Research article

Gender differences in performance under different evaluation schemes and the leaky pipeline in economics♦

Fabiana **ROCHA**[1]
frocha@usp.br | https://orcid.org/0000-0002-9314-0023

Paula **PEREDA**[1]
pereda@usp.br | https://orcid.org/0000-0002-6717-9522

Maria Dolores Montoya **DIAZ**[1]
madmdiaz@usp.br | https://orcid.org/0000-0003-2385-2073

Bruna **BORGES**[2]
bruna.borges@fgv.br | https://orcid.org/0000-0002-2434-9034

Gabriel Facundes **MONTEIRO**[3]
gmonteir@andrew.cmu.edu | https://orcid.org/0000-0001-7281-1548

Luiza **KARPAVICIUS**[4]
luiza.karpavicius@envs.au.dk | https://orcid.org/0000-0003-4596-5052

Liz **MATSUNAGA**[1]
liz.mtn@gmail.com | https://orcid.org/0000-0002-8731-1260

Clara **BRENCK**[5]
clarazb@cedeplar.ufmg.br | https://orcid.org/0000-0002-0110-8668

[1] Universidade de São Paulo, Faculdade de Economia, Administração, Contabilidade e Atuária, Departamento de Economia, São Paulo, SP, Brasil.
[2] Fundação Getúlio Vargas, Escola de Economia de São Paulo, São Paulo, SP, Brasil.
[3] Carnegie Mellon University, Heinz School of Public Policy and Management, Pittsburgh, PA, United. States.
[4] Aarhus University, Department of Environmental Sciences, Section for Environmental Social Science and Geography, Aarhus, Denmark.
[5] Universidade Federal de Minas Gerais, Departamento de Economia, Belo Horizonte, MG, Brasil.

Recebido: 03/10/2023.
Revisado: 10/05/2024.
Aceito: 26/06/2024.
DOI: https://doi.org/10.1590/1980-53575441fpmbgllc

Abstract

The leaky pipeline remains a persistent challenge to achieving gender diversity in the economics career. In this study, we contribute to the existing literature by investigating gender differences in academic performance in economics in Brazil in two distinct stages: undergraduate studies and graduate admission exams. We use individual-level data from the national admission exam for economics graduate programs (ANPEC exam) and undergraduate records from the University of São Paulo. Women are less likely to rank among the top 100 ANPEC applicants and perform worse than men in all exam subjects. Meanwhile, we find consistent evidence that female students perform similarly to their male counterparts in undergraduate courses with comparable content to those evaluated on the ANPEC exam. Since the students taking the ANPEC exam were exposed to the same higher education program, after controlling for observable characteristics, we can relate the differences in performance to the exam itself rather than to differences in learning abilities. While we cannot identify the source of the performance gap, as the ANPEC exam and undergraduate grading system vary in terms of stakes, grading scheme, risk, and competitiveness (all of which can potentially affect women and men differently), we argue that our evidence suggests the need to reconsider admission exam designs to address the leaky pipeline in economics.

Keywords

Gender; Exams; Higher education; Pipeline; Women in economics.

Diferenças de gênero no desempenho sob diferentes esquemas de avaliação e o *leaky pipeline* em economia

Resumo

O *leaky pipeline* ("vazamento no duto") continua sendo um desafio persistente para alcançar a diversidade de gênero na carreira de economia. Neste estudo, contribuímos para a literatura existente investigando as diferenças de gênero no desempenho acadêmico em economia no Brasil em duas etapas distintas: estudos de graduação e exames de admissão na pós-graduação. Utilizamos dados individuais do exame nacional de admissão para programas de pós-graduação em economia (exame da ANPEC) e registros de graduação da Universidade de São Paulo. As mulheres têm menor probabilidade de se classificar entre os 100 melhores candidatos da ANPEC e apresentam desempenho inferior aos homens em todas as disciplinas do exame. Enquanto isso, encontramos evidências consistentes de que as alunas apresentam desempenho semelhante ao de seus colegas do sexo masculino em disciplinas de graduação com conteúdo comparável aos avaliados no exame ANPEC. Uma vez que os alunos que fazem o exame da ANPEC foram expostos ao mesmo programa de ensino superior, após controlar por características observáveis, podemos relacionar as diferenças de desempenho ao próprio exame, e não às diferenças nas habilidades de aprendizagem. Embora não possamos identificar a fonte da diferença de desempenho, já que o exame ANPEC e o sistema de notas de graduação variam em termos de riscos, formato de avaliação, e competitividade (todos os quais podem potencialmente afetar mulheres e homens de forma diferente), argumentamos que nossas evidências sugerem a necessidade de reconsiderar os desenhos dos exames de admissão para abordar o problema do *leaky pipeline* em economia.

Palavras-chave
Gênero; Exames; Educação superior; *Pipeline*; Mulheres em economia.

JEL Classification
I23; I24; J16

1. Introduction

Women are not only underrepresented in academic economics, but they also experience slow career progression. In 2023, according to the Committee on the Status of Women in the Economics Profession (CSWEP), women accounted for only 33.7% of assistant professors, 27.6% of associate professors with tenure, and less than 17.5% of full professors in US Ph.D. granting economics departments. The widening gender gap in economics at each career stage is known as the "leaky pipeline."

A considerable and growing body of research discusses the barriers at the various stages of the training and promotional pipeline in academic economics. Examples are Ceci et al. (2014) and Ceci et al. (2015) for tenure promotions, Hengel (2022) for differences in top journals publications, Boudreau and Kaushik (2022) for women's preference towards collaborations and Sarsons (2017) for co-authorship impacts on women. As a matter of fact, as pointed out by Goldin (2023) and Avilova and Goldin (2018), underrepresentation is observable as early as the undergraduate level.

This paper aims to understand the initial stages of the pipeline, comparing the performance of male and female students throughout their undergraduate journey and on the Brazilian unified national master's entrance exam in economics (ANPEC exam).

Our sample consists of students who graduated in economics from the University of São Paulo (USP) and were admitted between 2000 and 2010. We analyze their scores on the admission exam to the undergraduate economics program at USP (FUVEST) and follow their performance throughout college. Finally, we evaluate the same student's performance in the ANPEC exam, the admission exam to the master's programs. The ANPEC exam, run by the National Association of Postgraduate Programs in Economics (*Associação Nacional dos Centros de Pós-Graduação em*

Economia - ANPEC), is highly competitive. For instance, to be accepted by one of the "top four" institutions, candidates must be among the 80 best-ranked out of more than 1,000 candidates. The stakes are high, as attending a top master's program impacts students' career prospects (Estevan and Santos, 2022). A distinctive characteristic of the exam is that the questions are mainly true or false, where wrong answers are penalized (they yield negative points). Omissions are not penalized and yield zero points.

We first guarantee that there are no significant differences in performance before admission to the economics bachelor program, i.e., that male and female students' rankings in FUVEST are similar. Next, we provide evidence that female students perform as well as, or even better than, male students in the mandatory courses during the undergraduate program (Microeconomics, Macroeconomics, Mathematics, and Statistics), topics which also make up the ANPEC exam.

Gender does not affect selection into the ANPEC exam. Although higher grades in compulsory courses are positively related to the probability of students taking the ANPEC exam, we find no gender differences in this likelihood. Within each gender, the percentage of students taking the ANPEC exam increases when we go from the bottom to the top quartile (higher GPAs). Still, there is no gender gap in the likelihood of taking the exam within each quartile.

Nevertheless, our results indicate that women are less likely to rank among the top 100 students in the ANPEC exam and perform worse than men in Mathematics, Statistics, Macroeconomics, and Microeconomics.[1] Consequently, they are more likely to be accepted into lower- ranked master's programs, which can affect their academic career. According to the Brazilian Women in Economics research group,[2] there is only a small percentage of women in higher-ranked master's programs in Brazil. In 2019 (2020), only 30% (27%) of students in the most prestigious master's programs were women, compared to 48% (41%) in less prestigious programs.

[1] In the text, we use the full names, but in some tables, for space reasons, we use Stat, Macro, Micro, Math.

[2] Brazilian Women in Economics (www.usp.br/bwe) is a research group that studies the various dimensions of the gender gap in Brazil. It is specifically interested in understanding the relative lack of women in economics in the different academic careers and labor market stages. It also intends to promote higher participation of women in economics by increasing the connections between those who work in the private and public sectors and academia. The group conducts an annual survey of 52 departments with graduate programs in economics in Brazil.

Finally, using a linear differences-in-differences model, we show a decline in female performance in the ANPEC exam compared to their undergraduate grades. In contrast, male performance increases by 11% of a standard deviation.

