)	-0.005 (0.029)
Female × Female peers > 66%	-0.022* (0.013)	-0.019 (0.013)	-0.019 (0.013)	-0.020 (0.013)	-0.019 (0.014)	-0.016 (0.013)
Female peers ≤ 33%	0.008 (0.017)	0.017 (0.018)	0.012 (0.022)	0.013 (0.022)	-0.003 (0.018)	0.007 (0.018)
Female peers > 66%	-0.005 (0.011)	0.007 (0.011)	0.007 (0.011)	0.009 (0.011)	-0.006 (0.012)	0.010 (0.012)
Female	-0.316*** (0.004)	-0.310*** (0.004)	-0.310*** (0.004)	-0.310*** (0.004)	-0.314*** (0.004)	-0.309*** (0.004)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.425	0.424	0.424	0.424	0.425	0.424
<i>Dependent variable: STEM Major</i>						
Female × Female peers ≤ 33%	-0.003 (0.022)	-0.012 (0.022)	-0.011 (0.023)	-0.012 (0.023)	-0.000 (0.022)	-0.009 (0.022)
Female × Female peers > 66%	-0.017 (0.015)	-0.013 (0.015)	-0.014 (0.015)	-0.014 (0.015)	-0.015 (0.015)	-0.011 (0.015)
Female peers ≤ 33%	0.004 (0.020)	0.012 (0.020)	0.008 (0.023)	0.010 (0.023)	-0.006 (0.024)	0.002 (0.023)
Female peers > 66%	-0.016 (0.012)	-0.006 (0.012)	-0.005 (0.012)	-0.005 (0.012)	-0.015 (0.013)	-0.002 (0.013)
Female	-0.253*** (0.005)	-0.247*** (0.004)	-0.247*** (0.004)	-0.247*** (0.004)	-0.251*** (0.005)	-0.246*** (0.005)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.465	0.465	0.465	0.465	0.465	0.465
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: Female peers > 66% and ≤ 33% dummies, and their interactions with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.5.11 – Female peers > 66% and ≤ 33% dummies: Impact on the likelihood of selecting a male-dominated major

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Female-dominated field</i>						
Female × Female peers ≤ 33%	0.027 (0.031)	0.030 (0.031)	0.029 (0.032)	0.029 (0.032)	0.028 (0.032)	0.032 (0.033)
Female × Female peers > 66%	0.007 (0.012)	0.006 (0.012)	0.006 (0.012)	0.005 (0.012)	0.011 (0.012)	0.011 (0.012)
Female peers ≤ 33%	-0.015 (0.013)	-0.016 (0.014)	-0.016 (0.015)	-0.015 (0.015)	-0.015 (0.014)	-0.015 (0.015)
Female peers > 66%	-0.011 (0.008)	-0.008 (0.008)	-0.008 (0.008)	-0.007 (0.008)	-0.017** (0.008)	-0.013 (0.008)
Female	0.196*** (0.005)	0.178*** (0.005)	0.178*** (0.005)	0.178*** (0.005)	0.195*** (0.005)	0.177*** (0.005)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.164	0.164	0.164	0.164	0.164	0.164
<i>Dependent variable: Male-dominated field</i>						
Female × Female peers ≤ 33%	-0.077** (0.038)	-0.085** (0.038)	-0.084** (0.039)	-0.084** (0.039)	-0.076* (0.039)	-0.083** (0.039)
Female × Female peers > 66%	0.004 (0.012)	0.007 (0.012)	0.007 (0.012)	0.006 (0.012)	0.005 (0.013)	0.009 (0.013)
Female peers ≤ 33%	0.047* (0.025)	0.054** (0.025)	0.052* (0.028)	0.053* (0.028)	0.031 (0.027)	0.040 (0.027)
Female peers > 66%	-0.021* (0.011)	-0.013 (0.011)	-0.012 (0.011)	-0.011 (0.011)	-0.023** (0.012)	-0.012 (0.012)
Female	-0.357*** (0.005)	-0.352*** (0.005)	-0.352*** (0.005)	-0.351*** (0.005)	-0.357*** (0.006)	-0.351*** (0.005)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.259	0.259	0.259	0.259	0.259	0.259
<i>Dependent variable: Gender-balanced field</i>						
Female × Female peers ≤ 33%	0.051** (0.020)	0.055*** (0.020)	0.055*** (0.020)	0.055*** (0.021)	0.048** (0.020)	0.052** (0.020)
Female × Female peers > 66%	-0.011 (0.016)	-0.013 (0.016)	-0.013 (0.016)	-0.011 (0.016)	-0.016 (0.017)	-0.020 (0.016)
Female peers ≤ 33%	-0.032** (0.016)	-0.038** (0.017)	-0.036* (0.019)	-0.038** (0.019)	-0.016 (0.019)	-0.025 (0.017)
Female peers > 66%	0.033*** (0.012)	0.021* (0.012)	0.020* (0.012)	0.018 (0.012)	0.041*** (0.013)	0.025** (0.013)
Female	0.161*** (0.005)	0.173*** (0.005)	0.173*** (0.005)	0.173*** (0.005)	0.161*** (0.006)	0.174*** (0.006)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.577	0.577	0.577	0.578	0.577	0.577
<i>Dependent variable: Average % of female applicants in career (2000-2008)</i>						
Female × Female peers ≤ 33%	0.041* (0.024)	0.045* (0.024)	0.045* (0.025)	0.045* (0.025)	0.042* (0.025)	0.046* (0.025)
Female × Female peers > 66%	0.001 (0.006)	-0.001 (0.006)	-0.001 (0.006)	-0.001 (0.006)	0.001 (0.007)	-0.001 (0.007)
Female peers ≤ 33%	-0.026* (0.013)	-0.030** (0.014)	-0.029** (0.015)	-0.030** (0.015)	-0.019 (0.015)	-0.023 (0.014)
Female peers > 66%	0.006 (0.006)	0.002 (0.006)	0.002 (0.006)	0.001 (0.006)	0.005 (0.006)	-0.000 (0.006)
Female	0.229*** (0.003)	0.223*** (0.003)	0.223*** (0.003)	0.223*** (0.003)	0.229*** (0.003)	0.223*** (0.003)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.499	0.499	0.499	0.499	0.499	0.499
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: Female peers > 66% and ≤ 33% dummies, and their interactions with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.5.12 – Female peers > 66% and ≤ 33% dummies: Impact on the choice of competitive majors

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Average # applicants per slot (2000-2008)</i>						
Female × Female peers ≤ 33%	-0.652 (0.740)	-0.649 (0.695)	-0.693 (0.700)	-0.645 (0.700)	-0.805 (0.759)	-0.804 (0.700)
Female × Female peers > 66%	0.906 (0.765)	0.850 (0.757)	0.877 (0.753)	1.000 (0.755)	0.814 (0.794)	0.734 (0.785)
Female peers ≤ 33%	0.339 (0.715)	0.211 (0.663)	0.441 (0.752)	0.381 (0.764)	0.783 (0.766)	0.557 (0.681)
Female peers > 66%	0.670 (0.531)	0.217 (0.516)	0.197 (0.517)	0.077 (0.522)	0.301 (0.556)	-0.277 (0.544)
Female	6.525*** (0.277)	7.409*** (0.288)	7.411*** (0.288)	7.397*** (0.288)	6.497*** (0.286)	7.404*** (0.298)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	33.447	33.467	33.467	33.473	33.447	33.467
<i>Dependent variable: Average final admission scores cutoff (2000-2008)</i>						
Female × Female peers ≤ 33%	-6.954*** (2.226)	-7.797*** (2.500)	-7.936*** (2.455)	-7.830*** (2.459)	-7.506*** (2.302)	-8.351*** (2.568)
Female × Female peers > 66%	1.455 (2.096)	1.489 (2.080)	1.552 (2.064)	1.773 (2.073)	1.205 (2.171)	1.211 (2.158)
Female peers ≤ 33%	4.288** (1.743)	4.642*** (1.776)	5.256*** (1.763)	5.108*** (1.788)	4.279** (1.822)	4.429** (1.832)
Female peers > 66%	0.829 (1.647)	0.208 (1.597)	0.146 (1.608)	0.045 (1.617)	-0.099 (1.715)	-0.816 (1.674)
Female	-4.681*** (0.711)	-0.862 (0.701)	-0.855 (0.701)	-0.880 (0.703)	-4.739*** (0.732)	-0.826 (0.722)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	525.665	525.739	525.739	525.755	525.665	525.739
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: Female peers > 66% and ≤ 33% dummies, and their interactions with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. P-values: * p<0.10, ** p<0.05, *** p<0.01.

information about universities and careers, and, therefore be less sensitive to their high-school peers' choices when selecting a profession. Also, mothers that have a higher-education degree may serve as female role models to their daughters.

