Helsinki GSE

HELSINKI GSE DISCUSSION PAPERS 54 · 2026

Gender Gap in High-Stakes Exams: What Role for Exam Preparation?

Cécile Bonneau
Léa Dousset

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HANKEN

Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 54 · 2026

Cécile Bonneau and Léa Dousset:
Gender Gap in High-Stakes Exams: What Role for Exam Preparation?

ISBN 978-952-7543-53-5 (PDF)
ISSN 2954-1492

Helsinki GSE Discussion Papers:
https://www.helsinkigse.fi/discussion-papers

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, February 2026

Gender Gap in High-Stakes Exams: What Role for Exam Preparation?

Cécile Bonneau* Léa Dousset†

February 2, 2026

Abstract:

In this paper, we show that gender gaps in performance observed in high-stakes exams, relative to low-stakes exams, are partially shaped by the context of exam preparation. We examine the role of exam preparation using data from French elite STEM higher education programs. First, we document a substantial gender gap in admissions to the most selective STEM graduate schools compared with slightly less selective institutions. Second, we provide causal evidence that the competitiveness of the learning environment during exam preparation affects the gender gap in performance in high-stakes exams. Our results further indicate that widening gender gaps in competitive learning environments primarily reflect disproportionate male gains rather than female underperformance. These findings have important implications for understanding the underrepresentation of women in elite programs and the gender pay gap among top STEM workers.

JEL codes: I23, J16, J24.

Keywords: Higher Education, Gender, STEM Fields, Selective Programs, High-Stakes Exams. [1]

*Aalto University. Email: *cecile.bonneau@aalto.fi.* Professional address: Ekonominaukio 1, 02150 Espoo, Finland.

†Paris School of Economics (PSE). Email: *lea.dousset@gmail.com.*

[1]We are thankful for the support of the EUR grant ANR-17-EURE-0001 and EDUCA Flagship, funded by the Research Council of Finland (#358924, #358946). Financial support for this study was also received from the Chair 'Educational Policies and Social Mobility' (Ardian-DEPP-PSE) and from the CEPREMAP. We would like to express our gratitude to the SCEI team for granting us access to the data. We also thank the DEPP for providing data on prior academic achievement, as well as the participating programs that agreed to share their school records for our data collection. We thank our great research assistants: Clémentine Bouleau, Maria Fe Garcia Seminario and Francesca Rinaldi. This paper benefited greatly from the helpful discussions and comments of Anne Boring, José de Sousa, Tarikua Erda, Gabrielle Fack, Hannah

1 Introduction

A growing literature documents a gender gap in performance on high-stakes exams relative to more continuous forms of assessment (Cai et al., 2019; Arenas and Calsamiglia, 2025). These gender performance gaps in high-stakes exams have important implications for the allocation of talent. They affect the representation of women in selective programs and competitive environments and can generate efficiency losses in admission systems that place substantial weight on high-stakes exams. While recent work has emphasized the lack of socio-economic diversity in elite higher education (Chetty et al., 2025), much less is known about the origins of gender disparities within these highly selective environments.[2] In particular, little is known about the mechanisms underlying women's lower relative performance on high stakes exams, particularly the role of exam preparation. Because selective institutions are unlikely to abandon exam-based admissions, identifying the sources of these gaps is essential for designing policies that can mitigate them.

Studying the role of exam preparation in generating gender gaps in high-stakes exam performance is empirically challenging. First, students usually prepare in broadly similar institutional and educational settings, generating limited variation in preparation contexts. Second, data on students' performance during the preparation period are rarely available, making it difficult to measure how preparation evolves and how it relates to entrance exam outcomes.

This paper opens the black box of the role of exam preparation in generating gender gaps in high-stakes exams. We overcome the first challenge thanks to our setting which provides quasi-exogenous variation in the selectivity of the preparation environment, generated by ability tracking of students during the exam preparation period. We

Farkas, Catalina Franco, Manon Garrouste, Marion Goussé, Julien Grenet, Marc Gurgand, Gustave Kenedi, José Montalbán Castilla, Donato Onorato, Thomas Piketty, Kamelia Stavreva, Camille Terrier and Georgia Thebault. This paper uses data collected from school records and administrative data, and we are not at liberty to publish online; however, we can provide the programs as well as detailed instructions and assistance in how to apply and access the data in order to enhance the replicability of our analysis. The project has received approval from the CNRS DPD (2-22076V2). All remaining errors are our own.

[2]The underrepresentation of women in elite higher education programs, either in absolute terms or relative to their share in the overall student population, has been documented in a range of settings, including the U.S. (Bielby et al., 2014), China (Cai et al., 2019; Han et al., 2025), France's elite schools (Bonneau et al., 2021), Colombia's most selective university entrance exam (Franco and Skarpeid, 2025), and Spain following a college-admission reform (Arenas and Calsamiglia, 2025).

further assemble large-scale administrative data on entrance exams and link them to self-collected school records that track students' performance throughout their preparation to overcome the second challenge.

Our study takes advantage of the institutional features of elite STEM higher education programs in France. Admission to French elite STEM graduate schools (*Grandes Écoles d'ingénieurs*) relies on competitive national entrance exams and requires high-school graduates to first complete two to three years in highly selective undergraduate STEM programs (*Classes Préparatoires aux Grandes Écoles*, hereafter prep programs), higher-education programs explicitly designed to prepare students for these entrance exams.[3] During the second year, prep programs implement ability tracking: within each program, "star" classes enroll the highest-performing students. This tracking creates variation in the selectivity and competitiveness of students' learning environments, as students in star classes have higher-ability peers, more demanding teachers, greater intellectual stimulation, and more competition stemming from the fact that students within each class are ranked against one another very frequently.

We rely on novel administrative data from the centralized admission process to STEM graduate schools organized by the *Service des concours écoles d'ingénieurs* (SCEI), covering 165,450 applicants from 2015 to 2023. We complement these data with administrative school records that we manually collected from 18 prep programs, which enable us to observe students' performance during exam preparation for roughly 10% of the full applicant sample. Finally, we incorporate administrative data on previous academic performance through high school graduation exams[4] and web-scraped information on the early-career salaries of STEM graduates, aggregated at the cohort $\times$ school $\times$ gender level.[5]

Using this extensive dataset, we document that women are less likely to enter top-tier STEM graduate schools: while they represent 25% of STEM graduate school students

[3]Only 2.5% of a birth cohort follows this path, and in 2021, 50% of STEM prep-program students graduated high school with highest honors, compared to 10% of all students in France (Table 1).

[4]These data come from the OCEAN databases of the *Direction de l'évaluation, de la prospective et de la performance* (DEPP).

[5]Sourced from the *Commission des Titres d'Ingénieur* website, based on a compulsory student survey for each graduate program.

overall, their share falls to 20% in the top 10% most selective programs—a 5-percentage point, or roughly 20%, decline. This underrepresentation is striking given that (i) women who choose to enroll in STEM prep programs are positively selected: 59% graduated high school with the highest honors, compared with 47% of men in these programs (and 10% of all students in France); and (ii) it has implications for the gender pay gap among STEM workers: in our sample of STEM graduates, the gender pay gap one year after graduation is €1,260 (3% of men earnings). When we adjust for the selectivity of the STEM graduate school attended, the gap narrows to €380, indicating that more than two-thirds of the observed gender pay disparity can be attributed to differential access to the most selective STEM graduate schools.[6]

Our objective is to uncover the mechanisms behind the gender gap in admission to the most selective STEM graduate schools, with a particular focus on the role of the exam preparation environment.

Thanks to our rich and novel administrative dataset, we first use a decomposition analysis to descriptively track gender gaps at each stage of the admission process to STEM graduate schools. We find that women's underrepresentation in the top-tier graduate programs arises from three main sources: (i) lower performance on high-stakes entrance exams, conditional on performance just before the exam; (ii) a lower probability of applying to the most selective entrance exams; and (iii) a gender gap in performance that emerges and widens during the exam preparation period, with this third factor being the most important. These patterns motivate our closer analysis of the role of the exam preparation environment.

To study more closely the role of exam preparation in generating gender gaps on high-stakes exams, we exploit the tracking system that assigns students to "star" or standard classes at the end of the first year in prep program. We use two complementary identification strategies to estimate the gender-differential impact of being placed in a star class. First, we compare gender performance gaps in high-stakes entrance exams across class types, controlling for detailed measures of prior academic achievement and prep-

[6] These figures represent wages excluding bonuses. When bonuses are included, the raw gender pay gap is €2,000, and the gap accounting for graduate school fixed effects drops to €960.

program fixed effects.[7] We refer to this approach as a double-difference specification (track $\times$ gender). Second, we estimate a regression discontinuity design at the admission cutoff for star classes. A key requirement for both of our identification strategies is that, conditional on prior academic performance, women and men have the same probability of admission to star classes. We verify that this condition holds in the data. These empirical strategies are complementary: the double difference yields estimates close to the average treatment effect of star-class enrollment and applies to our full student sample (N = 89,079),[8] whereas the regression discontinuity design identifies a local average treatment effect at the admission margin and can only be implemented for a balanced sample of students with available school-record data (N = 6,585).