We speculate that differences in performance can be mostly attributed to gender differences related to the evaluation setting.[2] Our findings suggest that the ANPEC exam is not gender neutral, as women perform similarly to men in most undergraduate disciplines comparable to the ones in the ANPEC exam, and therefore have similar learning capabilities and knowledge in economics.[3]

Our paper contributes to the literature examining grading schemes overall impact. While there are studies that compare teacher and test-based assessments for younger students at the basic educational level (Angelo and Reis, 2021; Falch and Naper, 2013; Marcenaro-Gutierrez and Vignoles, 2015), there is a lack of research examining gender differences associated with different grading schemes for higher education students and analyzing how they change across the academic path. As far as we know, only Jansson and Tyrefors (2022) point to grading bias as a possible early cause of the "leaky pipeline." Actually, grading schemes at universities have received little attention, exceptions being Breda and Ly (2015) and Feld et al. (2016). Furthermore, our study adds to the literature on gender differences in the performance of undergraduate students in economics (Rask and Tiefenthaler, 2008; Beneito et al., 2021).

Unfortunately, we cannot disentangle the reasons for female relative underperformance in the ANPEC exam, as it differs in distinct dimensions apart from the grading scheme, such as stakes, risk, competitiveness level, and anonymity. Previous evidence suggests that all these aspects can affect women and men differently.

Several studies find that women perform worse than men when stakes are high (Jurajda and Münich, 2011; Ors et al., 2013; Azmat et al., 2016; Cai et al., 2019; Borges et al., 2023). Risk aversion may also be a possible explanation behind grading gaps. Female test-takers skip significantly more questions than male test-takers in the quantitative track (Saygin and Atwater, 2021) and when questions are difficult (Riener and Wagner, 2017). Self-

[3] Rocha, Pereda, Matsunaga, Diaz, Narita and Borges (2021) present, but do not test, the hypothesis that the "leakage in the pipeline" of the economics career in Brazil seems to begin with this exam.

assessment seems also related to skipping behavior and could partially explain the gender gap (Saygin and Atwater, 2021). When tests include penalties for incorrect answers, as occurs in the ANPEC exam, the gender gap becomes particularly relevant, given the need to manage risk (Baldiga, 2014; Pekkarinen, 2015).

Regarding competition, Nierdele and Vesterlund (2007) show that women (men) seek to avoid (embrace) competitive situations and claim that this difference in behavior is due to lower female expectations about performance. Dohmen and Falk (2011), on the other hand, attribute this difference to gender-specific attitudes towards risk.

Finally, the comparison between blind and non-blind tests provides mixed results. Lavy (2008) finds evidence of a bias against boys, while Hinnerich et al. (2011) obtain evidence of no bias against boys. Specifically focusing on economics, Jansson and Tyrefors (2022) show that women benefit more from being graded anonymously, and when that happens, they are more likely to keep studying economics.

Lastly, it is important to acknowledge the limitations of our sample. To ensure the generalizability of our results, future studies should examine whether they apply to other universities, particularly those located in states with potentially higher levels of discriminatory social norms and gender stereotypes.

The remainder of this article is organized in five sections besides this introduction. Section 2 provides an overview of the ANPEC exam and the University of São Paulo undergraduate economics program. Section 3 discusses selection and gender differences in performance on undergraduate exams. Section 4 presents the methodology and the results on gender gaps comparing achievement on undergraduate exams and the ANPEC exam, and Section 5 brings the main conclusions.

2. Institutional background and Data

2.1. The ANPEC Exam

In Brazil, undergraduate students from economics who wish to pursue graduate studies are required to take the ANPEC exam. The exam is administered yearly, typically in September, and is composed by 6 different exams that cover Microeconomics, Macroeconomics, Mathematics, Statistics, Brazilian Economy, and English. All students take the same exams simultaneously, with each candidate receiving a standardized score (for each of the exams and an average score). The National Association of Postgraduate Programs in Economics then publishes a ranking of candidates based on their average standardized scores. All exams are equally weighted, except English, which no institution uses as an admission criterion.

However, each university can choose its own weights and ranking for students' selection. The "top 4" institutions – Fundação Getúlio Vargas São Paulo (FGV-EESP), Fundação Getúlio Vargas Rio de Janeiro (FGV-EPGE), Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO), and Universidade de São Paulo (IPE-USP) – consider only the core exams of Microeconomics, Macroeconomics, Mathematics, and Statistics, with each receiving the same weight.

Two aspects of the exam warrant special attention. The first is the strategic component of the exam, which involves mainly true-or-false questions with negative scores for incorrect answers. More precisely, incorrect answers yield negative points equivalent to the same points earned for a correct answer, while omissions do not receive any penalty and earn zero points. Open-ended questions comprise roughly 20% of the exam and are not subject to penalties. This scoring rule means that the number of omitted questions can affect a candidate's probability of admission.

The second is the admission procedure. When registering for the exam, candidates submit a list of up to six desired universities without an order of priority. After the exam results are released, universities have one week to contact candidates and make them an offer. Each applicant who receives an offer must complete a form indicating either "definite" or "conditional" acceptance. If the candidate chooses definite acceptance, the process ends, and he/she will no longer receive offers from other universities. If the can-

didate chooses conditional acceptance, he/she is temporarily accepting the university inviting him/her, but is also signaling that he/she would rather go to another university and is still hoping to be accepted. After this first round, the second round starts, and candidates can still choose between "definite" or "conditional" acceptance. In the third and final round, it is only possible to give a definite acceptance. Most of the top-ranked candidates aim for the most prestigious graduate programs, and the most prestigious programs also aim to be chosen by them. Along with the three rounds, universities advertise themselves, showcasing the qualities of their programs and the future career prospects they offer.

On average, 1,000 students take the ANPEC exam each year (see Appendix Figure F.1), making it highly competitive. The top 4 programs traditionally only accept the top 80 candidates. Table 1 shows that women are only 33% of the test-takers and, most importantly, are underrepresented in top positions. Only about 15% and 17% of the top 50 and top 100 highest-performing candidates are women, respectively. This skewed distribution of female candidates implies that women are underrepresented in the most prestigious economics graduate programs, as admission to most graduate programs is based only on performance in the ANPEC exam.[4] Therefore, the average grade in the undergraduate program or any extracurricular activities/academic curriculum will not directly impact graduate school admission.

Table 1 – Percentage of female candidates by ANPEC Ranking, 2004 to 2019

ANPEC Ranking	(%) Women
Top 50 in ANPEC Ranking	14.9
Top 100 in ANPEC Ranking	17.1
Top 200 in ANPEC Ranking	20.2
Total ANPEC Candidates	33.1

Notes: The total number of ANPEC exam candidates between 2004 and 2019, including retakers, was 16,985.

[4] Some institutions may require additional material when selecting students for their master's programs, such as letters of recommendation, transcripts, curriculum vitae, and research projects. However, performance in the ANPEC exam is the main criterion for the most prestigious research centers.

Table 2 presents data retrieved from Rocha, Diaz, Pereda, Narita and Facundes (2021), reporting women's representation in Brazilian undergraduate and graduate programs by CAPES score. CAPES is a Brazilian governmental agency that evaluates postgraduate courses every four years (master's and Ph.D.) on a scale ranging from 1 to 7 (7 being the highest score). Scores 1 and 2 imply course disqualification. A score of 3 is considered satisfactory, that is, the program meets the minimum quality standards. Score 4 indicates a good performance, while score 5 is the maximum score for programs that offer only the master's degree. Scores 6 and 7 define programs (master's and Ph.D) with a high standard. The evaluation is based on the quality of the publications by professors and students, as well as internationalization and research promotion.

Table 2 shows that the participation of women decreases as the universities' CAPES score increases. Universities with the lowest score have around 40% of women in their undergraduate and master programs, while in universities with the highest scores (6 and 7) these numbers reduce to 30% or less. It also indicates that the participation of women is smaller in master's programs than in undergraduate programs, regardless of the university's score – apart from universities that scored 3 in 2019. In 2019, there were 33.3% and 30.1% of women in undergraduate and master's programs 6 and 7, respectively. In 2020, these numbers were even smaller (28.9% in undergraduate courses to 27.3% in master's).

Table 2 - Female average participation in undergraduate and master's graduation by CAPES score, 2019 and 2020

	2019 Undergraduate	Master	2020 Undergraduate	Master
CAPES 3	42.3%	48.1%	42.6%	40.8%
CAPES 4	33.0%	26.6%	32.5%	31.1%
CAPES 5	34.0%	27.7%	38.7%	26.0%
CAPES 6 or 7	33.3%	30.1%	28.9%	27.3%

Notes: Data retrieved from Rocha, Diaz, Pereda, Narita and Facundes (2021). CAPES score is a proxy measure for the quality of the institution. CAPES is a Brazilian governmental agency that evaluates postgraduate courses every four years (master's and Ph.D.) on a scale ranging from 1 to 7 (7 being the highest score). Scores 1 and 2 imply course disqualification. A score of 3 is considered satisfactory, that is, the program meets the minimum quality standards. Score 4 implies a good performance, while score 5 is the maximum score for programs that offer only the master's degree. Scores 6 and 7 apply to programs with a high standard of performance for Ph.D. programs. The evaluation is based on data from the publications of the professors and students of the program, as well as internationalization and fostering research.