I present the heterogeneous gender peer effects by maternal education in Tables 3.6.1, 3.6.2 and 3.6.3. In general, the impacts are qualitatively the same found in the previous specifications. I observe larger impacts on the likelihood of selecting a male-dominated major for applicants with less-educated parents. Moreover, I only observe female peers positively impact the choice of more competitive majors for female applicants whose mothers do not have a higher education degree. For college-educated mother candidates, there are no impacts. The previous literature find similar results, [Brenøe e Zölitz \(2020\)](#) find that the gender peer effects are weaker for female students with college-educated parents than for females whose parents do not have a higher education degree, and [Lavy e Schlosser \(2011\)](#) find that gender peer impacts are more substantial for students with lower socioeconomic status.

I next investigate whether effects are heterogeneous for students from public and private schools. As I explained in Section 3.2, students from private schools can freely select schools, while for public school students, the choice depends on their residential address and slot availability. Although the empirical strategy deals with the selection of students into schools, it is still interesting to compare groups that differently select into high-schools. Also, as students who attend public schools are, on average, from lower socioeconomic backgrounds, it is plausible that they have less information about university admission processes and major choices. Therefore, peers might have a greater influence on them. In general, the impacts of the share of female schoolmates on major choice differ according to the type of schools, but I do not observe a clear pattern of larger (or smaller) impacts for students from public schools (see Tables 3.A.7, 3.A.8, and 3.A.9).

Lastly, I investigate if the gender peer effects vary by cohort size. One might worry that the results are driven by small schools, which have larger variations in female peers' share across cohorts. Another concern is that a few large schools lead the results, as they might have a higher number of applicants at the UNICAMP admission exam. In this exercise, I split the sample of applicants into quartiles,

Table 3.6.1 – Heterogeneous impacts of high-school female peers on the probability of choosing a math-intensive field, by maternal education

	Less High School		High School		Higher Education	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Math and/or Physics are priority subjects</i>						
Female $\times$ % female peers	0.134 (0.090)	0.131 (0.096)	0.041 (0.061)	0.026 (0.064)	0.053 (0.067)	0.053 (0.069)
% female peers - Year 3 High school	-0.150** (0.070)	-0.159** (0.078)	-0.176*** (0.056)	-0.149** (0.063)	-0.132** (0.055)	-0.105* (0.058)
Female	-0.383*** (0.050)	-0.382*** (0.054)	-0.346*** (0.033)	-0.337*** (0.034)	-0.326*** (0.036)	-0.326*** (0.038)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	0.431	0.431	0.437	0.437	0.411	0.411
<i>Dependent variable: STEM Major</i>						
Female $\times$ % female peers	0.029 (0.090)	0.014 (0.096)	0.046 (0.062)	0.035 (0.066)	0.101* (0.057)	0.108* (0.059)
% female peers - Year 3 High school	-0.114* (0.068)	-0.110 (0.079)	-0.132** (0.057)	-0.086 (0.067)	-0.159*** (0.055)	-0.141** (0.060)
Female	-0.265*** (0.050)	-0.256*** (0.053)	-0.280*** (0.034)	-0.273*** (0.035)	-0.293*** (0.031)	-0.297*** (0.032)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	0.491	0.491	0.485	0.486	0.435	0.435
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	No	Yes	No	Yes	No
School size (Pol.)	Yes	No	Yes	No	Yes	No
School linear trend	No	Yes	No	Yes	No	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.6.2 – Heterogeneous impacts of high-school female peers on the likelihood of selecting a male-dominated major, by maternal education

	Less High School		High School		Higher Education	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Female-dominated field</i>						
Female × % female peers	-0.042 (0.087)	-0.038 (0.094)	-0.047 (0.071)	-0.057 (0.076)	0.004 (0.056)	0.015 (0.058)
% female peers - Year 3 High school	0.030 (0.055)	0.049 (0.061)	0.012 (0.043)	0.002 (0.047)	0.016 (0.037)	-0.019 (0.041)
Female	0.244*** (0.049)	0.243*** (0.052)	0.218*** (0.040)	0.222*** (0.043)	0.147*** (0.031)	0.141*** (0.033)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	0.202	0.202	0.165	0.165	0.141	0.141
<i>Dependent variable: Male-dominated field</i>						
Female × % female peers	0.271*** (0.095)	0.281*** (0.104)	0.259*** (0.075)	0.260*** (0.080)	0.154** (0.074)	0.157** (0.077)
% female peers - Year 3 High school	-0.229*** (0.070)	-0.207*** (0.079)	-0.202*** (0.061)	-0.190*** (0.067)	-0.153** (0.060)	-0.138** (0.063)
Female	-0.485*** (0.055)	-0.491*** (0.059)	-0.511*** (0.042)	-0.511*** (0.045)	-0.424*** (0.041)	-0.425*** (0.043)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	0.254	0.254	0.271	0.271	0.251	0.251
<i>Dependent variable: Gender-balanced field</i>						
Female × % female peers	-0.229*** (0.066)	-0.243*** (0.070)	-0.212*** (0.064)	-0.203*** (0.066)	-0.158*** (0.061)	-0.172*** (0.063)
% female peers - Year 3 High school	0.199*** (0.064)	0.158** (0.073)	0.190*** (0.058)	0.189*** (0.066)	0.137** (0.056)	0.157** (0.061)
Female	0.242*** (0.035)	0.243*** (0.038)	0.292*** (0.034)	0.289*** (0.036)	0.277*** (0.033)	0.284*** (0.034)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	0.544	0.544	0.564	0.563	0.608	0.607
<i>Dependent variable: Average % of female applicants in career (2000-2008)</i>						
Female × % female peers	-0.149** (0.066)	-0.154** (0.071)	-0.120** (0.047)	-0.123** (0.050)	-0.081* (0.044)	-0.079* (0.046)
% female peers - Year 3 High school	0.119*** (0.042)	0.118** (0.046)	0.107*** (0.033)	0.099*** (0.036)	0.084*** (0.031)	0.066** (0.033)
Female	0.309*** (0.038)	0.313*** (0.041)	0.300*** (0.027)	0.301*** (0.028)	0.254*** (0.025)	0.254*** (0.026)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	0.507	0.507	0.494	0.494	0.499	0.499
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	No	Yes	No	Yes	No
School size (Pol.)	Yes	No	Yes	No	Yes	No
School linear trend	No	Yes	No	Yes	No	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.6.3 – Heterogeneous impacts of high-school female peers on the choice of competitive majors, by maternal education

	Less High School		High School		Higher Education	
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Average # applicants per slot (2000-2008)</i>						
Female × % female peers	0.280 (2.871)	0.715 (3.100)	3.268 (2.814)	3.130 (2.979)	-0.689 (2.938)	-0.497 (3.028)
% female peers - Year 3 High school	1.925 (2.720)	-0.090 (3.117)	1.780 (2.729)	0.500 (3.023)	4.994* (2.556)	4.660 (2.905)
Female	4.475*** (1.490)	4.050** (1.609)	5.601*** (1.509)	5.706*** (1.600)	9.252*** (1.587)	9.203*** (1.633)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	27.317	27.309	32.523	32.524	37.630	37.612
<i>Dependent variable: Average final admission scores cutoff (2000-2008)</i>						
Female × % female peers	19.499** (9.832)	19.396* (10.524)	27.583*** (9.138)	27.141*** (9.643)	7.457 (7.791)	8.462 (8.094)
% female peers - Year 3 High school	-13.800 (8.851)	-18.965* (9.930)	-16.273** (8.285)	-17.414** (8.742)	1.738 (6.466)	1.081 (7.111)
Female	-19.473*** (5.327)	-19.982*** (5.684)	-16.085*** (5.020)	-15.694*** (5.306)	-0.630 (4.272)	-1.089 (4.432)
Observations	33154	33297	43530	43779	59888	60232
Mean dependent variable	503.526	503.495	523.959	523.961	539.616	539.574
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	No	Yes	No	Yes	No
School size (Pol.)	Yes	No	Yes	No	Yes	No
School linear trend	No	Yes	No	Yes	No	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

using the total number of enrollments in the third year of high-school. For each subsample, I have run separate regressions (the model in equation (3.4.1)). For the first quartile (smaller school sizes), there are no gender peer effects. For the remaining subsamples, we find impacts similar to the gender peer effects for the full sample. Therefore, our findings are relevant to medium and large schools, but the results do not hold for small schools in our sample.^{3.6.1}

3.7 Mechanisms

The findings suggest that the gender composition of high-school peers influence major choices. The proportion of female peers decreases the probability of selecting a math-intensive field for female and male applicants. Also, the share of own-gender schoolmates increases the likelihood of choosing a male-dominated field and reduces the probability of applying for a gender-balanced field. Finally, women apply for more competitive majors when exposed to higher shares of female peers.

To explore potential mechanisms that can explain the main results, I empirically test some channels by estimating equation (3.4.1) using performance measures and declared reasons of major and university choices as our dependent variables.^{3.7.1} First, I examine if a higher share of female schoolmates influences students' achievement. [Lavy e Schlosser \(2011\)](#) finds that an increase in the percentage of female peers increase students' performance and improves the learning environment. To assess the proposed mechanism, I examine if high-school peers' gender composition affects ENEM scores, UNICAMP's Phase 1 performance, and the probability of passing Stage 1 of the UNICAMP admission exam.^{3.7.2} The evidence in Table 3.7.1 suggests that this mechanism may play a role: higher shares of female schoolmates

^{3.6.1}Results are available upon request.