Across the two identification strategies, we find that men's academic performance benefits more from star-class environments than women's. The double-difference estimates show that the gender gap in admission to the top 10% most selective STEM graduate schools is twice as large for students in star classes as for those in standard classes (6 versus 2.7 percentage points). A similar pattern emerges for expected earnings, with a significantly larger gender pay gap among star-class alumni compared to standard-class alumni (60% larger).

The regression discontinuity design shows that, among students at the margin of star class admission, women experience no significant change in their probability of entering the top 10% most selective STEM graduate schools, while men's probability of admission more than doubles. Our findings thus indicate that the widening gender gap in these competitive learning environments stems from disproportionate male gains rather than female underperformance.

We perform several robustness checks to validate our results. The results are consistent across various definitions of graduate school selectivity, whether based on the prior academic achievements of admitted students or applicants' revealed preferences (Avery et al., 2013). The results are also robust to a range of regression discontinuity specifica-

[7] In one specification, we fully interact gender with demographics, prior achievement, and program fixed effects to allow observable characteristics to have gender-specific effects.

[8] Our study sample (N = 89,079) is smaller than the full applicant sample (N = 165,450) due to our restriction to prep programs with a star–standard class tracking system. Table 1.2 in the Online Appendix compares the samples.

tions, including changes in bandwidth choice and polynomial order.

Our paper contributes to the literature on the success of women who pursue STEM studies. While a substantial body of work, summarized in Kahn and Ginther (2018), examines the underrepresentation of women in STEM fields, less attention has been paid to those who choose STEM pathways. Existing studies on female persistence in STEM programs show that women are more likely to leave STEM (Ellis et al., 2016; Landaud and Maurin, 2025; Kugler et al., 2021) and are less likely to remain in STEM occupations after graduation (Beede et al., 2011; Delaney and Devereux, 2022). Our article extends this literature by investigating women's underrepresentation in the most selective STEM programs, rather than average ones. This is crucial for several reasons. First, this underrepresentation likely contributes to the substantial gender pay gap observed among STEM professionals and at the top of the income distribution. Secondly, low representation of women in these top programs suggests substantial scope for increasing participation. Lastly, while the broader underrepresentation of women in STEM may reflect differences in preferences (Ahimbisibwe et al., 2025)—factors that are difficult to address in the short run—the specific underrepresentation in top-tier programs may not be driven solely by preferences. The highly detailed nature of our administrative data, including the rarely available information on student preferences, allows us to examine the role of these preferences in depth.

Our article also contributes to the literature on gender differences in competitive and high-stakes environments. A large body of work documents women's underperformance in such settings, particularly in mixed-gender contexts (Gneezy et al., 2003; Niederle and Vesterlund, 2007; Gneezy et al., 2009), with recent evidence pointing to larger gaps among young people (Flory et al., 2018) and highly competitive individuals (Saccardo et al., 2018). While most of this research is based on laboratory experiments, Buser et al. (2014) show that lab measures of competitiveness predict real educational choices, underscoring the need for field-based evidence. More recent studies explore gender differences in real educational environments by exploiting variation in exam stakes, competitiveness, and selectivity (Azmat et al., 2016; Montolio and Taberner, 2021; Landaud et al., 2020; Montalbán and Sevilla, 2023; De Sousa and Hollard, 2023). Although much of this research focuses on earlier schooling, several studies also document female underperformance in college entrance exams and other high-stakes contexts (Jurajda and Münich, 2011; Ors et al., 2013; Pekkarinen, 2015; Cai et al., 2019; Schlosser et al., 2019; Franco and Gomez-Ruiz, 2024; Arenas and Calsamiglia, 2025). Our study complements this literature in three distinct ways. First, by focusing on a highly selected population of high-achieving students who have chosen selective STEM studies, we show that even among this group, genuinely well-prepared for high-stakes exams, gender gaps in exam performance persist. Second, we are able to open the black box of exam preparation thanks to the varying learning environments during exam preparation and the availability of rare within-program student performance metrics that we collected from school records. Finally, our design shows that in this setting, highly competitive learning environments widen the gender gap not because women underperform, but because men disproportionately improve.

Lastly, our research engages with the literature on educational tracking, reviewed in Betts (2011). This work shows that high-achieving minority students benefit from tracking (Card and Giuliano, 2016), and that low-performing students can benefit from instruction tailored to their level (Duflo et al., 2011). Closely related, Landaud and Maurin (2022) finds that tracking in French prep programs increases social gaps in access to elite schools. We add to this literature by examining gender-differentiated effects of tracking.

The article proceeds as follows. Section 2 describes the institutional context of French higher education, with a focus on undergraduate STEM prep programs and their tracking system. Section 3 presents the data and descriptive statistics. Section 4 decomposes the gender gap in admission to top-tier STEM graduate schools into several explanatory factors. Section 5 exploits tracking between standard and star classes to assess how the competitiveness of the learning environment during exam preparation affects the gender performance gap on high-stakes exams. Section 6 concludes.

2 Institutional Background

French Higher Education System. The French higher education system is characterized by a significant degree of academic hierarchy. After the high school graduation exam

(*Baccalauréat*), students who pursue higher education can choose between three main tracks: (i) technical and vocational programs, enrolling about 30% of first-year students in 2021–2022; (ii) a non-selective academic track represented by public universities, which enrolled around 50%;[9] and (iii) a selective academic track composed of preparatory programs (*Classes Préparatoires aux Grandes Écoles (CPGE)*) and elite graduate schools (*Grandes Écoles*), enrolling 7%. The coexistence of selective and non-selective academic tracks is a distinctive feature of the French higher education system.[10] Figure 1.A1 illustrates the structure of French higher education and its various tracks.

Undergraduate Prep Programs and Elite Graduate Schools. Prep programs are intensive two- to three-year higher education programs (see Figure 1), mostly hosted in prestigious high schools, preparing students for the competitive national entrance exams to elite graduate schools. Their curriculum, which is equivalent to roughly two years of a bachelor's degree in several subjects (mathematics, physics, chemistry, engineering science, and/or computer science), is highly demanding. Students typically complete four to six hours of written mock exams each week (often on Saturdays) and two individual or small-group oral assessments, with frequent peer-based ranking within their class. Admission to these programs is highly selective, based on grades and teachers' assessments from the final two years of high school. They attract top performers: over our study period, 50% of prep students graduated high school with highest honors, compared to 10% of all students in France (Table 1). Prep program students who fail to gain admission to elite graduate schools are granted direct entry into the third year of a university program after completing their two-year prep program, ensuring that no academic year is lost for these top students.

Elite graduate schools, created after the French Revolution to train leaders in poli-

[9] Until 2018, public universities were non-selective, and access was formally granted to anyone holding the high school graduation exam. In the case of oversubscribed universities and programs, random lotteries were drawn to select students. Since 2018, universities have been allowed to select their students based on their own criteria. However, most university programs remain undersubscribed and, therefore, not selective in practice: Bechichi et al. (2021) estimated that in 2018 and 2019, 84% of university programs are non-selective, in the sense that they refuse less than 5 percent of applicants.

[10] The remaining 13% of first-year students enter other specialized programs (e.g., paramedical training or specialized schools).

tics, business, science, the military, and academia, remain central in shaping the elite in France. Admission is *meritocratic*, in contrast to the *aristocratic* selection practices of the past. Today roughly 6% of a birth cohort graduates from these institutions, including 3% from STEM ones. Although the number of places in elite STEM graduate schools is sufficient to accommodate all students from STEM prep programs, substantial variation in selectivity and a pronounced hierarchy persist, making entrance exams particularly high-stakes.[11]

Tracking in STEM Prep Programs. This study focuses on Science, Technology, Engineering, and Mathematics (STEM) preparatory programs, which enroll about 25,000 students annually—roughly two-thirds of all prep program students. These programs comprise five subfields: mathematics–physics, physics–chemistry, physics–engineering science, engineering science, and biology (see Diagram 1.A3 in the Online Appendix). All subfields except biology use an ability tracking system in the second year, separating students into more selective "star" classes (*classes étoile*) and standard classes, a practice most common in larger and more prestigious prep programs.

Although the official curriculum is theoretically identical across standard and star classes, star classes typically go beyond it and provide more intensive preparation for the most competitive entrance exams. Weekly written assessments mimic these exams, and frequent peer-based rankings within the class foster a more competitive environment. Star classes also tend to be smaller and taught by more experienced teachers. While star-class students are more likely to sit for the most selective exams, overlap remains substantial: 98% of star-class students take at least one highly selective exam, compared with 89% in standard classes.