Our empirical analysis uses microdata for the 2004-2019 exams provided directly by ANPEC.[5] The available dataset contains information from 1997 to 2019. Still, we have limited our analysis to 2004 on because the oldest USP cohort in our sample (students admitted in the year 2000) start to graduate in 2003 (daytime students are expected to complete college in 4 years) and then apply for a master program in economics. The data provides information on the candidates' names,[6] gender, age, the year and institution they finished their undergraduate studies, how many times they took the exam, their performance on each test (Microeconomics, Macroeconomics, Mathematics, Statistics, Brazilian Economy, English), their university choice, and their rank in the official ANPEC Ranking.

2.2. Undergraduate students

The University of São Paulo (USP) is a large and prestigious public university in Brazil. According to the Times Higher Education Latin America University Rankings 2020, USP is the second-best university in the Latin America and Caribbean region, just behind the Pontifical Catholic University of Chile (PUC-Chile),[7] while in the SCIMAGO Institutions Rankings 2020,[8] USP occupies the 61st position among 500 international institutions. It is a research-intensive state university with ten campi in the state of São Paulo.[9] The main campus is in one of the wealthiest cities in the country, São Paulo. Our database is from the main campus, and from now on, we refer to USP when we mean the São Paulo city campus.

Entrance to the undergraduate programs at USP involves a rigorous and competitive process. During our period of analysis (2000 - 2010), the main admission system used by USP was a university-specific exam, named *vestibular*, more specifically FUVEST due to the foundation – *Fundação*

[5] Exams occur in the second semester of the year before admission. For example, the 2019 ANPEC exam was conducted on September 26th and 27th, 2018. Link: http://www.anpec.org.br/novosite/br/exame.

[6] Except for the year 2005.

[7] The ranking is based on the same 13 performance indicators as the Times Higher Education World University Rankings, but uses different weights to reflect the specific characteristics of the region's universities. The 2020 ranking includes 166 universities across 13 countries.

[8] Research and Innovation Rankings 2020 www.scimagoir.com, March 20, 2021.

[9] Public universities represented almost one-quarter of all tertiary students in Brazil in 2019. Federal, state, or municipal governments can fund public universities. Data available in https://download.inep.gov.br/educacao_superior/censo_superior/documentos/2020/Apresentacao_Censo_da_Educacao_Superior_2019.pdf

Universitária para o Vestibular - in charge of elaborating the questions, applying and correcting the exam. It is held once a year and it has two phases, the first in November/December and the second in January.[10]

The first phase involves only multiple-choice general knowledge questions, covering subjects like Portuguese, Mathematics, History, Geography, and Natural Sciences. Students who achieve a minimum score in this phase qualify for the second phase, which comprises more specific and challenging subjects. This minimum score is determined by all candidates' ranking in the first phase and the number of slots available for each specific study program (which the students choose when they register to take the FUVEST exam). For admittance to the economics, business, accounting, or actuarial sciences undergraduate programs in the São Paulo city campus in 2023, students needed a score of 63% in the first phase.[11] A total of 1,775 students were approved for the second phase, where they competed for 413 spots for the mentioned programs.

The second phase consists of detailed tests with open-ended questions, varying according to the chosen major program, and an essay on a given topic regarding actualities. Students who want to study economics must answer questions about geography, history, and mathematics. Scores from the first and second phases are used to rank the candidates. The ranking serves as the basis for admission. From 2009 on candidates applying for business, economics, accounting, or actuarial sciences undergraduate programs indicate their preferences among the four degrees.

The university then admits the top-ranked students and offers them a place in their first-choice program. The process continues by moving down the ranking list and offering admission to the next highest-scoring student until all available slots for a degree program are filled. Then, the student will be offered admittance to their second preferred program, and this process continues until slots in all programs are filled.

As a result, candidates who achieve higher scores in both the first and second phases have a better chance of being admitted to the program, as they will be given priority in the selection process. It's important to note that the number of available slots in the economics program is limited,

[10] Since 2015 the students interested in attending an undergraduate program at USP can also use for admission process their scores in a centralized standardized national exam called ENEM.

[11] Seehttps://jornal.usp.br/universidade/estude-na-usp/fuvest-divulga-notas-de-corte-do-vestibular-2023/

and the competition is usually intense. Therefore, students aiming to join USP economics program must obtain high scores in the FUVEST exam to increase their chances of being admitted.

Besides being renowned institutions in Brazil, public universities do not charge tuition or other fees. Therefore, they tend to be the preferred choice among high-school students, and their admission process is highly competitive (Borges and Estevan, 2023).

Unlike most American universities, the coursework structure in economics at USP is quite rigid, as about 57% of credits are in mandatory courses (Borges and Estevan, 2023). The undergraduate program is highly specialized in economics, so it does not have business-related disciplines as core courses. Besides, students must pass introductory disciplines to take more advanced ones, implying that students typically follow classes with their admission cohorts. If they fail one class, they will take longer to graduate. Students must also write a bachelor's thesis under a professor's supervision to complete their degree.

Daytime and evening students are split into two sections per shift (morning or evening classes). There are two classrooms in the morning and evening, with around 45 students attending parallel classes during the first semester. Conditional on the shift, the assignment criterion of students to each of the classrooms follows alphabetical order.[12] After the first semester, the students' composition in each classroom remains similar, but there is some change because some students fail the mandatory courses, and students can enroll in sections without following the initial classes' assignment. Later in the course, the classrooms are more mixed because most courses are not mandatory.

The grading scheme for undergraduate courses differs depending on the subject and the instructor in charge of the course. Despite that, grades in mandatory courses are mainly the average of two to three no-aids- allowed written assessments done in class (1 or 2 midterms and 1 final exam). Handing out different assignments during the semester may also compose the final grade specially of non-mandatory courses.

[12] Borges and Estevan (2023) explore the exogenous assignment of students to classes at the Department of Economics from USP to estimate female peers' effects and professors' effects on the labor market and educational outcomes.

The undergraduate grades were collected from the academic record system of undergraduate students at USP.[13] It contains information about the student's attendance, grades, result (approved or failed), FUVEST admission ranking, age, city, and state of residence when the student applied to USP.

Our sample consists of students that were admitted into the undergraduate economics program during the period 2000 to 2010. However we will focus on the subsample of students admitted to USP through the *vestibular*,[14] that completed the degree, and for which we have information on grades for the subgroup of subjects included in the ANPEC exam: 1,431 students (27% female students).

Since we are concerned with the relationship between the evaluation scheme and the leaky pipeline, we want to follow students exposed to the same higher education program. This way, after controlling for other observable characteristics, gender differences in performance can be mostly credited to the different evaluation systems. Therefore, our sample of interest is composed by undergraduate students from USP who took the ANPEC exam (296 students, 21% of the restricted sample).

Each student in USP administrative records is identified by a university ID number while the ANPEC microdata contains information on each candidate's name. To merge the two data sets we had to request the Department of Economics the ID numbers of FEA students that applied to the ANPEC exam.

For each student, we keep only her/his best (highest-ranking) result in the ANPEC exam if she/he took the exam more than one time. This procedure avoids both double-counting or capturing 'trainees' lower-effort results – a possible concern if selecting only first attempts.

We end up with a panel of students with two performance measures per student, undergraduate grades and scores on the ANPEC exam. We evaluate their performance in the undergraduate disciplines that make up the ANPEC exam: Mathematics, Statistics, Microeconomics, and Macroeconomics.

[13] The system is called JupiterWeb platform.

[14] Other possible admittance paths in our period of analysis involved exchange agreements with partner international universities and student transfer process from other programs and/or universities in Brazil. However, we only consider in our analysis students admitted specifically through the FUVEST exam.

Given USP's prestige and the selective nature of its admittance exam, its undergraduate students (both female and male) tend to be quite distinct from other students who take the ANPEC exam. Table 3 compares the performance of all students from USP who took the ANPEC exam and all applicants who took the exam (including and excluding USP students, broken down by gender) and examines their overall scores as well as their scores on the specific exams. To ensure comparability, we restrict the sample of ANPEC applicants to those who graduated with a degree in economics after 2002 (the same year our sample of USP students began graduating). For all samples, we find statistically significant gender differences in performance, with men outperforming women in all the exams except Mathematics, where no statistically significant difference is found among USP students (all students and for our sample of analysis – USP and ANPEC). Besides, female students who graduated from USP tend to outperform their counterparts (both female and male students) from other universities.