^{3.7.1}Another potential explanation could be students' selection into the UNICAMP admission exam. To test for this hypothesis, I have run a regression where the response variable is the number of students from each school and cohort that applied for UNICAMP major divided by the total number of students in the third year of high-school and the explanatory variable is the share of female students. I have found no correlation.

^{3.7.2}Applicants must pass Phase 1 to attend Phase 2, and only 30% of UNICAMP applicants pass to the second stage. Therefore, I will not have Phase 2 scores for the full sample and will not use them as a performance measure.

increase ENEM and Phase 1 scores, while there is no impact on the probability of passing Stage 1. For ENEM scores, the estimates suggest that the positive effects on performance are larger for men. I would expect that higher performance could influence the major choice, as students with a higher achievement are more likely to be admitted into competitive majors. However, as previously described, I only observe impacts on women's major choices: female students exposed to higher percentages of female peers choose majors with a higher admission cutoff. Higher female responsiveness to changes in performance could explain this pattern (RASK; TIEFENTHALER, 2008; KUGLER; TINSLEY; UKHANEVA, 2017). Though, I am not able to empirically test this hypothesis.^{3.7.3}

Another set of potential explanations for the main results focuses on the applicants' declared reasons for choosing the university and the research field. I first investigate whether the gender composition of high school peers affects students' declared reasons to choose UNICAMP (Table 3.7.2). The student could select one of the following reasons for the university choice: UNICAMP is the best university for the student's major choice, UNICAMP is tuition-free, or UNICAMP has a good reputation. Female applicants are more likely to choose UNICAMP because of the university's reputation or because it is tuition-free, and male applicants tend to choose UNICAMP because it is the best university option for their chosen major. The proportion of female high-school peers slightly increases the likelihood that female students' will choose UNICAMP because it is the best option for the major choice: a one-standard-deviation share of female peers decreases by 0.3 percentage points this likelihood. For men, an increase by the same amount decreases this declared reason by 0.6 p.p. (column (4)). Also, the percentage of female high-school peers reduces the likelihood that a female applicant selects UNICAMP because it is tuition-free, but this effect is only marginally significant (at a 10% level).

I next investigate if the percentage of female high-school peers influences the degree of career certainty and the circumstances of career selection. The

^{3.7.3}I have also checked whether the impacts on performance differs by subject. I run the same regressions, measuring performance as Phase 1 score in each subject. I report the estimates in Table 3.A.10. In general, coefficients are positive, although they are not statistically significant in all models, and there are no gender differences. The main exception is Physics: a higher proportion of female peers increases performance in this subject only for women.

Table 3.7.1 – Impact of high-school female peers on performance

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Normalized ENEM scores</i>						
Female × % female peers	-0.076 (0.065)	-0.138** (0.062)	-0.149** (0.062)	-0.156** (0.062)	-0.067 (0.067)	-0.125* (0.064)
% female peers - Year 3 High school	0.336*** (0.073)	0.281*** (0.061)	0.307*** (0.062)	0.358*** (0.062)	0.372*** (0.090)	0.294*** (0.072)
Female	-0.296*** (0.035)	-0.264*** (0.034)	-0.259*** (0.034)	-0.255*** (0.034)	-0.299*** (0.036)	-0.271*** (0.035)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.003	0.004	0.004	0.005	0.003	0.004
<i>Dependent variable: Stage 1 Survivors</i>						
Female × % female peers	-0.050* (0.030)	-0.038 (0.027)	-0.041 (0.028)	-0.042 (0.028)	-0.040 (0.032)	-0.029 (0.029)
% female peers - Year 3 High school	0.070** (0.030)	-0.002 (0.025)	0.006 (0.025)	0.004 (0.025)	0.047 (0.037)	-0.034 (0.028)
Female	-0.054*** (0.016)	0.016 (0.015)	0.017 (0.015)	0.018 (0.015)	-0.059*** (0.017)	0.012 (0.015)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.317	0.317	0.317	0.317	0.317	0.317
<i>Dependent variable: Phase 1 scores</i>						
Female × % female peers	-0.718 (1.246)	0.196 (0.765)	0.098 (0.753)	0.020 (0.753)	-0.838 (1.274)	-0.045 (0.785)
% female peers - Year 3 High school	5.415*** (1.262)	1.486* (0.807)	1.734** (0.791)	1.960** (0.793)	5.347*** (1.485)	0.933 (0.862)
Female	-0.755 (0.679)	2.493*** (0.413)	2.544*** (0.406)	2.587*** (0.406)	-0.700 (0.695)	2.605*** (0.423)
Observations	137070	136327	136327	135605	137070	136327
Mean dependent variable	50.883	50.915	50.915	50.926	50.883	50.915
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

questionnaire asked how confident the applicant was about her major choice on a 1 to 5 scale, where 1 means the applicant is very undecided and 5 means absolutely decided. Female candidates are, on average, less sure about their career choice than males. The percentage of female peers raises the degree of career certainty for women, while there is no impact on male decision levels. One could interpret this impact as higher female confidence when women's representation is more extensive during high-school. I have also estimated the effects on career choices' reasons and find no gender peer effects. Women are more likely to declare that they have chosen the career for preferences or to contribute to society. Men are more prone to state that they have selected a major because they have a comparative advantage in terms of ability. High-school peers' gender composition does not seem to influence the reasons to choose the area (see Table 3.7.3).

3.8 Conclusion

In this paper, I empirically measure gender peer effects on major choice. In Brazil, students select their major at the end of high-school, when registering for university admission exams. Therefore, high-school peers' gender composition is particularly relevant in our scenario, potentially influencing students' career choices. Using data from Brazilian high-schools linked with admission data from a large selective university called UNICAMP, I estimate the impacts of the female peers' proportion on the likelihood of selecting math-intensive, male-dominated, and competitive majors. To identify gender peer effects, I exploit the idiosyncratic variation in students' gender composition within schools across adjacent cohorts.

The results show that a higher percentage of own-gender schoolmates increases the likelihood of choosing a male-dominated field and decreases the probability of selecting a gender-balanced major. I also find that female high-school peers' share reduces the applicants' likelihood of applying for a math-intensive field both for female and male students. For female applicants, the proportion of women among high-school peers leads to more competitive major choices (higher average admission cutoffs). The gender peer effects are more sizable for students who do

Table 3.7.2 – Reasons: UNICAMP Choice

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Reason: Best university for students' career choice</i>						
Female × % female peers	0.079** (0.032)	0.085*** (0.032)	0.087*** (0.032)	0.087*** (0.032)	0.082** (0.033)	0.089*** (0.033)
% female peers - Year 3 High school	-0.067** (0.030)	-0.054* (0.030)	-0.057* (0.030)	-0.059** (0.030)	-0.069** (0.032)	-0.051 (0.032)
Female	-0.101*** (0.018)	-0.119*** (0.018)	-0.120*** (0.018)	-0.120*** (0.018)	-0.102*** (0.018)	-0.121*** (0.018)
Observations	138860	138324	138324	137581	138860	138324
Mean dependent variable	0.357	0.358	0.358	0.358	0.357	0.358
<i>Reason: Tuition-free</i>						
Female × % female peers	-0.043 (0.030)	-0.051* (0.030)	-0.052* (0.030)	-0.053* (0.030)	-0.048 (0.031)	-0.057* (0.031)
% female peers - Year 3 High school	-0.006 (0.025)	-0.002 (0.025)	-0.001 (0.024)	0.004 (0.024)	0.010 (0.025)	0.013 (0.025)
Female	0.030* (0.017)	0.032** (0.016)	0.032** (0.016)	0.033** (0.016)	0.033* (0.017)	0.036** (0.017)
Observations	138860	138324	138324	137581	138860	138324
Mean dependent variable	0.196	0.196	0.196	0.196	0.196	0.196
<i>Reason: University reputation</i>						
Female × % female peers	-0.012 (0.023)	-0.003 (0.024)	-0.004 (0.024)	-0.005 (0.024)	-0.007 (0.024)	0.001 (0.024)
% female peers - Year 3 High school	0.046** (0.023)	0.028 (0.023)	0.031 (0.023)	0.032 (0.024)	0.028 (0.026)	0.008 (0.026)
Female	0.057*** (0.012)	0.069*** (0.013)	0.069*** (0.013)	0.070*** (0.013)	0.054*** (0.013)	0.066*** (0.013)
Observations	138860	138324	138324	137581	138860	138324
Mean dependent variable	0.248	0.248	0.248	0.248	0.248	0.248
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.7.3 – Reasons: Career Choice