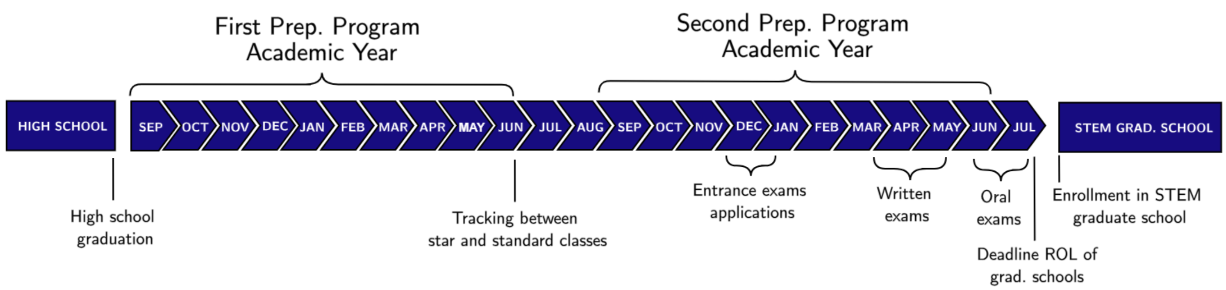
Admission to elite STEM graduate schools. At the end of their second year, prep students sit competitive entrance examinations for more than 200 elite graduate schools. To manage volume and avoid scheduling conflicts, most graduate schools group their exams into five large consortia, applying different weightings to the same written tests within

[11] The average percentile rank at the high school graduation exam is 91 among students enrolled in the most selective 10% of schools, compared with 38 in the least selective 10% (Bonneau et al., 2021).

each consortium. Students select among these exams according to their preferences and goals. Admission is centralized through a clearinghouse that uses a college-proposing Gale–Shapley Deferred Acceptance mechanism (Gale and Shapley, 1962), with no limit on the number of programs a student may rank.

The admission timeline (shown in Figure 1) begins with exam registration in December–January, followed by written exams in April–May. Consortia typically administer two mathematics tests, two physics tests, a foreign language test, a literature test, and, depending on subfield, additional exams in chemistry, engineering science, and/or computer science. Candidates who pass the written stage proceed to oral exams in June–July. By late July, students submit their rankings to the clearinghouse. Offers are then issued in five successive rounds from late July to early September, partly because about 20% of students repeat their second prep year and thus release places to other candidates.

Figure 1: Schedule of Prep Program and STEM Graduate Schools Admission



Notes: This diagram illustrates the two-year timeline of preparatory programs and the admission process to elite STEM graduate schools. Some students repeat the second year to improve their chances of admission to a more selective program. ROL denotes the rank-ordered list of STEM graduate schools submitted by students.

3 Data

3.1 Data Sources

Our analyses rely on four distinct data sources, combining administrative records with self-collected data.

Administrative data from the centralized admission process to STEM graduate schools. We use novel and exhaustive administrative data on STEM graduate school admission from the *Service de Concours Écoles d'Ingénieur* (SCEI) covering 2015–2023. The dataset includes detailed demographics (age, gender, social background, geographical origin), academic information (prep program, subfield, star or standard class), entrance exam applications and results, students' rank-ordered lists (ROL) of graduate schools, and admission offers.

Administrative data from the Ministry of Education. To measure prior academic achievement, we link the SCEI data to administrative records from the Ministry of Education (*Direction de l'évaluation, de la prospective et de la performance* — DEPP) using encrypted identifiers. These records include results from the middle-school and high-school graduation exams for cohorts 2010–2021, providing both GPAs and subject-specific grades.

School records data. In the spring semesters of 2022 and 2023, we collected school records from 18 STEM prep programs that agreed to participate in the study (Figure 1.A2a shows their locations). Depending on availability, we obtained data for recent cohorts and current students, including demographics (gender, date of birth) and academic information (subfield, class, and subject-level grades). These records were statistically matched to the administrative data using gender, date of birth, class, and program identifiers, yielding a 96% match rate.[12] Overall, the records provide detailed grade information for 21,532 students, of whom 8,857 form a balanced sample that we can follow from the beginning to the end of preparation—representing about 10% of the administrative sample. These data allow us to observe students' performance over the two years of exam preparation.

Aggregated earnings data (CTI). We web-scraped data on median earnings of alumni from STEM elite graduate schools—by school, cohort, and gender—from the *Commission des Titres d'Ingénieurs* (CTI) for graduates between 2015 and 2025. For each graduate school, the CTI reports median earnings by gender from compulsory student surveys for the previous graduating cohort.[13]

[12] One preparatory program declined to participate in the matching procedure; its data are therefore used only in analyzes relying solely on school records.

[13] When data for a given school–cohort were missing, we inferred values from adjacent cohorts; for the most recent cohorts not yet graduated, we projected earnings using graduate school-specific average annual

3.2 Samples of Analyses

We restrict the sample to STEM preparatory programs in France,[14] for which we have richer information on students' prior academic achievement. We further limit the data to prep program–subfield combinations that (i) are observed in all ten years (2015–2024), (ii) enroll at least ten students per year, and (iii) offer both a standard and a star class,[15] which is an essential feature of our identification strategy. These restrictions yield a main analysis sample of 89,079 students. The biology subfield is excluded because (i) our analysis focuses on programs where women are underrepresented, and (ii) this subfield does not feature the star/standard class distinction central to our research design. We also use a smaller subsample of students for whom we collected within-program grade data from school records. The overlap between these samples is illustrated in Figure 1.B1 in the Online Appendix.

3.3 Descriptive Statistics

Table 1 reports descriptive statistics for our main analysis sample from the SCEI administrative data. Prep students form a highly selected group and are not representative of the broader student population. Women account for only 26% of STEM prep students, compared to 54% of students nationally. Prep students also come from more advantaged backgrounds: 67% (resp. 52%) have a father (resp. mother) in a high-socioeconomic status (SES) occupation, versus 32% (resp. 21%) in the overall student population. Academically, almost two-thirds obtained the highest honors at the high school graduation exam (vs. 10% nationally). Students from Paris and those enrolled in Paris-based programs are overrepresented as well. Within our study sample, 47% of students are enrolled in star classes. Star classes have fewer women (21% vs. 30%), more high-SES students, and stronger academic profiles.

growth. After these imputations, earnings data remain missing for around 3% of individuals admitted to a graduate school in our sample.

[14] There are a few programs abroad, mainly in Morocco.

[15] This restriction covers 66 out of roughly 200 preparatory programs in France, which together enroll 54% of all prep students. Appendix Table 1.2 compares the full population of STEM prep students to our restricted study sample.

Table 1: Descriptive Statistics (2015-2023)

	Prep. program students (2015-2023)			All students (2016-2017)
	All (1)	Star classes (2)	Standard classes (3)	(4)
A. Students				
Female	0.26	0.21	0.30	0.54
Age	19.5	19.5	19.6	19.2
	(0.8)	(0.8)	(0.7)	(1.4)
Need-based scholarship holder	0.26	0.22	0.30	0.38
Father is high SES	0.67	0.72	0.62	0.32
Mother is high SES	0.52	0.57	0.48	0.21
High School Graduation Exam				
Highest honors	0.64	0.74	0.54	0.10
High honors	0.27	0.21	0.32	0.18
Honors	0.08	0.04	0.12	0.31
Without Honors	0.01	0.00	0.02	0.42
Percentile rank	0.84	0.88	0.80	0.51
	(0.15)	(0.12)	(0.16)	(0.29)
Percentile rank re-weighted (exams. coeffs.)	0.92	0.95	0.89	–
	(0.12)	(0.08)	(0.14)	
From Paris (in high school)	0.09	0.11	0.07	0.04
From Parisian area - outside Paris (in high school)	0.16	0.16	0.16	0.15
Enrolled in Paris	0.23	0.25	0.21	0.10
Enrolled in Parisian area - outside Paris	0.10	0.11	0.09	0.11
Star class	0.47	1.00	0.00	–
Repeater in prep program	0.18	0.19	0.18	–
B. Prep Programs				
Number of prep programs	66	66	66	–
in Paris	13	13	13	–
in Parisian area (outside Paris)	7	7	7	–
Number of classes	262	130	132	–
Number of students	89,079	42,018	47,061	1,090,356

Notes: This table reports descriptive statistics for all students in our sample, and separately for those in star and standard classes. The sample is drawn from SCEI administrative data covering all applicants to elite STEM graduate schools from 2015 to 2023, excluding the biology subfield. We retain only program–subfield combinations offering both star and standard classes in the second year. Descriptive statistics for all STEM prep students appear in Appendix Table 1.2. Socioeconomic status (SES) follows the Ministry of Education statistical service classification. Column (4) compares these figures with all first- and second-year higher-education students in France in 2016–2017, from Bonneau et al. (2021).

Appendix Table 1.2 reports summary statistics for our different samples: the full population of STEM prep students and the subset of prep programs from which we collected school records. The full population is slightly less socially and academically advantaged than our study sample, as programs without a star class tend to be smaller and somewhat less selective, though both groups remain far more privileged than the overall student population. The school record sample also modestly overrepresents Parisian programs and recent cohorts, reflecting differences in record availability. To address these imbalances, we apply inverse probability weights to reweight underrepresented observations so that the school record sample resembles our study sample.[16] The final column of the table reports descriptive statistics for this reweighted school record sample.

3.4 Outcome Variables

Our main outcome is a binary indicator equal to 1 if a student's final offer in the centralized admission process is from a top-decile STEM graduate school, which we defined using two selectivity measures described below, and 0 otherwise. For both definitions, selectivity is computed separately by subfield, as different subfields lead to slightly different sets of STEM graduate schools. Our rankings align closely with popular online rankings available to students such as those on the *L'Étudiant* website.