Table 4 presents additional descriptive statistics for our main sample of undergraduate students. In Panel A, we observe no significant difference in exam participation rates between female and male students. Approximately 21% of USP's students take the exam (81 out of 383 female students and 215 out of 1,048 male students). When we examine the subset of students who took the exam (296 students) in Panel B, we observe that USP students perform exceptionally well compared to other test takers. Over 70% of USP undergraduate students rank among the top 200 best candidates. However, when we compare the representation of female and male students among the top 50, 100, and 200 highest achievers in the ANPEC exam, we observe a decreasing proportion of women, suggesting a lack of gender balance among the high-performing test takers. Women account for 12.35%, 32.10%, and 69.14% of the top 50, 100, and 200 best candidates in the ANPEC exam, respectively. Although we observe women's under-representation among the top 50 and 100 highest-ranked applicants, there is no statistically significant difference in the percentage of female and male students among the top 200 positions.

Table 3 - Gender performance differences on the ANPEC exam (grades), all ANPEC applicants × all USP students

	All Economics			All Economics except USP			USP students			USP students (sample)		
	Women	Men	Diff.	Women	Men	Diff.	Women	Men	Diff.	Women	Men	Diff.
Total ANPEC grade	9.17	13.16	-3.99***	8.87	12.72	-3.85***	21.21	24.68	-3.47***	20.65	24.61	-3.96***
	(8.60)	(10.67)		(8.38)	(10.45)		(9.05)	(10.17)		(8.66)	(10.32)	
Microeconomics	2.38	3.36	-0.98***	2.32	3.25	-0.93***	5.18	6.28	-1.10***	5.02	6.33	-1.31***
	(2.45)	(2.94)		(2.40)	(2.88)		(2.54)	(2.90)		(2.41)	(2.99)	
Macroeconomics	2.89	4.11	-1.23***	2.82	4.02	-1.19***	5.57	6.68	-1.11***	5.38	6.58	-1.20***
	(2.57)	(3.04)		(2.53)	(3.00)		(2.72)	(2.80)		(2.62)	(2.76)	
Mathematics	1.61	2.38	-0.77***	1.53	2.27	-0.74***	4.70	5.19	-0.49	4.69	5.32	-0.63
	(2.13)	(2.64)		(2.06)	(2.56)		(2.72)	(3.01)		(2.64)	(3.06)	
Statistics	2.29	3.31	-1.02***	2.21	3.19	-0.98***	5.76	6.53	-0.77**	5.55	6.37	-0.82**
	(2.80)	(3.29)		(2.74)	(3.24)		(2.91)	(3.10)		(2.8)	(3.07)	
Observations	3,665	6,861	10,526	3,577	6,610	10,187	88	251	339	81	215	296

Notes: We present descriptive statistics (mean and standard deviation in parentheses) for ANPEC grades for each exam separately (Microeconomics, Macroeconomics, Mathematics, and Statistics) and the sum of all four exams (Total ANPEC grade). The group "All Economics" comprises all Economics students who took the exam between 2004 and 2019 and completed the course after 2002. The group "USP students" comprises the subsample of students who completed the economics course at USP. "All Economics except USP" represents all Economics applicants excluding the undergraduate students from USP. "USP students (sample)" displays the statistics for our main sample of analysis. Columns 'Women' display the statistics for female applicants, 'Men' for male applicants, and 'Difference' reports the coefficient of a t-test of mean differences between groups. P-values: $*p<0.1$; $**p<0.05$; $***p<0.01$.

Table 4 - Gender participation rates in the ANPEC exam (ranking), USP students

	Women	Men	Difference
Panel A: Total students from USP			
Percentage of Students that took ANPEC	21.15	20.52	0.63
	(40.89)	(40.40)	
Number of observations (total students)	383	1,048	1,431
Panel B: Students from USP who took the exam			
Percentage in Top 50 Ranking at ANPEC	12.35	33.02	-20.68***
	(33.10)	(47.14)	
Percentage in Top 100 Ranking at ANPEC	32.10	49.30	-17.20***
	(46.98)	(50.11)	
Percentage in Top 200 Ranking at ANPEC	69.14	74.42	-5.28
	(46.48)	(43.73)	
Number of observations (students that took the exam)	81	215	296

Notes: Panel B's sample is restricted to students from USP who took the ANPEC exam and were admitted from 2000 to 2010 via the FUVEST exam and concluded the course. We calculate the percentages in top positions conditional on the students that took the exam. Columns 'Women' display the statistics for female applicants,'Men' for male applicants, and 'Difference' reports the coefficient of a t-test of mean differences between groups. P-values: $*p<0.1$; $**p<0.05$; $***p<0.01$.

3. Gender differences in perfomance

3.1. Selection before undergrad: the FUVEST exam

We first analyze if there is any statistically significant gender difference in the FUVEST admission exam and in undergraduate core courses. Table 5 brings the average FUVEST ranking and average grades at the undergrad level for the full sample of economics students at USP and our USP/ANPEC subset of students. Notably, we do not find any statistically significant difference in the FUVEST ranking for men and women for the two samples. This indicates no significant pre-existing differences in performance before admission to the economics bachelor program.

However, we find that men have lower average grades in compulsory courses, though the difference is only significant at a marginal level for the ANPEC subset. This is suggestive that, despite no observed differences before admission, women tend to outperform men in undergraduate courses. Nonetheless, we must interpret these numbers carefully, as we are not controlling for other observable characteristics that might affect the results.

Table 5 - Gender differences before and during undergrad

	Full sample			ANPEC sample		
	Women	Men	Diff.	Women	Men	Diff.
FUVEST Ranking	114.29	109.57	4.72	103.44	96.81	6.63
	(73.34)	(73.96)		(60.47)	(77.75)	
Avg grade compulsory courses	6.62	6.45	0.17***	7.17	6.94	0.24*
	(0.92)	(0.99)		(0.74)	(1.00)	
Observations	383	1,048	1,431	81	215	296

Notes: A lower FUVEST ranking means a better placement at the USP admission exam.

3.2. Academic performance during college

To further examine the academic performance of female students in the undergraduate program at USP in more depth, we build a panel of student-course performance over time for our restricted sample of USP/ANPEC

students. We then evaluate gender differences in course grades and pass rates among undergraduate economics students, including classroom fixed effects which absorb potential observable and unobservable differences.

We estimate the following regression model:

$$y_{icst} = \beta_0 + \beta_1 women_i + \delta_{cst} + \gamma_{t*} + \theta' X_i + \epsilon_{icst} \quad (1)$$

Where y_{icst} is one of our two performance measures: either the normalized final course grades (mean zero and standard deviation of one) or pass rates which are a binary variable equal to 1 if the student passed and 0 if the student failed. These measures are obtained by student i in year t (admitted in $t*$) in the subject s in the classroom c. The variable $women_i$ is a dummy variable equal to 1 if the student is female and 0 if the student is male. The parameters δ_{cst} denotes the subject-classroom-semester fixed effect, corresponding to the section (daytime or evening) where student i took discipline s in semester/year t. The parameter $\gamma t*$ represents the cohort fixed effect based on students' admission year. The matrix X_i contains control variables for student i, including age, ranking in the FUVEST exam, and a dummy variable indicating whether the student's city of origin is São Paulo. In this equation, each student may appear multiple times, corresponding to the number of undergraduate courses she/he took in each group of subjects.

Table 6 displays the regression results from estimating equation (1) using ordinary least squares (OLS). Columns (1) to (4) show the results for the average normalized undergraduate grades in the courses that make up the ANPEC exam, while columns (5) to (8) report the average pass rates in the same group of courses. The results indicate that female students perform as well as or better than male students in the mandatory disciplines at USP. Women also tend to have higher pass rates in Microeconomics and Statistics, and similar pass rates in Macroeconomics and Mathematics.[15]

[15] We also analyze the total distribution of grades for men and women in our sample. Figure F.2 describes performance distribution for the subsample of students who took the ANPEC exam, using the grading systems of USP mandatory courses. We observe that the distribution of women's grades is less spread out than men's.

Table 6 - Gender differences in final grade and approval rate, mandatory subjects per category

	Normalized course grade				Pass rate			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Macro	Micro	Math	Stat	Macro	Micro	Math	Stat
Women	-0.048	0.049	0.103	0.110**	0.014	0.043**	-0.010	0.035**
	(0.060)	(0.058)	(0.074)	(0.054)	(0.012)	(0.017)	(0.014)	(0.014)
Number of observations	1,046	1,040	681	1,587	1,046	1,040	681	1,587
Admission year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Classroom FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Our sample consists of a panel of undergraduate students' grades from 2000 to 2016. These students were admitted via the FUVEST exam between 2000 and 2010, completed their courses, and took the ANPEC exam. Columns (1)-(4) present the OLS regression estimates for the normalized course grade (within classroom) as the dependent variable, while columns (5)-(8) present the linear probability models for the approval rates. All regressions include the complete specification with admission year fixed effects, classroom fixed effects, and all control variables. The classification of the subjects is shown in Appendix Table F.1. Standard errors are clustered at the classroom level. P-values: *p<0.1; **p<0.05; ***p<0.01.