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Degree of career certainty (1 to 5 scale)</i>						
Female × % female peers	0.255*** (0.065)	0.182*** (0.062)	0.176*** (0.062)	0.178*** (0.062)	0.259*** (0.068)	0.188*** (0.065)
% female peers - Year 3 High school	0.051 (0.077)	0.021 (0.062)	0.035 (0.062)	0.038 (0.062)	0.070 (0.087)	0.014 (0.070)
Female	-0.147*** (0.035)	-0.103*** (0.033)	-0.100*** (0.033)	-0.101*** (0.033)	-0.149*** (0.036)	-0.106*** (0.035)
Observations	139375	138818	138818	138074	139375	138818
Mean dependent variable	4.101	4.101	4.101	4.101	4.101	4.101
<i>Reason: Personal ability</i>						
Female × % female peers	-0.005 (0.034)	-0.003 (0.034)	-0.004 (0.033)	-0.007 (0.034)	-0.008 (0.035)	-0.005 (0.034)
% female peers - Year 3 High school	-0.031 (0.030)	-0.043 (0.029)	-0.039 (0.029)	-0.032 (0.029)	-0.034 (0.033)	-0.048 (0.032)
Female	-0.086*** (0.019)	-0.075*** (0.018)	-0.074*** (0.018)	-0.073*** (0.018)	-0.084*** (0.019)	-0.073*** (0.018)
Observations	139275	138701	138701	137958	139275	138701
Mean dependent variable	0.479	0.479	0.479	0.479	0.479	0.479
<i>Reason: Preferences</i>						
Female × % female peers	0.021 (0.029)	0.027 (0.028)	0.028 (0.028)	0.027 (0.028)	0.023 (0.029)	0.030 (0.029)
% female peers - Year 3 High school	0.022 (0.025)	0.024 (0.025)	0.023 (0.025)	0.016 (0.025)	0.016 (0.028)	0.019 (0.028)
Female	0.071*** (0.015)	0.066*** (0.015)	0.066*** (0.015)	0.066*** (0.015)	0.069*** (0.016)	0.065*** (0.015)
Observations	139275	138701	138701	137958	139275	138701
Mean dependent variable	0.284	0.284	0.284	0.284	0.284	0.284
<i>Reason: Contribute to society</i>						
Female × % female peers	0.013 (0.021)	0.004 (0.021)	0.005 (0.021)	0.009 (0.021)	0.012 (0.022)	0.004 (0.021)
% female peers - Year 3 High school	0.035** (0.016)	0.033** (0.016)	0.031* (0.017)	0.029* (0.017)	0.030* (0.018)	0.025 (0.018)
Female	0.042*** (0.011)	0.045*** (0.011)	0.045*** (0.011)	0.042*** (0.011)	0.041*** (0.012)	0.045*** (0.011)
Observations	139275	138701	138701	137958	139275	138701
Mean dependent variable	0.117	0.117	0.117	0.117	0.117	0.117
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

not have college-educated mothers, signaling that female role models might be relevant in career decisions.

When female peers' share is linearly included in the main regressions, I observe modest effect sizes. However, the impacts of having a female majority in the same school-cohort have a slightly larger size, suggesting the existence of non-linear gender peer effects. Besides, when I examine the non-linearity of the peer effects, I find that the results are driven by women exposed to extremely low shares of female peers during high-school. Female applicants with a percentage of female peers below 33% are more likely to choose gender-balanced fields and select less competitive majors at the UNICAMP admission exam.

The percentage of female peers increases students performance, measured by ENEM and UNICAMP scores. As applicants with higher test scores are more likely to be admitted into more prestigious majors, I would expect candidates exposed to higher female peers' shares to select more competitive majors. Still, I only find positive impacts on women's choice of competitive majors. One possibility is that women have a higher responsiveness to changes in performance. Moreover, I note that female schoolmates' proportion increases women's degree of certainty in their career choice, suggesting that exposure to a higher female representation could render women more confident about their choices.

As female peers affect the probability of female applicants choosing competitive and male-dominated majors, it would be interesting to assess if the impacts will result in differences in the labor market. At the moment, I do not have available information on long-term outcomes. Nevertheless, I would like to extend the analysis by merging UNICAMP data with Brazilian formal labor market data - *Relação Anual de Informações Sociais* (RAIS) - to obtain the long-term results.

Appendix – Figures and Tables

Figures

Figure 3.A.1 – Distribution of female students' share in our sample of Brazilian high-schools (Year 3), by graduation year

Figure 3.A.2 – Year-to-year variation in the proportion of female high school students, control for school FE and linear trend

Note: This graph shows year-to-year variation in the percentage of female students in the third year of high-school school peers, plotted along with a normal distribution. To calculate the within-school variation, I have regressed female share on school and cohort fixed-effects plus a school-specific linear trend, and plotted the residuals. Each high school-cohort represents one observation.

Figure 3.A.3 – Robustness check – Exclude majors: Impacts on the likelihood of selecting a math-intensive field

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample) and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.4 – Robustness check – Exclude majors: Impacts on the likelihood of selecting a STEM major

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.5 – Robustness check – Exclude majors: Impacts on the likelihood of selecting a female-dominated field

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.6 – Robustness check – Exclude majors: Impacts on the likelihood of selecting a male-dominated field

Note: Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.7 – Robustness check – Exclude majors: Impacts on the likelihood of selecting a gender-balanced field

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.8 – Robustness check – Exclude majors: Impacts on the share of female applicants in the chosen major

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.9 – Robustness check – Exclude majors: Impacts on major choice's average number of applicants per slot

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Figure 3.A.10 – Robustness check – Exclude majors: Impacts on major choice's average admission cutoff

Note: The scatter plot reports the coefficients of regressions excluding each major of the estimations (59 regressions, one for each major). The horizontal solid red line represents the point estimate for the main regression (full sample), and the dashed black lines indicate the confidence intervals of the primary estimates. The dashed orange line highlights the zero coefficients (no impact). The first graph plots the coefficients for the percentage of female peers in high-school and the second graph reports the interaction term of female peers and the gender dummy.

Tables

Table 3.A.1 – UNICAMP data: Summary statistics

	All	Female	Male	Diff.
Female	0.51 (0.50)			
Grade retention	0.03 (0.18)	0.03 (0.16)	0.04 (0.20)	-0.02***
Admission exam fee exemption	0.04 (0.20)	0.05 (0.23)	0.03 (0.17)	0.02***
Family income - Less than 5 minimum wages	0.16 (0.36)	0.17 (0.38)	0.14 (0.34)	0.04***
Family income - 5 to 10 minimum wages	0.21 (0.41)	0.22 (0.41)	0.20 (0.40)	0.01***
Family income - 10 to 15 minimum wages	0.16 (0.37)	0.16 (0.36)	0.16 (0.37)	-0.01***
Family income - 15 to 20 minimum wages	0.13 (0.34)	0.13 (0.33)	0.13 (0.34)	-0.01***
Family income - 20 to 30 minimum wages	0.16 (0.36)	0.15 (0.36)	0.16 (0.37)	-0.01***
Family income - More than 30 minimum wages	0.19 (0.39)	0.17 (0.38)	0.20 (0.40)	-0.02***
Work	0.17 (0.37)	0.15 (0.35)	0.19 (0.40)	-0.05***
Only public high school	0.29 (0.45)	0.29 (0.46)	0.28 (0.45)	0.01***
Only private elementary school	0.52 (0.50)	0.50 (0.50)	0.54 (0.50)	-0.04***
Only public elementary school	0.31 (0.46)	0.33 (0.47)	0.28 (0.45)	0.05***
Mother - Higher Education	0.44 (0.50)	0.42 (0.49)	0.46 (0.50)	-0.04***
Mother - Upper Secondary School	0.32 (0.47)	0.32 (0.47)	0.32 (0.47)	-0.00
Mother - Lower Secondary School	0.10 (0.31)	0.11 (0.31)	0.10 (0.29)	0.02***
Mother - Primary School	0.10 (0.30)	0.11 (0.31)	0.09 (0.28)	0.02***
Mother - No Primary School	0.04 (0.20)	0.04 (0.20)	0.04 (0.19)	0.01***
Father - Higher Education	0.50 (0.50)	0.47 (0.50)	0.53 (0.50)	-0.06***
Father - Upper Secondary School	0.27 (0.45)	0.28 (0.45)	0.26 (0.44)	0.01***
Father - Lower Secondary School	0.09 (0.28)	0.10 (0.30)	0.08 (0.27)	0.02***
Father - Primary School	0.10 (0.29)	0.11 (0.31)	0.09 (0.28)	0.02***
Father - No Primary School	0.05 (0.21)	0.05 (0.22)	0.04 (0.20)	0.01***
Mother - High-Earning Occupation	0.01 (0.09)	0.01 (0.08)	0.01 (0.10)	-0.00***
Mother - White-Collar Worker	0.44 (0.50)	0.43 (0.50)	0.44 (0.50)	-0.01***
Mother - Grey-Collar Worker	0.13 (0.34)	0.14 (0.34)	0.13 (0.34)	0.01***
Mother - Blue-Collar Worker	0.06 (0.23)	0.06 (0.24)	0.05 (0.22)	0.01***
Mother - House Wife	0.35 (0.48)	0.36 (0.48)	0.35 (0.48)	0.00
Mother - Other occupations	0.01 (0.11)	0.01 (0.11)	0.01 (0.11)	-0.00
Father - High-Earning Occupation	0.02 (0.14)	0.02 (0.13)	0.02 (0.15)	-0.01***
Father - White-Collar Worker	0.69 (0.46)	0.68 (0.47)	0.71 (0.45)	-0.04***
Father - Grey-Collar Worker	0.14 (0.35)	0.15 (0.36)	0.13 (0.34)	0.02***
Father - Blue-Collar Worker	0.12 (0.32)	0.13 (0.33)	0.11 (0.31)	0.02***
Father - Other occupations	0.01 (0.11)	0.01 (0.12)	0.01 (0.11)	0.00***
Age	18.31 (1.36)	18.28 (1.32)	18.34 (1.40)	-0.06***
Household Size	4.23 (1.00)	4.24 (0.99)	4.21 (1.00)	0.03***
Observations	139,896	71,742	68,154	139,896