Selectivity. The first measure is *objective*, based on the average percentile rank at the high school graduation exam of students admitted in the first round of admission to the STEM graduate school. We refer to this as *selectivity*.

Desirability. The second measure, following Avery et al. (2013), is *subjective*, and derived from students' revealed preferences. We estimate a rank-order logit model using data on students' preferences:

$$U_{i,j} = \theta_j + \epsilon_{i,j}, \tag{1}$$

[16]Weights are estimated via a logistic model predicting inclusion in the school-record sample as a function of gender, star-class status, geographic origin, low-income status, parental SES, nationality, repeater status, disability status, high-school track and option, and the gender composition of the program–subfield–cohort.

where $U_{i,j}$ is student i's rank of graduate school j and θ_j a school fixed effect. Standardized estimates $\hat{\theta}_j$ serve as a proxy for desirability: more desirable schools attract more applications and are ranked higher.[17] As a robustness check, we compute separate fixed effects for women ($\hat{\theta_{jF}}$) and men ($\hat{\theta_{jM}}$), allowing us to assess gender-specific preferences and their role in female underrepresentation in the most selective STEM graduate schools. We refer to $\hat{\theta}_j$ as *desirability* and to ($\hat{\theta_{jF}}$) as *desirability for female students*.

Motivation of the variables of interest. Admission to a top-tier STEM graduate school has substantial consequences for labor-market outcomes. Using aggregate data by graduate school, cohort, and gender, we find a clear earnings premium for the top 10% of graduate schools (Figure 1.2 in the Appendix): graduates from this decile earn about €4,400 more annually than those from the ninth decile just one year after graduation when bonuses are excluded, and about €6,000 more when bonuses are included. Among STEM graduates, the raw gender wage gap is €1,260, but it falls to €380 after controlling for graduate school fixed effects, indicating that women's underrepresentation in the most selective programs contributes meaningfully to the wage gap among top STEM earners.[18] Top-tier schools also channel graduates into leadership roles: as of October 2022, 20 of the 40 CEOs of firms listed in the CAC 40, France's leading stock market index, had a science background, and 19 trained at one of the top 10% of STEM graduate schools. Selectivity can also be measured by acceptance rates, which average 12% among the top 10% of STEM graduate schools (Online Appendix Figure 1.B3), even though applicants are already a highly selected student population, making these rates broadly comparable to Ivy League admissions (5.2% in 2022).

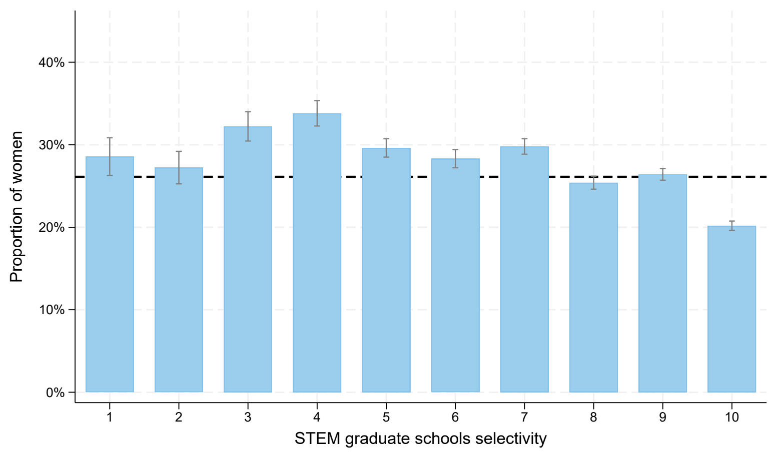
[17]Such methods have been criticized for preference misreporting, even in strategy-proof mechanisms such as deferred acceptance, particularly when students face limits on the number of ranked choices (Fack et al., 2019). In our setting, no such limit exists, making truthful reporting more likely. Consistent with this, the two selectivity measures are highly correlated (see Table 1.A2 in the Appendix), suggesting that our desirability measure is not substantially affected by misreporting.

[18]These figures represent wages excluding bonuses. When bonuses are included, the raw gender pay gap is €2,000, and the gap accounting for graduate school fixed effects drops to €960.

4 Gender Gap in Admission to Top STEM Graduate Schools

In this section, we provide a descriptive analysis of the gender gap in admission to the 10% most prestigious STEM graduate programs. As shown in Figure 2, the share of female students declines as graduate-school selectivity increases, reaching 20% in the top 10% most selective schools compared with 28% on average.[19] Access to top-tier STEM graduate schools has substantial labor market implications, including higher wages and increased access to leadership positions (see more details in Section 3.4).

Figure 2: Proportion of Female Students, by Decile of Selectivity of STEM Graduate Schools



Notes: This figure displays the proportion of female students admitted to STEM graduate schools, across deciles of selectivity. This figure use data from our study sample, restricted to prep-program and field combinations that include both star and standard classes. The black line represents the average proportion of female students. Program selectivity is measured by the average percentile rank of admitted students on the high school graduation examination. Online Appendix Figure 1.B4 presents results for the full sample of applicants—without restriction to program–field combinations with star classes— and separately by subfield, and confirms that the underrepresentation of women in the top 10% most selective graduate programs is observed across all subfields.

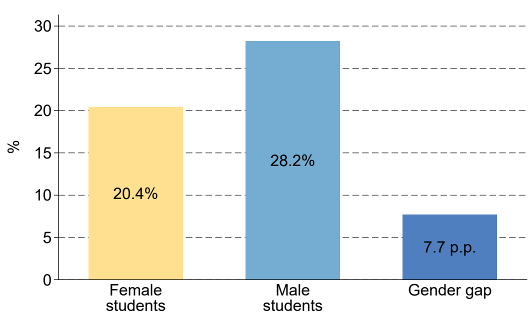
Interestingly, Online Appendix Figure 1.B4 shows that this gender gap in access to the top 10% most selective graduate programs is present across virtually all prep program subfields, including the female-dominated biology track, and is similar when using alternative measures of graduate-school selectivity based on revealed preferences of all students and of female students only (Avery et al., 2013). This gap may stem from (i) a gender gap in achievement before prep programs; (ii) a gender gap in achievement during prep programs; (iii) a gender gap in high-stakes exam performance, or (iv) gendered

[19] These figures refer to the subsample of preparation program–track combinations with both star and standard classes. Results for the full population, shown in Figure 1.B4 in the Online Appendix, display a similar pattern: women comprise 25% of STEM graduate students overall and 20% in top 10% programs.

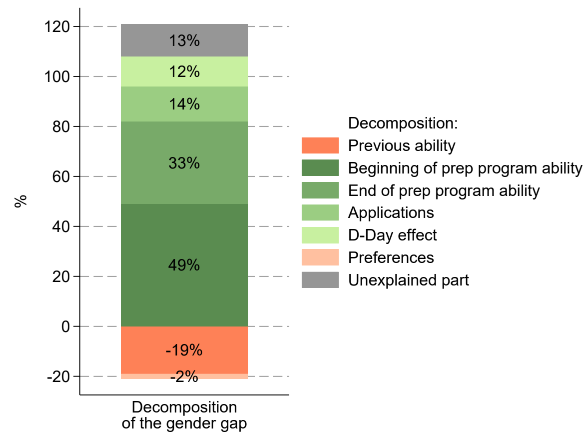
preferences over STEM schools. We quantify the contribution of each factor using a decomposition, leveraging a unique dataset that covers the full admission process to elite STEM graduate schools.

For this decomposition exercise, we restrict the sample to the balanced sample of students for whom we collected school records (N = 8,857).[20] Figure 3 presents the main decomposition results.[21] In the weighted school record sample, the raw gender gap is about 7.7 percentage points: 20.4% of women reach top programs versus 28.2% of men.[22]

Figure 3: Decomposition of the Gender Gap in Access to the Top 10% Most Selective STEM Graduate Schools



(a) Probability of Access to the Top 10% Most Selective Graduate Schools



(b) Decomposition of the Gender Gap

Notes: This figure decomposes the gender gap in access to the top 10% most selective STEM graduate schools. The analysis uses the subsample of students with available prep-program grades (N = 8,857), reweighted to match the characteristics of the study sample. Selectivity is defined by the average percentile rank of admitted students at the high-school graduation exam. Panel (a) reports access probabilities to top 10% most selective STEM graduate schools for men and women and the raw gender gap. Panel (b) shows how the raw gender gap narrows as successive controls are added. Previous ability is measured using decile indicators of GPA, as well as quintile indicators of grades in each core subject in the high school and middle school graduation exams. Prep-program ability is measured using decile indicators of GPA and quintile indicators of grades in each core subject in the first and last semesters, along with a dummy indicator for star-class status. We then add dummies for entrance exams applications and exam performance (percentile rank) and for whether each top graduate school appears in students' rank-ordered lists and how high in the list (percentile rank). Controls follow the chronological order of students' decisions.

We then examine how the gender gap in access to the top tier STEM graduate schools evolves as we sequentially add controls for students' prior achievement and preferences.