3.3. Selection into the ANPEC exam

To better understand the selection of students who take the ANPEC exam we analyze the characteristics within gender groups. Table 7 compares some characteristics of the students that took and that did not take the ANPEC exam. The results show that, on average, students that took the exam have higher grades. Specifically, women have grades 0.7 higher, while men have grades 0.61 higher. There are also differences in graduation times with women and men who took the ANPEC exam graduating, on average, six months earlier than those who did not take the exam. Finally, both female and male students that take the ANPEC exam are generally younger.

Table 7 – Descriptive statistics of USP students by gender, took ANPEC × did not take the exam

	Women			Men		
	Took ANPEC	Did not take ANPEC	Diff.	Took ANPEC	Did not take ANPEC	Diff.
Age at the time of admission (years)	18.20	18.59	-0.40*	18.59	18.96	-0.38**
	(1.08)	(1.77)		(1.83)	(2.35)	
FUVEST (USP admission exam) Ranking	103.44	117.20	-13.75	96.81	112.86	-16.05***
	(60.47)	(76.25)		(77.75)	(72.63)	
% students from Sao Paulo (State)	100.00	98.68	1.32	97.67	98.44	-0.76
	(0.00)	(11.45)		(15.11)	(12.40)	
% students from Sao Paulo (City)	86.42	86.09	0.33	77.21	82.35	-5.14*
	(34.47)	(34.66)		(42.05)	(38.14)	
Time to graduate (in years)	3.94	4.47	-0.53***	4.11	4.75	-0.63***
	(0.62)	(1.13)		(0.94)	(1.24)	
Avg grade compulsory courses	7.17	6.47	0.70***	6.94	6.33	0.61***
	(0.74)	(0.91)		(1.00)	(0.94)	
Observations	81	302	383	215	833	1,048

Notes: The sample is restricted to students admitted to USP between 2000 and 2010 via the FUVEST exam who completed the course. The 'Women' columns display the statistics for female applicants, 'Men' for male applicants, and 'Diff.' reports the coefficients of a t-test of mean differences between groups. The 'Took ANPEC' columns report the results from a sample restricted to students who took the ANPEC exam. The Did not take ANPEC columns report the results from a sample of students who did not take the ANPEC exam. P-values: *p<0.1; **p<0.05; ***p<0.01.

To examine if students' characteristics affect their likelihood of taking the ANPEC exam, we estimate a probit model where the binary response variable indicates whether the student took the ANPEC exam or not. This analysis encompasses all undergraduate students admitted between 2000 and 2010 via the FUVEST exam who successfully completed the course.

It is worth mentioning that the explanatory variables in this regression are potentially endogenous. For example, students who undertook an internship while at college probably took longer to graduate and had no interest in taking the exam because they probably received job offers before graduating. We should, therefore, interpret the coefficients with caution, understanding them as correlations and not causal effects.

Table 8 reports the average marginal effects. Our results indicate that there are no gender differences in the probability of taking the ANPEC exam. Age and time to complete the course also do not appear to be significant factors. However, higher grades in compulsory courses are positively

related to the probability of students taking the ANPEC exam, indicating a selection of students interested in pursuing a master's program. Besides, students with better academic performance are more likely to take the ANPEC exam regardless of their gender (as shown in Table F.2).

Table 8 – Gender differences in the probability to take the ANPEC exam, USP students

	(1)	(2)	(3)	(4)
Dependent variable: Took ANPEC				
Women=1	0.010	-0.010	-0.013	-0.013
	(0.024)	(0.022)	(0.022)	(0.022)
Avg grade compulsory courses		0.112***	0.110***	0.105***
		(0.011)	(0.011)	(0.012)
Age (at admission)			-0.009	-0.009
			(0.007)	(0.007)
Took more than 5 years to graduate=1				-0.034
				(0.035)
Number of observations	1,431	1,431	1,431	1,431
Year FE	Yes	Yes	Yes	Yes

Notes: The reported results are average marginal effects from probit regressions conducted on a sample of undergraduate students admitted between 2000 and 2010 via FUVEST exam, who successfully completed the course. All regressions include fixed effects for the admission year at USP, as well as a binary gender variable as controls. In column (2), we add the average grade in mandatory courses as a control variable, in column (3) we add age, and in column (4) we add a binary variable indicating if the student took more than 5 years to graduate. P-values: *p<0.1; **p<0.05; ***p<0.01.

To further analyze if high-performing men and women differently select into the ANPEC exam, in Figure 1 we analyze the share of students taking the exam looking at the gender-specific GPA distribution. Within each gender, the share of students taking the ANPEC exam increases when we go from the bottom to the top quartile. Within each quartile, there are no gender differences in the share of candidates (as we observe in Table F.3). For instance, the share of male and female students in the highest quartile of the GPA distribution (Q4) is 36% and 41%, respectively, but this difference is not statistically significant.[16]

4. Grading systems and the gender gap

Having analyzed the performance during undergraduate courses and the selection into USP via FUVEST and into the ANPEC exam, we now turn

[16] In Table F.4, we report the coefficients of Probit regressions where we interact a dummy that identifies the top 25% students in the GPA distribution (both genders) with a female dummy. We verify that high-achieving students are much more likely to pursue graduate studies regardless of their gender.

to the main exercise proposed in this paper. To make comparison easier, we standardize the grades in USP core courses and the ANPEC exam, giving them a mean of zero and a standard deviation of one.

Table 9 compares the average performance of undergraduate students in the mandatory courses at college and their performance in the ANPEC exam.[17] For the same group of students, in columns 'USP,' we compare gender differences in performance in the undergraduate core courses, while in columns 'ANPEC,' we verify gender gaps in performance in the ANPEC exam. Finally, in columns 'ANPEC-USP,' we present the gender comparisons for the individual difference in performance between the two exams: ANPEC minus USP standardized scores.

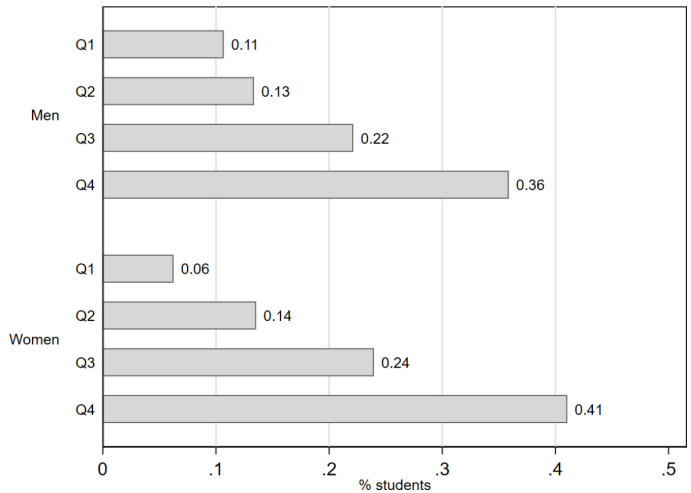

Figure 1 - Share of ANPEC applicants by within-gender GPA quartiles

The difference in the performance of male and female students is not statistically different for most undergraduate courses, although females have higher averages. One exception is Math where, with a 5% significance level, we find evidence that women outperform men with a 26% higher grade in standard deviation terms. Interestingly, when we compare the grades for the ANPEC exam, we find robust evidence that men outperform women in almost all subjects. For example, men score around 0.31 standard deviations higher in Microeconomics and 0.36 standard deviations in Macroeconomics. When we look at the difference in performance

[17]Appendix Table F.1 presents the complete list of undergraduate courses in each category.

between the two evaluation systems, we observe that while relative male performance increases in the ANPEC exam, the opposite happens for women, leading to statistically significant differences in all subjects. These findings suggest that there are indeed gender differences in performance related to grading schemes.

4.1. Empirical strategy

Since we can match the same student's grades in undergraduate courses and the ANPEC exam, we follow the methodology of Falch and Naper (2013) and estimate a linear differences- in-differences model to examine the gender gaps between teacher grading and scores on the graduate entrance exam.