Notes: The table presents the descriptive statistics of our sample of UNICAMP applicants. I report standard deviations in parenthesis. Column “All” reports the means for the full sample, and columns “Female” and “Male” report the averages by gender. The column “Difference” reports the coefficient of a t-test of mean differences between groups. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.2 – Brazilian School Census: Summary statistics

	All	Female	Male	Diff.
% female peers - Year 3 High school	0.52 (0.10)	0.53 (0.10)	0.51 (0.11)	0.02***
# employees	120.08 (94.09)	115.75 (88.11)	124.63 (99.79)	-8.88***
# high school teachers	40.44 (33.42)	39.15 (29.37)	41.80 (37.16)	-2.66***
# classes - Year 3 High school	5.72 (4.62)	5.71 (4.54)	5.74 (4.69)	-0.04
# students enrolled - Year 3 High school	219.66 (186.97)	219.85 (186.44)	219.45 (187.53)	0.39
# students enrolled - High school (all grades)	696.57 (580.01)	698.99 (579.80)	694.03 (580.21)	4.96
% students enrolled in daytime classes	0.69 (0.37)	0.68 (0.37)	0.69 (0.36)	-0.02***
School located in Sao Paulo state	0.92 (0.27)	0.93 (0.26)	0.91 (0.29)	0.02***
Private school	0.70 (0.46)	0.69 (0.46)	0.70 (0.46)	-0.01***
Federal school	0.02 (0.13)	0.01 (0.10)	0.02 (0.15)	-0.01***
State school	0.27 (0.45)	0.29 (0.45)	0.26 (0.44)	0.02***
City school	0.01 (0.12)	0.02 (0.12)	0.01 (0.11)	0.00***
Urban school	1.00 (0.07)	1.00 (0.07)	0.99 (0.07)	0.00
Observations	139,896	71,742	68,154	139,896

Notes: The table presents the descriptive statistics of our sample of UNICAMP applicants. I report standard deviations in parenthesis. Column “All” reports the means for the full sample, and columns “Female” and “Male” report the averages by gender. The column “Difference” reports the coefficient of a t-test of mean differences between groups. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.3 – Performance: Summary statistics

	All	Female	Male	Diff.
Normalized ENEM scores	0.00 (1.00)	-0.19 (1.00)	0.21 (0.95)	-0.40***
Phase 1 scores	50.88 (16.86)	49.84 (16.83)	51.96 (16.83)	-2.12***
Phase 1 Biology score	3.15 (2.36)	3.02 (2.35)	3.29 (2.36)	-0.27***
Phase 1 Chemistry score	3.00 (2.82)	2.77 (2.70)	3.23 (2.93)	-0.46***
Phase 1 Geography score	4.08 (2.23)	3.88 (2.23)	4.28 (2.22)	-0.40***
Phase 1 History score	3.95 (2.40)	3.76 (2.38)	4.15 (2.40)	-0.39***
Phase 1 Mathematics score	4.77 (3.29)	4.28 (3.22)	5.28 (3.28)	-1.00***
Phase 1 Physics score	4.37 (3.35)	3.89 (3.23)	4.86 (3.40)	-0.97***
Essay score	27.55 (8.60)	28.22 (8.31)	26.84 (8.84)	1.37***
Stage 1 Survivors	0.32 (0.47)	0.27 (0.44)	0.37 (0.48)	-0.09***
Observations	139,896	71,742	68,154	139,896

Notes: The table presents the descriptive statistics of our sample of UNICAMP applicants. I report standard deviations in parenthesis. Column “All” reports the means for the full sample, and columns “Female” and “Male” report the averages by gender. The column “Difference” reports the coefficient of a t-test of mean differences between groups. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.4 – Reasons UNICAMP choice: Summary statistics

	All	Female	Male	Diff.
Best university for students’ career choice	0.36 (0.48)	0.33 (0.47)	0.39 (0.49)	-0.06***
Tuition-free	0.20 (0.40)	0.20 (0.40)	0.19 (0.39)	0.02***
University reputation	0.25 (0.43)	0.27 (0.44)	0.23 (0.42)	0.04***
Observations	138,867	71,211	67,656	138,867

Notes: The table presents the descriptive statistics of our sample of UNICAMP applicants. I report standard deviations in parenthesis. Column “All” reports the means for the full sample, and columns “Female” and “Male” report the averages by gender. The column “Difference” reports the coefficient of a t-test of mean differences between groups. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.5 – Reasons major choice: Summary statistics

	All	Female	Male	Diff.
Degree of career certainty	4.10 (1.01)	4.10 (1.02)	4.11 (1.00)	-0.01*
Reason: Personal ability	0.48 (0.50)	0.43 (0.50)	0.53 (0.50)	-0.10***
Reason: Preferences	0.28 (0.45)	0.32 (0.47)	0.24 (0.43)	0.08***
Reason: Contribute to society	0.12 (0.32)	0.14 (0.35)	0.09 (0.29)	0.05***
Observations	139,277	71,458	67,819	139,277

Notes: Degree of career certainty: The questionnaire asked how confident the applicant was about her major choice on a 1 to 5 scale, where 1 means the very undecided and 5 means absolutely decided. Reasons to major choice: dummy variables for personal ability, preferences or desire to contribute to society. The table presents the descriptive statistics of our sample of UNICAMP applicants. I report standard deviations in parenthesis. Column “All” reports the means for the full sample, and columns “Female” and “Male” report the averages by gender. The column “Difference” reports the coefficient of a t-test of mean differences between groups. P-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.A.6 – Gender composition of field of study: Higher Education Census

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Female-dominated field</i>						
Female × % female peers	-0.037 (0.063)	-0.042 (0.064)	-0.041 (0.064)	-0.043 (0.064)	-0.036 (0.066)	-0.040 (0.067)
% female peers - Year 3 High school	0.012 (0.036)	0.025 (0.036)	0.024 (0.037)	0.019 (0.038)	-0.006 (0.039)	0.010 (0.040)
Female	0.259*** (0.036)	0.247*** (0.036)	0.247*** (0.036)	0.247*** (0.036)	0.258*** (0.037)	0.246*** (0.038)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.230	0.230	0.230	0.230	0.230	0.230
<i>Dependent variable: Male-dominated field</i>						
Female × % female peers	0.211*** (0.065)	0.235*** (0.065)	0.233*** (0.066)	0.234*** (0.066)	0.213*** (0.068)	0.237*** (0.068)
% female peers - Year 3 High school	-0.214*** (0.045)	-0.206*** (0.044)	-0.201*** (0.044)	-0.203*** (0.045)	-0.200*** (0.049)	-0.184*** (0.047)
Female	-0.452*** (0.037)	-0.462*** (0.037)	-0.461*** (0.038)	-0.461*** (0.038)	-0.453*** (0.039)	-0.462*** (0.039)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.239	0.239	0.239	0.239	0.239	0.239
<i>Dependent variable: Gender-balanced field</i>						
Female × % female peers	-0.174*** (0.040)	-0.193*** (0.040)	-0.192*** (0.039)	-0.191*** (0.040)	-0.177*** (0.042)	-0.196*** (0.041)
% female peers - Year 3 High school	0.202*** (0.037)	0.180*** (0.035)	0.177*** (0.035)	0.184*** (0.035)	0.206*** (0.041)	0.174*** (0.038)
Female	0.193*** (0.022)	0.215*** (0.022)	0.214*** (0.022)	0.213*** (0.022)	0.195*** (0.023)	0.217*** (0.023)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.531	0.531	0.531	0.531	0.531	0.531
<i>Dependent variable: % female undergraduate students (2000-2004)</i>						
Female × % female peers	-0.095** (0.039)	-0.106*** (0.039)	-0.105*** (0.040)	-0.106*** (0.040)	-0.097** (0.041)	-0.107*** (0.041)
% female peers - Year 3 High school	0.086*** (0.023)	0.088*** (0.023)	0.086*** (0.024)	0.086*** (0.024)	0.079*** (0.025)	0.078*** (0.025)
Female	0.240*** (0.022)	0.240*** (0.023)	0.240*** (0.023)	0.240*** (0.023)	0.241*** (0.023)	0.241*** (0.024)
Observations	139896	139118	139118	138372	139896	139118
Mean dependent variable	0.461	0.461	0.461	0.461	0.461	0.461
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: To create the response variables, I use data from the Brazilian Higher Education Census, on all undergraduate students from 2000 to 2004. Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.7 – Heterogeneous impacts of high-school female peers on the probability of choosing a math-intensive field, by school type