[20] To maintain representativeness, we apply inverse probability weights to the school record sample to match the study sample (see Section 3.3 for details).

[21] Table 1.C1 in the Online Appendix reports the same results, along with the evolution of the raw gender gap and the adjusted R^2.

[22] Although we focus on the top 10% of programs, these programs are generally larger and enroll more than 10% of students. We also restrict to program–subfield combinations with at least one star class, which are more selective and send more students to top programs (see Appendix Table 1.2).

These controls are introduced in the chronological order of students' trajectories and grouped by whether they capture performance or preferences. The baseline specification includes gender, demographic characteristics, and prep program, subfield, and cohort fixed effects, ensuring that all analyses are conducted within prep program and subfield.

4.1 Students' performance

We examine here the contribution of students' performance to the gender gap in access to top STEM programs.

Performance before prep program. On average, women entering STEM prep programs outperform men on the national high school graduation exam (Figure 1.1 in the Appendix): their average percentile rank at this national exam is 83.5 compared with 79.1 for men.[23] Thus, prior academic achievement does not explain the gender gap in access to the most selective programs. Based on prior achievement alone, women would be expected to enter the most selective STEM programs more often than men: controlling for detailed high school graduation exam grades actually increases the gender gap by 19% (Figure 3).

Performance during prep program. Using performance data from the school records we collected, we assess how achievement during the prep program contributes to the gender gap in admission to top STEM graduate programs. Such data, rarely available, provide a unique window into the black box of exam preparation. We control for achievement—decile indicators of GPA and quintile indicators of grades in each core subject (math, physics, chemistry, engineering science, computer science, French, and language)—at both the start (first-year, first semester) and end (second-year, second semester) of the prep program. The emergence and widening of the gender gap in performance account for much of the overall gap in access to top STEM graduate programs: 49% is explained by performance at the end of the first semester, and another 33% by the increase in the

[23]This holds both when using high school graduation exam GPA (Figure 1.1(a)) and when reweighting grades using coefficients from the most competitive entrance exams (Table 1.A1), which place greater weight on STEM subjects (Figure 1.1(b)), although the difference is less pronounced with this reweighting.

performance gap between the first and last semesters.[24] Overall, the prep-program period features a reversal and widening of the gender gap in performance, which accounts for an important share of the gender gap in admission to the most selective STEM graduate schools.

Performance on the day of the high-stakes exams. Performance on the day of the high-stakes exams explains 12% of the gender gap in access to the top 10% most selective STEM graduate schools.[25] This reflects a *D-Day effect*: conditional on similar achievement by the end of the prep program and similar application set (see discussion below), women slightly underperform in the most selective high-stakes exams. These results extend prior evidence of women's underperformance in high-stakes relative to continuous-assessment settings (Azmat et al., 2016; Arenas and Calsamiglia, 2025) to a highly selective population of STEM-oriented high achievers. This is notable given that these students are extensively prepared for such exams, taking weekly written and oral mock tests that replicate their exact format during two to three years.

Nevertheless, the D-Day effect plays a comparatively minor role relative to the widening gender gap in performance during exam preparation.

4.2 Students' preferences

We then examine the role of students' preferences, both in their application to competitive entrance exams and in the way they rank STEM graduate schools on their preference lists (see Diagram 1 for the timeline of application).

Application to competitive entrance exams. Differences in application behavior to competitive entrance exams[26] account for 14% of the gender gap in admission to top 10%

[24]For this specification, we also include an indicator for star-class enrollment, since GPA deciles and grade quintiles are defined at the class level and are therefore not comparable between students in star and standard tracks.

[25]We control for exam performance using (i) indicators for whether the applicant is ranked in competitive exams leading to top 10% graduate schools (students performing too poorly are not ranked) and (ii) their percentile rank among ranked candidates, controlled linearly within each exam.

[26]Applications are controlled using dummies for each entrance exam leading to top 10% STEM graduate schools, defined separately by subfield.

STEM graduate schools. Even at similar performance levels by the end of the prep program, within a certain prep program, women are less likely than men to apply to the most selective entrance exams. This finding aligns with prior evidence that women are less inclined to enter competitive environments (Niederle and Vesterlund, 2007), and extends it to a highly selected student population in a real-life setting.

Differences in preferences over STEM graduate schools. Differences in preferences are often cited to explain women's underrepresentation in STEM fields (Kahn and Ginther, 2018; Ahimbisibwe et al., 2025). We ask whether their underrepresentation in the most selective STEM programs also partly reflects gender differences in preferences over STEM graduate schools. Figure 3 shows that the role of preferences is minimal.[27] Among students with similar prep-program performance, entrance exam applications, and exam results, preferences over top STEM schools are very similar.

To corroborate these results, we exploit detailed data on students' preferences to (i) identify which characteristics of STEM graduate schools are valued by male and female students and (ii) run counterfactual simulations of the graduate school–student matching algorithm, estimating female students' preferences based on those of male students (see Online Appendix 1.C.2). These results are more precise, as they rely on the full structure of students' preferences.

We regress graduate school characteristics and their interactions with gender on school fixed effects estimated separately by gender from students' revealed preferences, ($\hat{\theta_{jF}}$ and $\hat{\theta_{jM}}$ in Equation 1). Overall, male and female students value similar graduate school characteristics (Table 2): both genders similarly value top 10% schools, overall selectivity, expected earnings, and military schools. This pattern largely holds when all characteristics are included jointly (Column 6, Table 2), despite a negative and significant coefficient for women on school selectivity decile. This suggests that, all else equal, women place somewhat less weight on selectivity, though not on whether a school falls within the top selectivity decile—consistent with the finding that the set of top 10% most desirable schools is nearly identical for men and women (Table 1.A2). The underrepresentation of women

[27] Preferences are controlled by adding dummies for whether each top 10% STEM graduate school appears in students' rank-ordered lists.

Table 2: STEM Graduate School Characteristics Valued by Students, by Gender

	(1)	(2)	(3)	(4)	(5)
	School FE (normalized)	School FE (normalized)	School FE (normalized)	School FE (normalized)	School FE (normalized)
Top Decile of Selectivity	2.19***				0.78***
	(0.079)				(0.066)
Top Decile of Selectivity × Female students	-0.022				0.078
	(0.11)				(0.093)
Decile of Selectivity		0.35***			0.22***
		(0.011)			(0.0081)
Decile of Selectivity × Female students		-0.021			-0.029**
		(0.016)			(0.011)
Expected Earnings			0.17***		0.069***
			(0.0076)		(0.0052)
Expected Earnings × Female students			0.00012		0.0037
			(0.011)		(0.0073)
Military School				1.53***	0.32***
				(0.19)	(0.074)
Military School × Female students				-0.063	-0.040
				(0.27)	(0.11)
Female students	-0.052	0.089	-0.059	-0.052	-0.017
	(0.048)	(0.12)	(0.44)	(0.068)	(0.29)
Subfield Fixed Effects	✓	✓	✓	✓	✓
Weights (number of admitted students)	✓	✓	✓	✓	✓
N	960	960	942	960	942
R^2	0.620	0.667	0.535	0.135	0.880

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table presents the regression coefficients of various STEM graduate school characteristics, as well as their interactions with a gender dummy variable, on school fixed effects computed separately for male and female students ($\hat{\theta_{jF}}$ and $\hat{\theta_{jM}}$ in Equation 1), allowing an analysis of the attributes valued by students in general and varying by gender of the students. The school fixed effects are calculated using a rank-ordered logit model based on students' ranked lists of preferred schools (Avery et al., 2013), and these standardized fixed effects also serve to quantify school desirability. In this analysis, we incorporate weights equal to the number of students admitted to each graduate school, to avoid placing too much weight on small programs. The school fixed effects are normalized to have a mean of 0 and a standard deviation of 1, and are computed separately for each subfield. The regression thus also includes subfield fixed effects. "Expected earnings" refers to the median gross annual salary, inclusive of bonuses, of the most recent alumni cohort from each STEM graduate school. These data are retrieved from the CTI website and are based on a compulsory survey of recent graduates conducted within each STEM school.

in the top 10% of STEM graduate schools cannot be explained by a lower preference for these graduate schools, as observed in students' rank-ordered lists of programs.

Taken together, these results suggest that women's underrepresentation is driven primarily by a gender performance gap that emerges and widens during exam preparation. Focusing solely on exam-day performance therefore overlooks a key mechanism: the growing gender gap in performance prior to the exam. This points to the role of the *learning environment* during exam preparation in shaping performance. We investigate this question by leveraging tracking into star and standard classes within prep programs.

5 Role of the Learning Environment

In this section, we aim to test whether the selective and competitive learning environment of prep programs affects male and female performance differently. Identifying this effect requires exogenous variation in the competitiveness and selectivity of students' learning environments. Our identification relies on a key institutional feature: at the start of the second year, students are tracked into star or standard classes based on their end-of-first-year performance. Star classes concentrate the highest-achieving students and foster a more competitive setting, where frequent within-class rankings increase peer pressure, raise academic stimulation, and strengthen preparation for highly selective entrance exams. This constitutes a bundled treatment of a more demanding learning environment, and we cannot isolate its individual components. However, because real-world variation in learning environments is inherently bundled, our results are well suited to capturing how increased selectivity and competitiveness in real-life learning environments shape the gender gap in performance.