Table 9 - Mean comparison tests by gender and evaluation setting, USP and ANPEC grades

	USP			ANPEC			ANPEC-USP		
	Women	Men	Diff.	Women	Men	Diff.	Women	Men	Diff.
Microeconomics	0.05	-0.02	0.07	-0.23	0.09	-0.31**	-0.28	0.10	-0.38***
	(0.81)	(1.04)		(0.87)	(1.00)		(1.08)	(1.04)	
Macroeconomics	0.03	-0.01	0.04	-0.26	0.10	-0.36***	-0.29	0.11	-0.40***
	(0.74)	(1.06)		(0.94)	(0.97)		(1.12)	(1.12)	
Statistics	0.07	-0.03	0.10	-0.14	0.05	-0.20	-0.22	0.08	-0.30**
	(0.84)	(1.03)		(0.95)	(0.98)		(1.15)	(1.03)	
Mathematics	0.19	-0.07	0.26**	-0.15	0.06	-0.21*	-0.34	0.13	-0.47***
	(0.93)	(1.00)		(0.86)	(1.01)		(1.21)	(1.09)	
Observations	81	215	296	81	215	296	81	215	296

Notes: The sample is restricted to students from USP admitted via FUVEST between 2000 and 2010 who completed the course and took the ANPEC exam. The 'Women' columns display the statistics for female applicants, 'Men' for male applicants, and 'Diff.' reports the estimated coefficients of a t-test of mean differences between groups. The 'USP' columns present the standardized grades in undergraduate courses. The 'ANPEC' columns present the standardized grades in the ANPEC exam. The 'ANPEC-USP' columns report the difference between the standardized grade in the ANPEC exam and USP courses. P-values: *p<0.1; **p<0.05; ***p<0.01.

Usually, in a differences-in-differences methodology, we compare results before and after the event of a natural experiment for the treatment and control groups. Here we observe men and women in two different exams and instants. We can consider gender as our treatment (with women as the treated group and men as the control group, for example), and "before" being students' performance throughout their undergraduate studies at

USP and "after" being performance in the ANPEC exam. However, it's important to note that the main variation we will analyze is in the exam format, not the timing of each assessment.

We estimate the following regression model:

$$y_{ie} = \delta_0 + \delta_1 women_i + \delta_2 ANPEC_e + \delta_3 (women_i \times ANPEC_e) + \gamma_{t*} + \epsilon_{ie} \quad (2)$$

Where y_{ie} is the standardized performance of student i in exam type e (USP mandatory courses or ANPEC exam). $women_i$ is a gender dummy that equals 1 if the student is a woman. $ANPEC_e$ is a dummy variable that indicates if the performance measure is from the ANPEC exam (=1) or USP (=0), γ_{t*} are cohort fixed effects (admission-year dummies), and ϵ_{ie} is the idiosyncratic error term. In all specifications, we cluster the standard errors at the individual level.

We estimate separate OLS regressions for each subject tested in ANPEC: Microeconomics, Macroeconomics, Statistics, and Mathematics. We also present a specification that pools all four subjects to evaluate the overall gender gap. Since we combine ANPEC and USP grades, we have two observations per student for each subject and end up with 592 observations. In the regression that includes all subjects, we have eight observations per student (four subjects × two evaluation schemes) and end up with a total of 2,368 observations.

The coefficient δ_1 captures gender differences in overall performance, and δ_2 displays differences in performance in ANPEC grades relative to USP grades not related to gender. δ_3 represents the differences-in-differences coefficient that identifies the gender gap in performance in the high-stakes ANPEC exam. A positive/negative δ_3 indicates that female performance relatively increases/decreases in the ANPEC exam compared to the grades obtained in the mandatory courses of the undergraduate program at USP.

4.2. Results

Table 10 presents the results for the differences-in-differences regressions. In column (1), we report the combined results for all ANPEC exams. In the subsequent columns, we present the results for each ANPEC exam separately: Microeconomics (2); Macroeconomics (3); Statistics (4); Mathematics (5).

In the first column (All) we find no gender differences in performance for the whole set of subjects evaluated in the ANPEC exam. However, we observe a decline in female performance in the ANPEC exam compared to their performance in undergraduate courses, while male performance increases by 11% of a standard deviation. Specifically, female performance drops by 28% (-0.387 + 0.106) of a standard deviation. In Microeconomics and Macroeconomics there are no statistically significant differences for men, while women decrease their performance in the ANPEC exam by around 40% of a standard deviation. In Statistics, we observe a similar pattern with men presenting no important changes in their performance and women decreasing their performance in the ANPEC exam by around 30%. In Mathematics, women have better grades in undergraduate courses (23% of a standard deviation larger). Nevertheless, in the ANPEC exam their relative performance declines by 34% of a standard deviation (-0.473 + 0.130) while male grades improve by 13% of a standard deviation.

Table 10 – Gender gaps in the ANPEC exam in comparison with undergraduate courses, USP students

	All (1)	Micro (2)	Macro (3)	Stat (4)	Math (5)
Women × ANPEC	-0.387***	-0.382***	-0.397***	-0.297**	-0.473***
	(0.122)	(0.141)	(0.147)	(0.147)	(0.155)
ANPEC	0.106*	0.104	0.109	0.081	0.130*
	(0.059)	(0.072)	(0.077)	(0.071)	(0.075)
Women	0.091	0.049	0.001	0.081	0.234*
	(0.103)	(0.118)	(0.115)	(0.122)	(0.127)
Number of observations	2,368	592	592	592	592
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: The sample includes students who were admitted via FUVEST from 2000 to 2010, completed the course, and took the ANPEC exam. The dependent variable is the standardized grade (mean zero, standard deviation one) in ANPEC and USP. In all regressions, we control for a gender variable, a binary variable that equals 1 for the ANPEC exam and zero for USP courses, and their interaction. We also control for cohort fixed effects in all regressions. P-values: *p<0.1; **p<0.05; ***p<0.01. Standard errors clustered at the individual level are presented in parentheses.

Table F.5 in the appendix provides additional evidence on women's performance in the ANPEC exam. It displays the results of OLS and quantile regressions where the dependent variable is ANPEC scores. The results corroborate that women perform worse than men on the ANPEC exam and, on average, obtain grades that are 4.75 points lower than men. Each exam (Mathematics, Statistics, Macroeconomics, and Microeconomics) is worth 15 points, and therefore the exam as a whole sum 60 points.

The mean of the scores during the period of analysis is 23.53, so the gender difference is substantial. Results from the quantile regressions (percentiles 10, 25, 50, 75, and 90) show that women perform worse than men across the entire distribution of grades. The gender gap increases in absolute terms and reaches its highest point in the last decile, where women's scores are 5.48 points lower than men's. However, when we compare the size of the coefficient to its statistics (average and each percentile), we note that the gender gap decreases along the distribution. For instance, -3.41 represents 36% of the ANPEC grade at the 10th percentile, while -5.48 corresponds to 15% at the 90th percentile.

Table 11 reports the average marginal effects from probit regressions, where the response variable is the probability of achieving a top 100 position in the ANPEC exam. Column (1) reports the raw gender gap, while in columns (2) to (5) we gradually include explanatory variables. The inclusion of covariates does not substantially affect the coefficients. Once again, we should be cautious, as we have correlations, not causal impacts. Consistent with our previous findings, holding other factors constant, female students from USP are less likely than their male counterparts to attain a top 100 position. Consequently, they will engage in lower-ranked graduate programs jeopardizing their academic careers.

Table 11 – Gender differences in the probability of achieving the ANPEC Top 100, 2004 to 2019

	(1)	(2)	(3)	(4)	(5)
I (achieving Top 100 in ANPEC rankings)					
Women=1	- 0.211***	-0.231***	-0.238***	-0.225***	-0.230***
	(0.058)	(0.056)	(0.056)	(0.056)	(0.055)
Avg grade compulsory courses		0.154***	0.145***	0.127***	0.100***
		(0.026)	(0.027)	(0.028)	(0.032)
Age (at admission)			-0.035*	-0.036*	-0.037*
			(0.019)	(0.020)	(0.020)
FUVEST (USP admission exam) Ranking				-0.001**	-0.001**
				(0.000)	(0.000)
Took more than 5 years to graduate=1					-0.288**
					(0.118)
Number of observations	296	296	296	296	296
Year FE	Yes	Yes	Yes	Yes	Yes

Notes: The reported results are average marginal effects from probit regressions, based on a sample of undergraduate students in economics from USP who were admitted through FUVEST, completed the economics majors, and took the ANPEC exam from 2014 to 2019. The dependent variable is a binary variable that indicates whether the student achieved a top 100 ranking in the ANPEC exam. In all regressions, we control for age at admission, FUVEST ranking, a binary variable indicating whether the student took more than five years to complete the undergraduate program, the average grade in compulsory courses during the undergraduate studies, and fixed effects for the year of the ANPEC exam. P-values: *p<0.1; **p<0.05; ***p<0.01.

5. Conclusions

A growing and important body of research has been discussing the barriers women face at the various stages of their academic career in economics and how they become underrepresented (the "leaky pipeline"). The purpose of this paper is to contribute to this literature by comparing the performance of male and female students from when they enter the university till they apply to a master's program.

We follow students who received a degree in economics from the University of São Paulo and compare their performance in the main undergraduate courses and the national admissions exam for graduate programs in economics. As they went to the same university, we expect differences in performance to be driven mainly by gender differences due to the evaluation schemes.