	Public School		Private School	
	(1)	(2)	(3)	(4)
<i>Dependent variable: Math or Physics are priority subjects</i>				
Female × % female peers	0.061 (0.108)	0.061 (0.111)	0.056 (0.052)	0.061 (0.054)
% female peers - Year 3 High school	-0.150** (0.074)	-0.119* (0.072)	-0.106** (0.043)	-0.116** (0.047)
Female	-0.337*** (0.060)	-0.336*** (0.061)	-0.342*** (0.029)	-0.344*** (0.030)
Observations	42108	42257	96264	96861
Mean dependent variable	0.469	0.469	0.404	0.405
<i>Dependent variable: STEM Major</i>				
Female × % female peers	-0.016 (0.087)	-0.015 (0.092)	0.096* (0.052)	0.095* (0.053)
% female peers - Year 3 High school	-0.101 (0.065)	-0.035 (0.080)	-0.140*** (0.045)	-0.144*** (0.048)
Female	-0.235*** (0.048)	-0.233*** (0.050)	-0.301*** (0.028)	-0.300*** (0.029)
Observations	42108	42257	96264	96861
Mean dependent variable	0.533	0.533	0.435	0.435
School Fixed-Effects	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes
Students' characteristics	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	No	Yes	No
School size (Pol.)	Yes	No	Yes	No
School linear trend	No	Yes	No	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.8 – Heterogeneous impacts of high-school female peers on the likelihood of selecting a male-dominated major, by school type

	Public School		Private School	
	(1)	(2)	(3)	(4)
<i>Dependent variable: Female-dominated field</i>				
Female × % female peers	0.043 (0.124)	0.038 (0.127)	-0.016 (0.040)	-0.014 (0.041)
% female peers - Year 3 High school	0.004 (0.070)	0.012 (0.068)	-0.017 (0.027)	-0.021 (0.031)
Female	0.214*** (0.070)	0.216*** (0.073)	0.166*** (0.023)	0.165*** (0.023)
Observations	42108	42257	96264	96861
Mean dependent variable	0.198	0.198	0.149	0.149
<i>Dependent variable: Male-dominated field</i>				
Female × % female peers	0.338** (0.139)	0.334** (0.146)	0.131** (0.056)	0.143** (0.056)
% female peers - Year 3 High school	-0.284*** (0.087)	-0.225** (0.093)	-0.106** (0.046)	-0.123** (0.048)
Female	-0.526*** (0.080)	-0.525*** (0.084)	-0.423*** (0.031)	-0.429*** (0.031)
Observations	42108	42257	96264	96861
Mean dependent variable	0.291	0.291	0.244	0.245
<i>Dependent variable: Gender-balanced field</i>				
Female × % female peers	-0.381*** (0.057)	-0.372*** (0.060)	-0.115** (0.052)	-0.129** (0.053)
% female peers - Year 3 High school	0.280*** (0.057)	0.213*** (0.067)	0.123*** (0.044)	0.145*** (0.046)
Female	0.313*** (0.031)	0.309*** (0.033)	0.257*** (0.029)	0.264*** (0.029)
Observations	42108	42257	96264	96861
Mean dependent variable	0.511	0.511	0.607	0.606
<i>Dependent variable: Average % of female applicants in career (2000-2008)</i>				
Female × % female peers	-0.152 (0.095)	-0.153 (0.099)	-0.067** (0.031)	-0.071** (0.032)
% female peers - Year 3 High school	0.137*** (0.053)	0.121** (0.055)	0.051** (0.024)	0.055** (0.026)
Female	0.316*** (0.054)	0.317*** (0.057)	0.254*** (0.018)	0.256*** (0.018)
Observations	42108	42257	96264	96861
Mean dependent variable	0.487	0.488	0.505	0.505
School Fixed-Effects	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes
Students' characteristics	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	No	Yes	No
School size (Pol.)	Yes	No	Yes	No
School linear trend	No	Yes	No	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.9 – Heterogeneous impacts of high-school female peers on choice of competitive majors, by school type

	Public School		Private School	
	(1)	(2)	(3)	(4)
<i>Dependent variable: Average # applicants per slot (2000-2008)</i>				
Female × % female peers	-3.522 (2.702)	-3.587 (2.787)	0.101 (2.377)	0.037 (2.431)
% female peers - Year 3 High school	-1.626 (2.601)	-5.398** (2.693)	5.888*** (2.072)	6.255*** (2.226)
Female	5.602*** (1.414)	5.583*** (1.454)	8.821*** (1.317)	8.828*** (1.347)
Observations	42108	42257	96264	96861
Mean dependent variable	24.807	24.799	37.264	37.248
<i>Dependent variable: Average final admission scores cutoff (2000-2008)</i>				
Female × % female peers	16.661 (10.189)	15.788 (10.514)	7.473 (6.269)	7.587 (6.377)
% female peers - Year 3 High school	-32.898*** (9.095)	-39.571*** (8.969)	8.153 (5.235)	8.614 (5.802)
Female	-18.423*** (5.488)	-18.144*** (5.650)	-1.445 (3.507)	-1.544 (3.569)
Observations	42108	42257	96264	96861
Mean dependent variable	498.029	497.981	537.883	537.849
School Fixed-Effects	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes
Students' characteristics	Yes	Yes	Yes	Yes
Peers' characteristics	Yes	No	Yes	No
School size (Pol.)	Yes	No	Yes	No
School linear trend	No	Yes	No	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

Table 3.A.10 – Impacts of high-school female peers on Phase 1 performance, by subject

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Dependent variable: Phase 1 score Biology</i>						
Female × % female peers	-0.187 (0.219)	-0.145 (0.168)	-0.149 (0.167)	-0.151 (0.168)	-0.137 (0.226)	-0.104 (0.173)
% female peers - Year 3 High school	0.676*** (0.195)	0.208 (0.143)	0.218 (0.140)	0.222 (0.139)	0.746*** (0.212)	0.203 (0.150)
Female	-0.101 (0.122)	0.280*** (0.094)	0.282*** (0.093)	0.284*** (0.093)	-0.127 (0.126)	0.256*** (0.096)
Mean dependent variable	3.151	3.154	3.154	3.155	3.151	3.154
<i>Dependent variable: Phase 1 score Chemistry</i>						
Female × % female peers	-0.095 (0.276)	-0.037 (0.226)	-0.053 (0.226)	-0.043 (0.227)	-0.122 (0.286)	-0.082 (0.236)
% female peers - Year 3 High school	0.857*** (0.230)	0.387** (0.178)	0.426** (0.177)	0.435** (0.176)	0.737*** (0.258)	0.190 (0.193)
Female	-0.280* (0.155)	0.107 (0.127)	0.116 (0.127)	0.110 (0.127)	-0.272* (0.161)	0.128 (0.132)
Mean dependent variable	2.996	2.997	2.997	2.998	2.996	2.997
<i>Dependent variable: Phase 1 score History</i>						
Female × % female peers	-0.101 (0.162)	-0.034 (0.121)	-0.026 (0.120)	-0.045 (0.120)	-0.089 (0.165)	-0.032 (0.125)
% female peers - Year 3 High school	0.523*** (0.151)	0.118 (0.123)	0.107 (0.121)	0.136 (0.124)	0.500*** (0.163)	0.042 (0.130)
Female	-0.237*** (0.088)	0.110* (0.066)	0.105 (0.065)	0.115* (0.065)	-0.241*** (0.090)	0.109 (0.067)
Mean dependent variable	3.953	3.957	3.957	3.958	3.953	3.957
<i>Dependent variable: Phase 1 score Physics</i>						
Female × % female peers	0.565*** (0.194)	0.681*** (0.180)	0.646*** (0.177)	0.660*** (0.178)	0.598*** (0.198)	0.698*** (0.184)
% female peers - Year 3 High school	0.369 (0.229)	-0.220 (0.167)	-0.137 (0.158)	-0.121 (0.159)	0.250 (0.276)	-0.409** (0.189)
Female	-1.092*** (0.105)	-0.626*** (0.099)	-0.608*** (0.097)	-0.617*** (0.097)	-1.109*** (0.107)	-0.635*** (0.101)
Mean dependent variable	4.367	4.369	4.369	4.370	4.367	4.369
<i>Dependent variable: Phase 1 score Geography</i>						
Female × % female peers	-0.157 (0.123)	-0.107 (0.099)	-0.117 (0.099)	-0.122 (0.099)	-0.173 (0.128)	-0.132 (0.103)
% female peers - Year 3 High school	0.320** (0.133)	0.021 (0.104)	0.045 (0.104)	0.054 (0.103)	0.382** (0.154)	0.041 (0.117)
Female	-0.244*** (0.066)	0.019 (0.053)	0.024 (0.053)	0.028 (0.053)	-0.236*** (0.068)	0.030 (0.055)
Mean dependent variable	4.077	4.082	4.082	4.082	4.077	4.082
<i>Dependent variable: Phase 1 score Mathematics</i>						
Female × % female peers	-0.248 (0.231)	-0.028 (0.180)	-0.039 (0.181)	-0.039 (0.182)	-0.258 (0.238)	-0.055 (0.185)
% female peers - Year 3 High school	0.636*** (0.189)	0.176 (0.161)	0.198 (0.160)	0.227 (0.163)	0.413* (0.213)	-0.071 (0.178)
Female	-0.682*** (0.124)	-0.283*** (0.095)	-0.277*** (0.096)	-0.276*** (0.096)	-0.678*** (0.128)	-0.270*** (0.098)
Mean dependent variable	4.769	4.773	4.773	4.775	4.769	4.773
<i>Dependent variable: Essay score</i>						
Female × % female peers	-0.502 (0.527)	-0.165 (0.458)	-0.196 (0.457)	-0.273 (0.457)	-0.670 (0.534)	-0.382 (0.467)
% female peers - Year 3 High school	2.052*** (0.567)	0.828* (0.502)	0.914* (0.502)	1.045** (0.504)	2.345*** (0.651)	0.976* (0.544)
Female	1.898*** (0.283)	2.914*** (0.246)	2.929*** (0.245)	2.971*** (0.245)	1.983*** (0.287)	3.021*** (0.251)
Mean dependent variable	27.547	27.560	27.560	27.565	27.547	27.560
Observations	137148	136404	136404	135682	137148	136404
School Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed-Effects	Yes	Yes	Yes	Yes	Yes	Yes
Students' characteristics	No	Yes	Yes	Yes	No	Yes
School size (Pol.)	No	No	Yes	Yes	No	No
Peers' characteristics	No	No	No	Yes	No	No
School linear trend	No	No	No	No	Yes	Yes