We assess whether placement in a star class, relative to a standard class, differentially affects male and female performance on the most selective high-stakes entrance exams using two complementary strategies.

First, using the full sample of students enrolled in prep programs–subfield with standard and star class (N = 89,065), we estimate the average relative admission to top 10% STEM graduate schools of men and women in star versus standard classes, controlling for demographics, detailed prior achievement, prep-program and cohort fixed effects. This estimates represents an Average Treatment Effect (ATE). We fully interact gender with demographics, prior achievement, and program fixed effects to allow observable characteristics to have gender-specific effects. We also verify that, conditional on performance, there is no gender differential in admission to star classes.

Second, we focus on students near the star-class admission cutoff, who are arguably similar in observed and unobserved characteristics, using a regression discontinuity design (RDD). This analysis uses the smaller subsample with within-program grades (N = 6,585) and identifies a Local Average Treatment Effect (LATE) that complements the ATE.

5.1 Double Difference: Gender Gap in Star and Standard Classes

We aim to uncover the gender-differential impact of preparing for exams in more selective and competitive environments on students' academic success through a "double" difference, comparing both gender (*male* versus *female*) and class status (*star* versus *standard*).

5.1.1 Empirical Strategy

We compare the average relative performance of male and female students preparing for exams in *star* versus *standard* classes using the following reduced-form specification:

$$y_{ikpc} = \alpha_0 + \alpha_1 F_i \times S_i + \alpha_2 F_i + \alpha_3 S_i + \gamma \boldsymbol{X_i} + \lambda_k + \lambda_p + \lambda_c + \epsilon_{ikpc} \quad (2)$$

where y_{ikpc} is an indicator equal to 1 if student i in subfield k, program p, and cohort c is admitted to a STEM graduate school in the top 10% of selectivity or desirability; F_i is an indicator for being female; S_i is an indicator for being enrolled in a *star* class; and $\boldsymbol{X_i}$ is a vector of student characteristics, including detailed measures of prior achievement, which are controlled for flexibly using decile indicators.[28] We include subfield (λ_k), program (λ_p), and cohort (λ_c) fixed effects. Our coefficient of interest is α_1, which represents the gender-specific effect of being enrolled in a *star* class. We estimate Equation 2 using a linear probability model and assess robustness with a probit specification, as the outcome variable is binary (Table 1.D7 in the Online Appendix).

To ensure that our results on the interaction between star class and gender are not confounded by gender-differential effects of other observable characteristics, we also estimate the following reduced-form specification:

[28]Demographic controls include: geographic origin (Paris/Paris area), low-income status, parental socioeconomic status (four categories), French nationality, repeater status, disability status, science track in high school, option (engineering science or computer science), and the gender composition in the *subfield* $\times$ *program* $\times$ *cohort*. Prior achievement is captured by (i) decile-rank dummies from middle- and high-school graduation GPAs, and (ii) quintile-rank dummies for grades in mathematics, physics-chemistry, engineering science, French (written and oral), and foreign languages.

$$y_{ikpc} = \alpha_0 + \alpha_1 F_i \times S_i + \alpha_2 S_i + \gamma_M \boldsymbol{X_i} + \gamma_F \boldsymbol{X_i} \times F_i + (1 + F_i) \times (\lambda_k + \lambda_p + \lambda_c) + \epsilon_{ikpc} \quad (3)$$

This specification is identical to Equation 2, except that it includes interactions of all observable characteristics and fixed effects with gender.

Validity of the empirical strategy. A potential concern is that our specification could capture gender-differential selection into star classes. Two opposite mechanisms could be at play: (i) female students could be favored if teachers or headmasters aim to improve gender balance, given women's underrepresentation in star classes; or (ii) male students could be favored if they are perceived as more willing to join a selective and competitive environment.

Using within–prep-program grades, we find no evidence of gender-based selection. Table 1.1 in the Appendix shows that women are admitted to star classes less often overall. However, once we control for end of first-year rank in the class, the gender coefficient becomes statistically insignificant and very close to zero. This indicates that, conditional on performance at the end of the first year of the prep program, access to star classes does not differ by gender, supporting the validity of our empirical strategy. Consistent with this, a 2021 survey of teachers from 14 prep programs (70% response rate) reported that admission to star classes is based solely on first-year academic results, with no gender-based affirmative action.

We further assess the absence of pre-existing gender differences by estimating Equation 2 using earlier academic performance measures as outcomes. These placebo tests should reveal no gender-specific performance gaps among future star-class students. Appendix Table 1.4, using high school graduation exam GPA and first-year preparatory GPA, confirms the absence of gender differences in initial performance between students later assigned to standard versus star classes.

5.1.2 Results

We present our main results in Table 3. Columns (1)–(2) use graduate school *selectivity* as the outcome, measured by the average percentile rank of admitted students on the high school graduation exam, while Columns (3)–(4) use school *desirability* as the outcome, based on applicants' revealed preferences (Avery et al., 2013). Columns (1) and (3) report estimates from Equation 2, whereas Columns (2) and (4) report results from Equation 3, which includes interactions of all observable characteristics and fixed effects with gender.

All specifications show that female students benefit less from the selective and competitive environment of star classes than their male counterparts. The gender gap in admission to the top 10% most selective STEM graduate schools is larger in star classes than in standard classes. Our preferred estimate in Column (2)—which includes interactions of all observable characteristics with gender—indicates that, relative to the gender gap in standard classes, female students in star classes have a 3.3 percentage point lower probability of admission to top 10% most selective STEM schools than their male classmates. This corresponds to a 15% decrease from a baseline admission probability of 22%. Similarly, for school desirability, female students have a 4.6 percentage point lower probability (Column 4), a 20% reduction from baseline.[29]

The gender gap in access to the top 10% STEM schools is more than twice as large for star-class students (6.0 percentage points: 2.7 + 3.3) as for standard-class students (2.7 percentage points).[30]

Impact on expected earnings. We do not have individual earnings data, but we retrieve median salaries one year after graduation—disaggregated by graduate school, cohort, and gender—from a from a compulsory student survey reported by the *Commission des Titres d'Ingénieur* (CTI), the body that certifies STEM graduate schools.[31]

[29] Table 1.D1 in the Online Appendix displays all coefficients for the control variables included in Equation 2.

[30] Interestingly, for both definitions of graduate school selectivity, the coefficient from the specification including interactions of all observable characteristics with gender is about 80% of the coefficient without these interactions ($\frac{3.3}{4.2}$ and $\frac{4.6}{5.8}$). This suggests that differences in the effects of observable baseline characteristics—especially the impact of prior ability on later outcomes or gender specific prep-program fixed effects—can account for at most 20% of the additional gender gap observed in star classes.

[31] https://www.cti-commission.fr/accreditation.

Table 3: Admission to Top 10% STEM Graduate Schools

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Selectivity)	(3) Top 10% grad. schools (Desirability)	(4) Top 10% grad. schools (Desirability)
Baseline proba. of access	0.22	0.22	0.23	0.23
Female student × Star class	-0.042***	-0.033***	-0.058***	-0.046***
	(0.0061)	(0.0064)	(0.0060)	(0.0064)
Star class	0.27***	0.27***	0.29***	0.28***
	(0.0061)	(0.0062)	(0.0061)	(0.0062)
Female student	-0.027***		-0.028***	
	(0.0031)		(0.0030)	
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
Demographic Controls × Female student		✓		✓
MS & HS Exam Score × Female student		✓		✓
Year Fixed-Effects × Female student		✓		✓
Subfield Fixed-Effects × Female student		✓		✓
Program Fixed-Effects × Female student		✓		✓
N	89,065	89,065	89,065	89,065

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table illustrates the change in the probability of admission to top-tier STEM graduate schools for female students in star classes over the period 2015-2023. Selectivity of schools is measured in two ways: Columns (1) and (2) use the average percentile rank of admitted students at the high school graduation exam, while Columns (3) and (4) are based on the revealed preferences of applicants (Avery et al., 2013). In Columns (2) and (4), we include interactions between all controls and fixed effects and a gender dummy variable, allowing observable characteristics to have gender-specific impacts on the outcome. Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs: mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level. For a comprehensive list of coefficients on the demographic control variables, refer to Table 1.D1 in the Appendix.

Table 1.3 in the Appendix reports the effect of studying in a more selective and competitive environment on expected salaries one year after graduation. Columns (1)–(2) use earnings without bonuses, and Columns (3)–(4) include bonuses. The gender gap in expected earnings is around 60% larger among former star-class students than among those from standard classes. Our preferred estimate in Column (4)—which includes bonuses, the most competitive component of remuneration—indicates that preparing in a more competitive environment increases the gender pay gap by about €800 annually ($\approx 2\%$ of baseline earnings of €42,000).