The Brazilian context is particularly interesting for examining the "leaky pipeline." Since students apply to universities and majors simultaneously, their performance on exams does not lead to early sorting into fields. Moreover, the main admission criteria for a graduate program in economics is the score in the ANPEC exam, the national unified admission exam. Therefore, students' undergrad grades or any extracurricular activities/ academic curriculum do not directly impact their admission to a master's program.

We provide evidence that there are no significant differences between men and women in FUVEST, the admission exam to USP. Besides, female students perform as well as, or even better, than male students in the mandatory undergrad courses (Microeconomics, Macroeconomics, Mathematics, Statistics) and there is no selection into the ANPEC exam. Finally, we use a differences-in-differences approach to compare USP and ANPEC standardized grades. In all subjects evaluated in the ANPEC exam, we find that women's grades decrease by 28% of a standard deviation, and men's grades increase by 11%.

We cannot disentangle the source of the gender gap, as undergraduate courses' evaluation schemes are very different from the ANPEC exam in terms of stakes, risk, competitiveness, all of which could potentially affect women and men differently. Therefore, further studies are necessary to understand which factors are important and how they influence women's confidence and self-esteem.

Our study, however, shows that evaluation schemes can have consequences for the leaky pipeline in economics. Besides affecting the probability of continuing with economics at the undergraduate level, as shown by Jansson and Tyrefors (2022), it can also affect the likelihood of continuing in economics at the graduate level, and especially the probability of attending a prestigious master's program, influencing the decision to pursue an academic career path. As Claudia Goldin notes: "we're teaching economics the same way we did when women didn't matter. But now women do matter. So how do we translate economics into 'girlish'?". According to the findings from our study, we must translate exams into 'girlish' as well.

There are several paths for future research. First, a natural extension is to verify if the result remains the same if we look at different universities. Second, it would be interesting to compare female and male candidates' performance in ANPEC's open questions. Since there is no penalty for open questions, this would allow us to verify if risk indeed affects women's decision-making on how to answer exams, as predicted by the existing literature.

References

Angelo, Catarina and Ana Balcão Reis. 2021. "Gender Gaps in Different Grading Systems." *Education Economics* 29 (1): 105–119.

Avilova, Tatyana and Claudia Goldin. 2018. "What Can UWE Do for Economics?" *AEA Papers and Proceedings* 108: 186–90.

Azmat, Ghazala, Caterina Calsamiglia, and Nagore Iriberri. 2016. "Gender Differences in Response to Big Stakes." *Journal of the European Economic Association* 14 (6): 1372–1400.

Baldiga, Katherine. 2014. "Gender Differences in Willingness to Guess." *Management Science* 60 (2): 434–448.

Beneito, Pilar, José E. Boscá, Javier Ferri, and Manu García. 2021. "Gender Imbalances across Subfields in Economics: When Does it Start?" *Journal of Human Capital* 15 (3): 469-511.

Borges, Bruna, and Fernanda Estevan. 2023. "Does Exposure to More Women in Male-Dominated Fields Render Female Students More Career-Oriented?" SSRN working paper 3789018.

Borges, Bruna, Fernanda Estevan, and Louis-Philippe Morin. 2023. "Gender Differences in Prioritizing Rewarding Tasks." SSRN working paper no. 4354209.

Boudreau, Kevin, and Nilam Kaushik. 2022. "Gender Differences in Response to Competitive Organization? Differences Across Fields from a Product Development Platform Field Experiment." Working Paper 30062, National Bureau of Economic Research.

Breda, Thomas, and Son Thierry Ly. 2015. "Professors in Core Science Fields Are Not Always Biased Against Women: Evidence from France." *American Economic Journal: Applied Economics* 7(4): 53–75.

Cai, Xiqian, Yi Lu, Jessica Pan, and Songfa Zhang. 2019. "Gender Gap Under Pressure: Evidence from China's National College Entrance Examination." *Review of Economics and Statistics* 101(2): 249–263.

Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2014. "Women in Academic Science: A Changing Landscape." *Psychological Science in the Public Interest* 15 (3): 75-141.

Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams. 2015. "Women in science: The path to progress." *Scientific American Mind* 26(1), 62–69.

Dohmen, Thomas, and Armin Falk. 2011. "Performance Pay and Multidimensional Sorting: Productivity, Preferences, and Gender." *American Economic Review* 101 (2): 556–90.

Estevan, Fernanda and Kelly Santos. 2022. "Does It Matter Which Institution You Choose? A Case Study of Brazilian Graduate Admissions." SSRN working paper no. 4225787.

Falch, Torberg, and Linn Renée Naper. 2013. "Educational Evaluation Schemes and Gender Gaps in Student Achievement." *Economics of Education Review* 36: 12–25.

Feld, Jan, Nicolás Salamanca, and Daniel S. Hamermesh. 2016. "Endophilia or Exophobia: Beyond Discrimination." *The Economic Journal* 126, no. 594: 1503–1527.

Goldin, Claudia. 2013. "Notes on Women and the Undergraduate Major." *CSWEP Newsletter* 15: 4–6.

Hengel, Erin. 2022. "Publishing While Female: Are Women Held to Higher Standards? Evidence from Peer Review." *The Economic Journal* 132, no. 648: 2951–2991.

Hinnerich, Björn Tyrefors, Eric Höglin, and Magnus Johannesson. 2011. "Are Boys Discriminated in Swedish High Schools?" *Economics of Education Review* 30, no. 4: 682-690.

Jansson, Joakim, and Björn Tyrefors. 2022. "Grading Bias and the Leaky Pipeline in Economics: Evidence from Stockholm University." *Labour Economics* 78: 102212.

Jurajda, Štěpán, and Daniel Münich. 2011. "Gender Gap in Performance under Competitive Pressure: Admission to Czech Universities." *American Economic Review* 101 (3): 514–18

Lavy, Victor. 2008. "Do Gender Stereotypes Reduce Girls' or Boys' Human Capital Outcomes? Evidence from a Natural Experiment." *Journal of Public Economics* 92 (10-11): 2083-2105.

Marcenaro-Gutierrez, Oscar, and Anna Vignoles. 2015. "A Comparison of Teacher and Test-Based Assessment for Spanish Primary and Secondary Students." *Educational Research* 57 (1): 1–21.

Nierdele, Muriel, and Lise Vesterlund. 2007. "Do Women Shy Away from Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122 (3): 1067-1101.

Ors, Evren, Frédéric Palomino, and Eloïc Peyrache. 2013. "Performance Gender Gap: Does Competition Matter?" *Journal of Labor Economics* 31 (3): 443-499.

Pekkarinen, Tuomas. 2015. "Gender Differences in Behaviour under Competitive Pressure: Evidence on Omission Patterns in University Entrance Examinations." *Journal of Economic Behavior & Organization* 115: 94-110.

Rask, Kevin, and Jill Tiefenthaler. 2008. "The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics." *Economics of Education Review* 27 (6): 676-687.

Riener, Gerhard, and Valentin Wagner. 2017. "Shying Away from Demanding Tasks? Experimental Evidence on Gender Differences in Answering Multiple-Choice Questions." *Economics of Education Review* 59: 43-62.

Rocha, Fabiana, Maria Dolores Montoya Diaz, Paula Pereda, Renata Narita, and Gabriel Facundes Monteiro. 2021. "As mulheres nos diferentes estágios da carreira acadêmica em economia no Brasil." Technical report, EconomistAs - Brazilian Women in Economics.

Rocha, Fabiana, Paula Pereda, Liz Matsunaga, Maria Dolores Montoya Diaz, Renata Narita, and Bruna P. Borges. 2021. "Gender Differences in the Academic Career of Economics in Brazil." *Cuadernos de Economía* 40 (84): 815–54.

Sarsons, Heather. 2017. "Recognition for Group Work: Gender Differences in Academia." *American Economic Review* 107(5): 141-45.

Saygin, Perihan O., and Ann Atwater. 2021. "Gender Differences in Leaving Questions Blank on High-Stakes Standardized Tests." *Economics of Education Review* 84: 102162.

Appendix

Figures

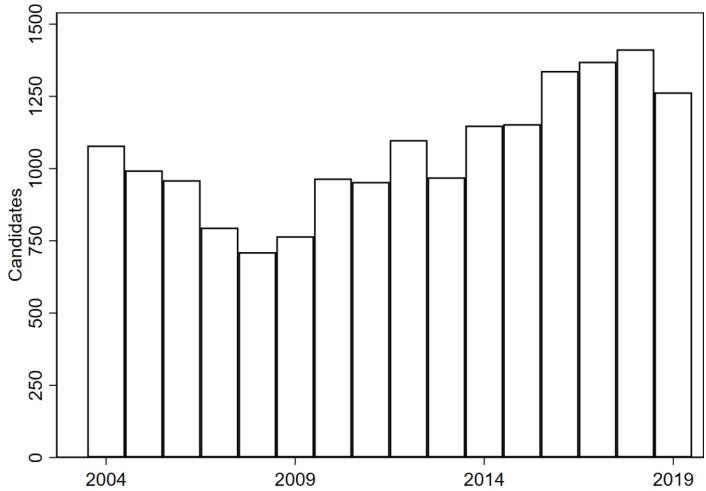

Figure F.1 - Total ANPEC candidates by year, 2004-2019

Source: ANPEC microdata.