Notes: Key explanatory variables: percentage of female peers (leave-me out share) in the third year of high-school and its interaction with the gender dummy. I run OLS regressions. Clustered standard-errors in parenthesis (at school-level). Students' covariates: age, squared age, normalized ENEM scores, work dummy. School size: Third-order polynomial of the number of enrollments in the third year of high-school. Peers' characteristics: age and share of daytime enrollments. P-values: * p<0.10, ** p<0.05, *** p<0.01.

REFERENCES

- AKYOL, P.; KRISHNA, K.; WANG, J. *Taking PISA Seriously: How Accurate are Low Stakes Exams?* [S.l.], 2019. (Working Paper Series, 24930). Disponível em: <http://www.nber.org/papers/w24930>.
- ALTMEJD, A. et al. Siblings' Spillover Effects on College and Major Choice: Evidence from Chile, Croatia and Sweden. Unpublished manuscript. 2019.
- ANDERSON, M. L. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, v. 103, n. 484, p. 1481–1495, 2008.
- ANELLI, M.; PERI, G. The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, v. 129, n. 618, p. 553–602, 2019.
- AZMAT, G.; CALSAMIGLIA, C.; IRIBERRI, N. Gender differences in response to big stakes. *Journal of the European Economic Association*, v. 14, n. 6, p. 1372–1400, 2016.
- BABCOCK, L.; RECALDE, M. P.; VESTERLUND, L. Gender differences in the allocation of low-promotability tasks: The role of backlash. *American Economic Review*, v. 107, n. 5, p. 131–135, 2017.
- BABCOCK, L. et al. Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, v. 107, n. 3, p. 714–747, 2017.
- BERTRAND, M. Coase Lecture – The Glass Ceiling. *Economica*, v. 85, n. 338, p. 205–231, 2018.
- BIFULCO, R.; FLETCHER, J. M.; ROSS, S. L. The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, v. 3, n. 1, p. 25–53, 2011.
- BLACK, S. E.; DEVEREUX, P. J.; SALVANES, K. G. Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics*, v. 31, n. 1, p. 119–152, 2013.
- BOOTH, A.; CARDONA-SOSA, L.; NOLEN, P. Gender differences in risk aversion: Do single-sex environments affect their development? *Journal of Economic Behavior and Organization*, v. 99, p. 126–154, 2014.

- BORDALO, P. et al. Beliefs about gender. *American Economic Review*, v. 109, n. 3, p. 739–773, 2019.
- BOSTWICK, V. K.; WEINBERG, B. A. *Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs*. [S.l.], 2018. (Working Paper Series, 25028). Disponível em: <<http://www.nber.org/papers/w25028>>.
- BOTTIA, M. C. et al. Growing the Roots of STEM Majors: Female Math and Science High School Faculty and the Participation of Students in STEM. *Economics of Education Review*, v. 45, p. 14–27, 2015.
- BRENØE, A. A.; ZÖLITZ, U. Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics*, v. 38, n. 4, p. 1–46, 2020.
- BRODATY, T.; GURGAND, M. Good Peers or Good Teachers? Evidence from a French University. *Economics of Education Review*, v. 54, p. 62–78, 2016.
- BROWN, C.; CORCORAN, M. Sex-Based Differences in School Content and the Male-Female Wage Gap. *Journal of Labor Economics*, v. 15, n. 3, p. 431–465, 1997.
- BUSER, T.; NIEDERLE, M.; OOSTERBEEK, H. Gender, competition and career choices. *Quarterly Journal of Economics*, v. 129, n. 3, p. 1409–1447, 2014.
- BUSER, T.; PETER, N.; WOLTER, S. C. Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review*, v. 107, n. 5, p. 125–130, 2017.
- BWE. [Brazilian Women in Economics (BWE) Annual Report] *As mulheres nos diferentes estágios da carreira acadêmica em Economia no Brasil*. 2018. 1–21 p. Available at http://paineira.usp.br/bwe/?page_id=454. Last accessed on July 26, 2020.
- CAI, X. et al. Gender gap under pressure: Evidence from China's national college entrance examination. *Review of Economics and Statistics*, v. 101, n. 2, p. 249–263, 2019.
- CAPES. [Qualis Journal Rankings] *Qualis Periódicos*. 2020. Available at <https://sucupira.capes.gov.br/sucupira/public/index.xhtml>. Last accessed on July 26, 2020.
- CARRELL, S. E.; PAGE, M. E.; WEST, J. E. Sex and Science: How Professor Gender Perpetuates the Gender Gap. *Quarterly Journal of Economics*, v. 125, n. 3, p. 1101–1144, ago. 2010.

CNPQ. [Lattes CV Platform] *Plataforma Lattes*. 2020. Available at <http://lattes.cnpq.br>. Last accessed on May 1, 2020.

CSWEP. *Committee on the Status of Women in the Economics Profession Annual Report*. 2017. 1–16 p. Available at <https://www.aeaweb.org/content/file?id=6388>. Last accessed on July 26, 2020.

DAHL, G. B. Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica*, v. 70, n. 6, p. 2367–2420, 2002.

DASGUPTA, N.; MCMANUS, M.; HUNSINGER, M. Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *PNAS*, v. 112, n. 16, p. 4988–4993, 2015.

DEE, T. S. A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review*, v. 95, n. 2, p. 158–165, maio 2005.

DEE, T. S. Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, v. 42, n. 3, p. 528–554, 2007.

ELLISON, G.; SWANSON, A. The gender gap in secondary school mathematics at high achievement levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, v. 24, n. 2, p. 109–128, 2010.

ELLISON, G.; SWANSON, A. *Dynamics of the Gender Gap in High Math Achievement*. [S.l.], 2018. (Working Paper Series, 24910). Disponível em: <http://www.nber.org/papers/w24910>.

ESTEVAN, F.; GALL, T.; MORIN, L.-P. Redistribution without distortion: Evidence from an affirmative action program at a large Brazilian university. *Economic Journal*, v. 129, n. 619, p. 1182–1220, 2019.

FAIRLIE, R. W.; HOFFMANN, F.; OREOPOULOS, P. A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom. *American Economic Review*, v. 104, n. 8, p. 2567–2591, 2014.

FELD, J.; ZÖLITZ, U. Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics*, v. 35, n. 2, p. 387–428, 2017.

FIRPO, S.; JALES, H.; PINTO, C. Measuring peer effects in the Brazilian school system. *Applied Economics*, v. 47, n. 32, p. 3414–3438, 2015.

FUNK, P.; IRIBERRI, N.; SAVIO, G. Does Scarcity of Female Instructors Create Demand for Diversity among Students? Evidence from Observational and Experimental Data. Unpublished manuscript. 2019.

GAULE, P.; PIACENTINI, M. An Advisor Like Me? Advisor Gender and Post-Graduate Careers in Science. *Research Policy*, v. 47, n. 4, p. 805–813, 2018.

GNEEZY, U.; NIEDERLE, M.; RUSTICHINI, A. Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics*, v. 118, n. 3, p. 1049–1074, 2003.

GOULAS, S.; MEGALOKONOMOU, R.; ZHANG, Y. *Does the Girl Next Door Affect Your Academic Outcomes and Career Choices?* [S.l.], 2018. (Discussion Paper Series, 11910). Disponível em: <<http://www.iza.org/de/publications/dp/11910>>.

GRIFFITH, A. L. Faculty Gender in the College Classroom: Does It Matter for Achievement and Major Choice? *Southern Economic Journal*, v. 81, n. 1, p. 211–231, 2014.

GURYAN, J.; KROFT, K.; NOTOWIDIGDO, M. J. Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics*, v. 1, n. 4, p. 34–68, 2009.