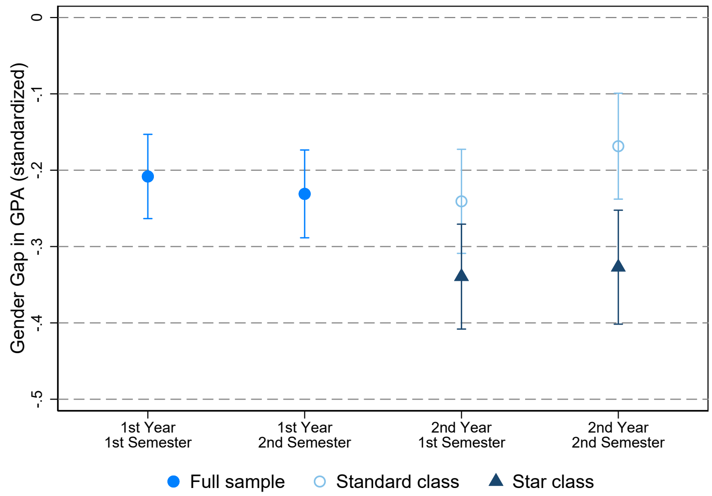
These results are consistent with prior evidence of persistent gender pay gaps, which are particularly pronounced at the top of the earnings distribution and among STEM professionals (Blau and Kahn, 2017).

During the Prep Program. Using the administrative school records we collected, we examine gender gaps in academic performance during the prep program by star-class status. Figure 4 presents the evolution of the gender gap in standardized GPA, based on regressions that control for track, subfield, and cohort fixed effects. We find that the gender gap widens over the course of the first year and is larger in the second year for students in star classes than for those in standard classes. In the final semester, the gender gap in GPA is -0.33 standard deviations for star-class students, compared with -0.17 standard deviations for students in standard classes. The star-class-specific gender gap observed in the most competitive entrance exams (Table 3) thus already begins to emerge during the exam preparation period.

Heterogeneity. We examine heterogeneity of our main results by parental income, prior academic achievement, repeater status and share of women in the prep program. Results are displayed in the Online Appendix (Tables 1.D2 to 1.D5).

The increased gender gap in access to top STEM graduate schools among star-class students is larger for low-income than for high-income female students, both in absolute and relative terms (Table 1.D2), and also larger for initially lower-achieving female students, in relative terms (Table 1.D3).

Figure 4: Gender Gap in GPA (standardized)



Notes: The figure presents the gender gap in standardized GPA among students in STEM prep program, using the balanced sample of school-record data we collected (N = 8,857), reweighted to match the full study sample. In the second year, results are split between star class and standard class students. GPAs are standardized within each *subfield* $\times$ *program* $\times$ *cohort* cell to have a mean of 0 and a standard deviation of 1. Each point represents the coefficient on a female dummy from a separate regression of GPA for each semester, controlling for preparatory program fixed effects, cohort fixed effects, and subfield (track) fixed effects. Standard errors are clustered at the *subfield* $\times$ *program* $\times$ *cohort* level.

About 20% of students (21% of men and 16% of women) repeat the second year of the prep program to retake the competitive entrance exams. The increased gender gap among star-class students is similar for first-time and repeat candidates (Table 1.D4), suggesting that the gender performance gap neither narrows nor widens among repeaters.

In Table 1.D5, we split the sample at the median share of women within prep-program $\times$ subfield cells. We find increasing gender gaps in star classes in both subsamples, with larger effects in programs with a higher share of women, although the effects in the two subsamples are not statistically significantly different.

Robustness checks. The Appendix reports several robustness checks. First, using revealed preferences based only on female students' rank-order lists (Table 1.D6), we find similar results and confirm that the gender gap in access to top 10% of STEM graduate schools is not driven by gender differences in preferences, as detailed in the descriptive section of the paper (Section 4).

Second, we test alternative measures of school selectivity: average graduate school ranking and top 20% (instead of top 10%). We obtain similar results, though effect sizes are smaller when using average selectivity (Table 1.D9), suggesting that the gender gap is concentrated in access to top-tier STEM programs.

Third, because our outcome is binary, we re-estimate the models using a probit specification (Table 1.D7) instead of a linear probability model, and the results remain virtually unchanged. We also include a specification interacting cohort, prep-program, and subfield fixed effects to allow prep-program quality to vary over time and across subfields (Table 1.D8), which yields consistent estimates.

Finally, to ensure that our results are driven by differences in exam preparation environments rather than by differences in the selectivity of exams taken by star- versus standard-class students, we flexibly control for the number of selective exams taken.[32] As shown in Table 1.D9, the results closely match our baseline estimates.

5.2 Regression Discontinuity at the Margin of Star Class Admission

In this final section, we use the subsample of students for whom we collected school records to estimate a regression discontinuity design at the margin of star-class admission. This approach complements the average treatment effect estimated displayed above with the double difference approach.

5.2.1 Empirical Strategy

We aim to estimate the effect of enrollment in a star class on the probability of admission to one of the top 10% most selective STEM graduate schools, separately by gender. This regression discontinuity design compares students who are just admitted to star classes with those who are not.[33] Unlike most regression discontinuity settings, we do not directly observe either the running variable or the threshold, as admission to star class is not determined by an explicit administrative cutoff but within each prep program. The first step is therefore to identify the appropriate running variable and the corresponding threshold for star class admission.

[32] We include fixed effects for the number of (i) very selective exams (leading only to top-tier schools) and (ii) selective exams (leading to at least one top-tier school).

[33] A similar approach was employed by Landaud and Maurin (2022) to estimate the overall effect of star class admission in one of the most selective preparatory programs in Paris.

Definition of the running variable. We know from the institutional context that admission to star classes is mostly determined by students' rank at the end of their first year in the prep program. Using subject-specific grades in science subjects (mathematics, physics, chemistry, engineering sciences, computer science), we construct students' weighted GPA using the coefficients from the most selective competitive entrance exams (Online Appendix Table 1.A1).[34] We then rank students within each class according to this weighted GPA.

Definition of the threshold. We determine the threshold for star class admission within each class in a data-driven manner, building on the methodology introduced by Hansen (2000). This approach has been adopted in contexts similar to ours (Hoekstra, 2009; Landaud et al., 2020; Bütikofer et al., 2023), particularly following Porter and Yu (2015) who showed its applicability to regression discontinuity designs. We implement an algorithm that, for each class, prep-program, and cohort, identifies the student rank at which the probability of star class admission exhibits the sharpest discontinuity.[35] We normalize the estimated threshold to zero and define the running variable as each student's relative rank distance from this estimated threshold.

RDD sample restrictions. We initially collected school-record grades for 21,532 students, but restrict the sample to those observed in both years of the preparatory program and matched to administrative entrance exam data (SCEI). Moreover, the threshold algorithm almost systematically identifies a cutoff between one admitted and one non-admitted student, while students further away from the threshold do not display a deterministic (0/1) probability of admission. Marginal cases immediately around the threshold are always-takers or never-takers and may thus exhibit characteristics that would bias our RD estimates (De Chaisemartin and Behaghel, 2020). We therefore exclude these observations from the main specification. We then construct a balanced RD sample, with

[34] We significantly reduce the weights assigned to French and foreign languages, as these subjects receive very little weight in star-class admission.

[35] Specifically, we select the cutoff that maximizes the R^2 from a regression of a star-class admission indicator on an indicator for being above a given rank. This procedure is repeated across all possible ranks within a class.

equal numbers of students on each side of the cutoff within each class, and restrict the running variable to the interval $[-20, 20]$, ensuring at least 100 students per rank value across programs.[36] The final sample includes 6,585 students.

First stage. Our method identifies a fuzzy regression discontinuity design: not all students above the estimated threshold are admitted to star classes, and not all students below it are assigned to standard classes. This fuzziness is expected for (at least) three reasons: (i) GPA is the main criterion, but additional non-observed factors, such as teachers' assessments, also play a role; (ii) GPA is not directly observed but reconstructed from raw grades, which introduces measurement error; and (iii) students transition from three subfields in the first prep year to four subfields in the second year (see Online Appendix Figure 1.A3), further contributing to the fuzziness of the estimated star-class admission cutoff. The first panel of Table 5 reports the first-stage results: the coefficient is 0.72, indicating a large and significant discontinuity in star-class admission and suggesting that our procedure is effective in recovering a valid running variable and threshold. Estimates are similar across genders, consistent with comparable admission criteria at the threshold, although they are less precise for women because of smaller sample size. Online Appendix Figure 1.E2 displays the first stage, separately by gender.

Density of the running variable. Figure 1.E1 in the Online Appendix shows that the density of the running variable around the star class admission threshold is continuous.

Balance of observable characteristics. Table 4 reports balance tests for observable student characteristics. For our gender-specific RD estimates, the key identifying requirement is the absence of gender-differentiated selection at the cutoff. Panel A shows perfect balance in demographic characteristics for the full sample of students, as well as for men and women separately. Panel B indicates that prior academic achievement is largely balanced, with a few small discontinuities that are expected given the number of coefficients tested. With respect to overall GPA, if anything, women just admitted to star classes

[36] Average class size is 35 (max. 48, min. 10).