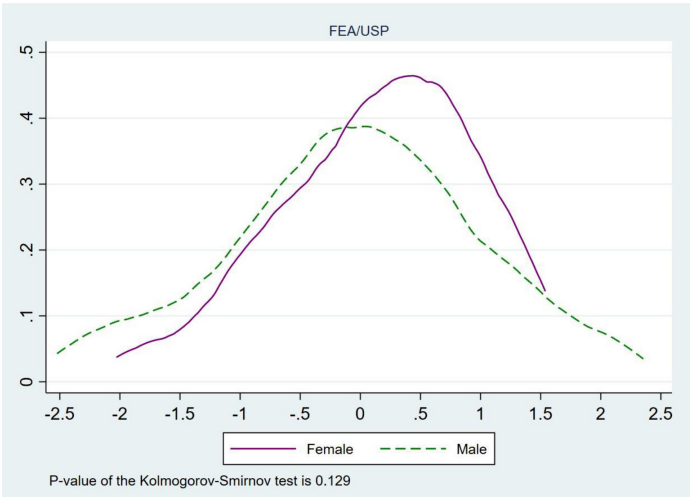

Figure F.2 – Normalized average grade distribution at USP by gender, 2000-2016

Sources: FEA-USP administrative datasets. We report the normalized average grade in compulsory courses by gender.

Tables

Table F.1 – Classification of Economics courses into categories at USP

Category	Course name
Econometrics	Econometrics I
Econometrics	Econometrics II
Econometrics	Econometrics III
Econometrics	Introduction to Probability and Statistics I
Econometrics	Introduction to Probability and Statistics II
Macro	Social Accounting and Balance of Payments
Macro	Macroeconomics I
Macro	Macroeconomics II
Macro	Accounting and Balance Analysis
Math	Calculus I
Math	Calculus II
Math	Calculus I for Economics
Math	Calculus II for Economics
Math	Math Applied to Economics
Micro	Microeconomics I
Micro	Microeconomics II
Micro	Economics of the Public Sector
Micro	International Economics I

Table F.2 – Probability to take the ANPEC exam, USP students

	(1)	(2)	(3)
Women			
Avg grade compulsory courses	0.140***	0.136***	0.122***
	(0.022)	(0.022)	(0.025)
Age (at admission)		-0.009	-0.008
		(0.012)	(0.013)
Took more than 5 years to graduate=1			-0.129*
			(0.071)
Number of observations	383	383	383
Mean dependent variable	0.21	0.21	0.21
Year FE	Yes	Yes	Yes
Men			
Avg grade compulsory courses	0.104***	0.102***	0.099***
	(0.012)	(0.012)	(0.014)
Age (at admission)		-0.008	-0.008
		(0.007)	(0.007)
Took more than 5 years to graduate=1			-0.016
			(0.039)
Number of observations	1,048	1,048	1,048
Mean dependent variable	0.21	0.21	0.21
Year FE	Yes	Yes	

Notes: The reported results are average marginal effects from probit regressions. The sample is a panel of undergraduate students from 2000 to 2016. These students were admitted from 2000 to 2012 via FUVEST and concluded the course. All regressions include admission year at USP fixed effects and a gender dummy. In column (1) we add average grade in mandatory courses, in column (2) age and in column (3) a dummy indicating if the student took more than 5 years to graduate. P-values: *p<0.1; **p<0.05; ***p<0.01.

Table F.3 – Share of ANPEC applicants by within-gender GPA quartiles

	Q1			Q2			Q3			Q4		
	Women	Men	Diff	Women	Men	Diff	Women	Men	Diff	Women	Men	Diff
Took the ANPEC exam	0.06	0.11	-0.04	0.14	0.13	0.00	0.24	0.22	0.02	0.41	0.36	0.05
	(0.24)	(0.31)		(0.34)	(0.34)		(0.43)	(0.42)		(0.49)	(0.48)	
Observations	96	262	358	96	262	358	96	262	358	95	262	357

Notes: The sample is restricted to students from USP admitted via FUVEST between 2000 and 2010 who completed the course. The 'Women' columns display the statistics for female applicants, 'Men' for male applicants, and 'Diff.' reports the estimated coefficients of a test of mean differences between groups. Q1, Q2, Q3 and Q4 represent the within-gender GPA distribution, considering the average grade in the compulsory courses of the undergraduate program from University of São Paulo. P-values: *p<0.1; **p<0.05; ***p<0.01.

Table F.4 – Probability to take the ANPEC exam, High achieving USP students

	(1)	(2)	(3)
Dependent variable: Took ANPEC			
Women=1	-0.070	-0.080	-0.107
	(0.111)	(0.111)	(0.111)
Top 25% FEA=1	0.642***	0.633***	0.547***
	(0.100)	(0.099)	(0.101)
Women=1 × Top 25% FEA=1	0.132	0.116	0.139
	(0.185)	(0.185)	(0.186)
Age (at admission)		-0.047*	-0.043*
		(0.025)	(0.025)
Took more than 5 years to graduate=1			-0.523***
			(0.133)
Number of observations	1,431	1,431	1,431
Year FE	Yes	Yes	Yes

Notes: The reported results are coefficients from probit regressions. The sample is a panel of undergraduate students from 2000 to 2016. These students were admitted from 2000 to 2012 via FUVEST and concluded the course. All regressions include admission year at USP fixed effects and a gender dummy. In column (2) we add age and in column (3) a dummy indicating if the student took more than 5 years to graduate. P-values: *p<0.1; **p<0.05;***p<0.01.

Table F.5 – Gender performance differences in the total ANPEC grades by quantile

	OLS	Quantile regressions				
		10	25	50	75	90
Total ANPEC grade						
Women	-4.750***	-3.414*	-3.798***	-5.119***	-4.455***	-5.483***
	(1.156	(1.949)	(1.348)	(1.461)	(1.517)	(1.901)
Age (at admission)	-0.729***	-0.044	-0.254	-0.555	-0.681*	-0.746
	(0.265)	(0.530)	(0.367)	(0.397)	(0.412)	(0.517)
Took more than 5 years to graduate	-5.275***	-1.580	-3.693	-3.839	-6.726**	-11.875***
	(1.934)	(4.013)	(2.776)	(3.009)	(3.124)	(3.914)
FUVEST (USP admission exam) Ranking	-0.025***	-0.021	-0.023**	-0.033***	-0.024**	-0.021
	(0.008)	(0.013)	(0.009)	(0.010)	(0.010)	(0.013)
Avg grade compulsory courses	3.986***	5.117***	4.360***	4.449***	4.847***	3.784***
	(0.573)	(1.054)	(0.729)	(0.791)	(0.821)	(1.028)
Number of observations	296	296	296	296	296	296
Statistics (Mean and Percentiles)	23.53	9.40	16.25	23.57	30.63	36.50

Notes: Sample restricted to students from USP who took the ANPEC exam, were admitted from 2000 to 2010 through the FUVEST exam, and successfully completed their undergraduate degree. The dependent variable is the total ANPEC grade, which is calculated by summing the scores of Microeconomics, Macroeconomics, Statistics, and Mathematics. In all regressions, we control for age at admission, FUVEST ranking, a binary variable indicating whether the student took more than five years to complete the undergraduate program, and the average grade in compulsory courses during the undergraduate studies. The 'OLS' column presents the results of OLS regressions, while the 'Quantile regressions' columns present estimates for quantile regressions at percentiles 10, 25, 50, 75,and 90. P-values: *p<0.1; **p<0.05; ***p<0.01. In the last row we present descriptive statistics for the response variable: average for the OLS regression and the percentile related to each quantile regression (10, 25, 50, 75, 90).

♦ We are grateful for the data that was kindly provided by ANPEC and the Department of Economics at the University of São Paulo. We are also grateful to the Department of Economics at the University of São Paulo, FIPE and FAPES grant # 2019/16952-7 for their financial support. Bruna received funding from São Paulo Research Foundation (FAPESP - grant # 2021/08880-6) to conduct academic research. Paula received funding from CNPq(grant 304221/2022-8) to conduct her research.

CONFLITO DE INTERESSE

Os autores declaram não terem quaisquer conflitos de interesse.

EDITOR-CHEFE

Dante Mendes Aldrighi https://orcid.org/0000-0003-2285-5694
Professor - Department of Economics University of São Paulo (USP)

Fábio Waltenberg https://orcid.org/0000-0003-3404-7424