HAN, L.; LI, T. The Gender Difference of Peer Influence in Higher Education. *Economics of Education Review*, v. 28, n. 1, p. 129–134, 2009.

HECKMAN, J.; VYTLACIL, E. 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*, v. 33, n. 4, p. 974–987, 1998.

HECKMAN, J. J.; URZUA, S.; VYTLACIL, E. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics*, v. 88, n. 3, p. 389–432, 2006. Disponível em: <<https://doi.org/10.1162/rest.88.3.389>>.

HILL, A. J. The girl next door: the effect of peer gender composition on high school achievement. *American Economic Journal: Applied Economics*, v. 7, n. 3, p. 147–177, 2015.

HOFFMAN, F.; OREOPOULOS, P. A Professor Like Me: The Influence of Instructor Gender on College Achievement. *Journal of Human Resources*, v. 44, n. 2, p. 479–494, 2009.

- HOXBY, C. *Peer Effects in the Classroom: Learning from Gender and Race Variation*. [S.l.], 2000. (Working Paper Series, 7867). Disponível em: <http://www.nber.org/papers/w7867>.
- HU, F. Do girl peers improve your academic performance? *Economics Letters*, v. 137, p. 54–58, 2015.
- INEP/MEC. [Brazilian School Census] *Censo Escolar da Educação Básica (in Portuguese)*. 1998–2003. Available at <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-escolar>. Last accessed on December 29, 2020.
- INEP/MEC. [Brazilian Higher Education Census] *Censo Escolar da Educação Superior (in Portuguese)*. 2000–2004. Available at <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior>. Last accessed on January 19, 2020.
- INEP/MEC. [2008 Higher Education Statistics] *Sinopse Estatística da Educação Superior 2008 (in Portuguese)*. 2008. Available at http://download.inep.gov.br/download/censo/2008/sinop_sup_2008_versao_preliminar.zip. Last accessed on July 26, 2020.
- IRIBERRI, N.; REY-BIEL, P. Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *Economic Journal*, v. 129, n. 620, p. 1863–1893, 2019.
- JACKSON, C. K. Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics*, v. 96, n. 1-2, p. 173–187, 2012.
- JOENSEN, J. S.; NIELSEN, H. S. Spillovers in education choice. *Journal of Public Economics*, v. 157, p. 158–183, 2018.
- KATO, T.; SONG, Y. *Advisor Like Me: Does Gender Matter?* [S.l.], 2018. 1–39 p. (Discussion Paper Series, 11575). Disponível em: <http://ftp.iza.org/dp11575.pdf>.
- KIRKEBOEN, L. J.; LEUVEN, E.; MOGSTAD, M. Field of study, earnings, and self selection. *Quarterly Journal of Economics*, v. 131, n. 3, p. 1057–1111, 2016.
- KUGLER, A. D.; TINSLEY, C. H.; UKHANEVA, O. *Choice of Majors: Are Women Really Different from Men?* [S.l.], 2017. (Working Paper Series, 23735). Disponível em: <http://www.nber.org/papers/w23735>.
- LAVY, V.; SCHLOSSER, A. Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, v. 3, n. 2, p. 1–33, 2011.

- LEPINE, A.; ESTEVAN, F. Do Ability Peer Effects Matter for Academic and Labor Market Outcomes? Unpublished manuscript. 2020.
- MACHADO, C.; SZERMAN, C. *Centralized Admission and the Student-College Match*. [S.l.], 2016. (Discussion Paper Series, 10251). Disponível em: <https://www.iza.org/publications/dp/10251>.
- MACHIN, S.; PUHANI, P. A. Subject of degree and the gender wage differential: evidence from the UK and Germany. *Economics Letters*, v. 79, p. 393–400, 2003.
- MANI, A.; RILEY, E. *Social Networks, Role Models, Peer Effects, and Aspirations*. [S.l.], 2019. 1–24 p. (WIDER Working Paper, 120). Disponível em: <http://www.wider.unu.edu/publication/social-networks-role-models-peer-effects-and-aspirations>.
- MANSKI, C. F. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, v. 60, p. 531–542, 1993.
- MANSOUR, H. et al. *The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students*. [S.l.], 2020. (Working Paper Series, 26822). Disponível em: <http://www.nber.org/papers/w26822>.
- MARTORELL, P.; McFarlin Jr., I. Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes. *Review of Economics and Statistics*, v. 93, n. 2, p. 436–454, 2011.
- MOUGANIE, P.; WANG, Y. High-Performing Peers and Female STEM Choices in School. *Journal of Labor Economics*, v. 38, n. 3, p. 805–841, 2020.
- NIEDERLE, M. A Gender Agenda: A Progress Report on Competitiveness. *American Economic Review*, v. 107, n. 5, p. 115–119, 2017.
- NIEDERLE, M.; VESTERLUND, L. Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, v. 122, n. 3, p. 1067–1101, 2007.
- NIEDERLE, M.; VESTERLUND, L. Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, v. 24, n. 2, p. 129–144, 2010.
- OOSTERBEEK, H.; EWIJK, R. V. Gender Peer Effects in University: Evidence from a Randomized Experiment. *Economics of Education Review*, v. 38, p. 51–63, 2014.
- ORS, E.; PALOMINO, F.; PEYRACHE, E. Performance gender gap: Does competition matter? *Journal of Labor Economics*, v. 31, n. 3, p. 443–499, 2013.

PAREDES, V. A Teacher Like Me or a Student Like Me? Role Model versus Teacher Bias Effect. *Economics of Education Review*, v. 39, p. 38–49, 2014.

PAREDES, V. A.; PASERMAN, M. D.; PINO, F. *Does Economics Make You Sexist?* [S.l.], 2020. (Working Paper Series, 27070). Disponível em: <http://www.nber.org/papers/w27070>.

PORTER, C.; SERRA, D. Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*, v. 12, n. 3, p. 226–254, 2020.

RASK, K.; TIEFENTHALER, J. The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, v. 27, n. 6, p. 676–687, 2008.

Receita Federal. [Brazilian Revenue Service: CNPJ Public Data] *Dados Públicos CNPJ*. 2020. Available at <http://receita.economia.gov.br/orientacao/tributaria/cadastros/cadastro-nacional-de-pessoas-juridicas-cnpj/dados-publicos-cnpj>. Last accessed on Apr 15, 2020.

ROSA, L. The unintended consequences of public school subsidies on educational markets. Unpublished manuscript. 2019.

SACERDOTE, B. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, v. 116, n. 2, p. 681–704, 2001.

SACERDOTE, B. Experimental and Quasi-Experimental Analysis of Peer Effects : Two Steps Forward ? *Annual Review of Economics*, v. 6, p. 253–272, 2014.

SANTOS, K. G. dos. *Does It Matter which Top Institution You Choose? A Case Study of Brazilian Graduate Admissions*. Dissertação (Master's Thesis) — Sao Paulo School of Economics - Fundação Getulio Vargas, 2020.

SCHLOSSER, A.; NEEMAN, Z.; ATTALI, Y. Differential performance in high vs. low stakes tests: Evidence from the GRE test. *Economic Journal*, v. 129, n. 623, p. 2916–2948, 2019.

SCHNEEWEIS, N.; ZWEIMÜLLER, M. Girls, girls, girls: Gender composition and female school choice. *Economics of Education Review*, v. 31, n. 4, p. 482–500, 2012.

SCHØNE, P.; SIMSON, K. V.; STRØM, M. Peer gender and educational choices. *Empirical Economics*, v. 59, p. 1763–1797, 2020.

SCOPPA, V.; PAOLA, M. D. Peer Group Effects in the Academic Performance of Italian Students. *Applied Economics*, v. 42, n. 17, p. 2203–2215, 2010.

STINEBRICKNER, R.; STINEBRICKNER, T. R. Working during School and Academic Performance. *Journal of Labor Economics*, v. 21, n. 2, p. 473–491, 2003.

Times Higher Education. *Latin America University Rankings*. 2019. Available at www.timeshighereducation.com/world-university-rankings/2019/latin-america-university-rankings. Last accessed on July 26, 2020.

UNESCO. *International Standard Classification of Education - ISCED 2011*. 2011. Available at <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>. Last accessed on December 29, 2020.

WHITMORE, D. Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review*, v. 95, n. 2, p. 199–203, 2005.

WOOLDRIDGE, J. M. Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, v. 87, n. 2, p. 385–390, 2005.

WU, A. H. Gendered Language on the Economics Job Market. *AEA Papers and Proceedings*, v. 108, p. 175–179, 2018.

YOON, T. et al. Sex differences in response to cognitive stress during a fatiguing contraction. *Journal of Applied Physiology*, v. 107, n. 5, p. 1486–1496, 2009.

YOUNG, A. Channelling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results. *Quarterly Journal of Economics*, v. 134, n. 2, p. 557–598, 2019.

ZÖLITZ, U.; FELD, J. *The Effect of Peer Gender on Major Choice in Business School*. [S.l.], 2020. (Discussion Paper Series, 13396). Disponível em: <http://www.iza.org/de/publications/dp/13396>.