Table 4: Balance of Observable Characteristics Around the RD Threshold, by Gender

	All Students		Male Students		Female Students	
	Baseline mean (1)	RD Estimate (2)	Baseline mean (3)	RD Estimate (4)	Baseline mean (5)	RD Estimate (6)
Panel A. Demographics						
Female	0.27	-0.02	–	–	–	–
		(0.06)				
Age	19.33	-0.00	19.34	0.02	19.30	-0.08
		(0.08)		(0.09)		(0.15)
Need-based scholarship holder	0.24	-0.08	0.24	-0.06	0.24	-0.14
		(0.05)		(0.06)		(0.09)
High SES	0.88	0.03	0.88	0.02	0.87	0.07
		(0.04)		(0.05)		(0.08)
From Paris	0.10	0.04	0.10	0.04	0.09	0.06
		(0.04)		(0.05)		(0.07)
Panel B. Previous Academic Achievement						
Highest honor at HS graduation exam	0.76	0.03	0.73	0.00	0.85	0.11
		(0.06)		(0.07)		(0.10)
Average GPA in 1st Year - 2nd Semester	11.57	0.27*	11.67	0.16	11.29	0.60**
		(0.16)		(0.19)		(0.29)
Average GPA in Math in 1st Year - 2nd Semester	11.61	0.53	11.76	0.31	11.19	1.19
		(0.35)		(0.37)		(0.82)
Average GPA in Physics in 1st Year - 2nd Semester	11.68	0.05	11.77	-0.09	11.45	0.47
		(0.23)		(0.27)		(0.41)
Average GPA in French in 1st Year - 2nd Semester	10.77	0.41	10.48	0.65*	11.53	-0.19
		(0.32)		(0.38)		(0.57)
Average GPA in Foreign Language in 1st Year - 2nd Semester	11.90	0.65*	11.64	0.72*	12.61	0.53
		(0.37)		(0.41)		(0.81)
Panel D. All Baseline Characteristics Jointly						
F-Stat		1.145		1.091		0.977
P-value		0.321		0.366		0.462
N.		2,275		1,668		607

Notes: This table presents non-parametric regression discontinuity biased-corrected estimates based on Calonico et al. (2017, 2019) to compare the characteristics of students near the cutoff for star class admission, for all students in the sample and separately by gender. Panel A and Panel B each assess different aspects of students' characteristics at this cutoff. Each coefficient results from a separate regression, where the student's relative distance to the cutoff serves as the running variable. The running variable is defined as the distance between a student's GPA rank and the estimated threshold rank for star-class admission, computed at the *class* × *subfield* × *prep program* × *cohort* through an algorithm. Non-parametric estimates use a triangular kernel, with bandwidths set to 6, i.e. the optimal bandwidth used in our main RD estimation. Columns 1, 3, and 5 display the mean value of the dependent variable for students below the cutoff. Standard errors are in parentheses.

appear slightly stronger rather than weaker. Joint tests of covariates balance reveal no statistically significant discontinuities at the cutoff.

Estimation. We estimate the effect of star-class enrollment on subsequent admission to the top 10% of STEM graduate schools using the `rdrobust` package (Calonico et al., 2014, 2017). The procedure implements bias-corrected local linear regressions on either side of the cutoff, constructs robust confidence intervals that account for variance and local-approximation bias, and uses data-driven bandwidth selectors to optimally balance bias and variance. This approach delivers consistent estimates of the treatment effect at the cutoff and valid inference under weak smoothness assumptions.

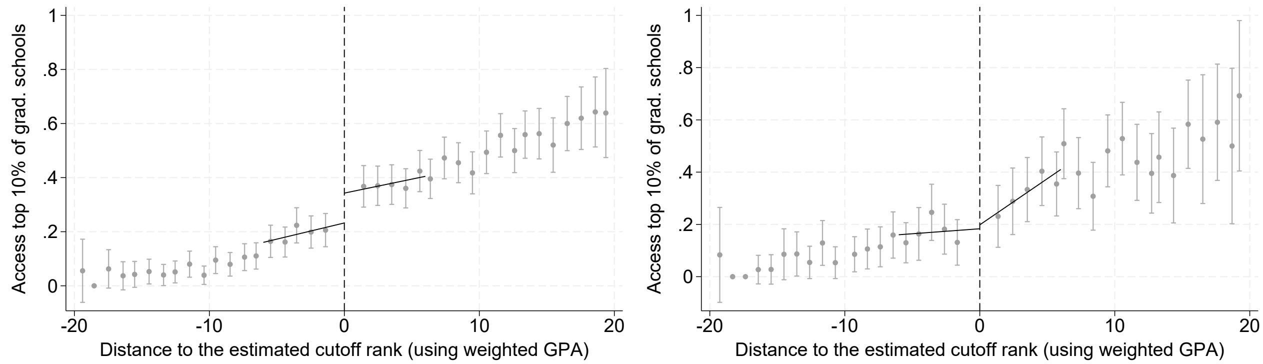
5.2.2 Results

Figure 5 and Table 5 present the main results on admission to top 10% STEM graduate schools. Two-stage least squares estimates accounting for RD fuzziness show that star-class enrollment increases admission by 15.9 percentage points, relative to an 11% baseline just below the cutoff. This effect is consistent with the stronger preparation provided by star classes and aligns with the findings of Landaud and Maurin (2022) for a single preparatory program.

Our estimates reveal pronounced gender heterogeneity. Star-class enrollment increases admission to top 10% STEM graduate programs by 18.7 percentage points for men, while the effect for women is smaller and statistically insignificant. Figure 5 nevertheless shows that, although star-class enrollment has no statistically significant effect for women at the margin, the slope of admission to top STEM graduate schools is steeper for female star-class students than for their peers in standard classes. This pattern suggests that star classes may enhance female students' performance trajectories, even if the effect is not statistically significant for marginal admits, in contrast to the clear effect observed for marginal male students.

Most research on gender performance gaps in competitive or high-stakes settings cannot distinguish whether the gap arises from women underperforming under pressure or from men performing better, since differences in evaluation methods capture only rel-

Figure 5: Admission to Top 10% Most Selective STEM Graduate Schools, by Gender



(a) Male Students

(b) Female Students

Notes: These figures depict the probability of admission probability to top 10% STEM graduate schools, by gender. Selectivity of graduate schools is measured using the percentile rank from high school graduation exam results of admitted students. The running variable is defined as the distance between a student's GPA rank and the estimated threshold rank for star-class admission, computed at the *class* × *subfield* × *prep program* × *cohort* through an algorithm.

Table 5: Regression Discontinuity: Admission to Top 10% Most Selective STEM Graduate Schools, by Gender

	All Students (1)	Male Students (2)	Female Students (3)
First Stage	Enrolled in Star Class		
Baseline mean (Below cutoff)	0.08	0.08	0.08
RD estimate	0.721***	0.738***	0.673***
	(0.044)	(0.050)	(0.089)
Regression Discontinuity	Top 10 % of grad. schools (Selectivity)		
Baseline mean (Below cutoff)	0.11	0.12	0.11
RD estimate (**ITT**)	0.114**	0.138**	0.036
	(0.055)	(0.066)	(0.098)
RD estimate (**ATT**)	0.159**	0.187**	0.054
	(0.075)	(0.088)	(0.144)
Obs. used in estimation	2,275	1,668	607
Total number of obs.	6,585	4,788	1,797

Notes: This table displays non-parametric regression discontinuity estimates of admission probability to top 10% STEM graduate schools. These estimates are based on Calonico et al. (2017, 2019). STEM graduate school selectivity is measured using the percentile rank from high school graduation exam results of admitted students. The running variable is defined as the distance between a student's GPA rank and the estimated threshold rank for star-class admission, computed at the *class* × *subfield* × *prep program* × *cohort* through an algorithm. For results comparability, the bandwidth for the estimation is fixed at -6 and 6, the one selected by the optimal bandwidth algorithm for the ATT on the full sample.

ative—not absolute—gender differences. Our results suggest that in this setting, more selective and competitive *learning environments* do not reduce women's performance, but disproportionately enhance men's performance, thereby widening the gender gap.

Robustness tests. We conduct several robustness checks of the main results (Online Appendix Tables 1.E1 and 1.E2), including varying the bandwidth, using a second-order polynomial, excluding two observations instead of one on each side of the cutoff, reweighting the sample using estimated weights to match the full sample, clustering standard errors at the running-variable level, and applying a local randomization approach.[37] Across specifications, the results consistently point to a larger effect of star-class assignment for male students than for female students, although the magnitude and statistical significance of the estimates vary.

In this section, we use a regression discontinuity design to show that, for students at the margin of enrolling in star classes, enrollment significantly improves male students' performance but has no significant effect on female students.

6 Conclusion

In this paper, we document that women are underrepresented in the most selective STEM graduate programs relative to their representation in slightly less selective ones. Using rich administrative data, we investigate the drivers of this disparity. We find that it is partly explained by lower relative performance by women on the exact day of the high-stakes entrance examinations, the "D-Day effect." However, the dominant factor is a gender gap in performance that emerges and widens during the preparation of high-stakes exams.

We thus examine the role of the learning environment during exam preparation in shaping the performance gap. We show that the gender gap is more pronounced among students preparing in more selective and competitive environments—specifically, star

[37]With this methodology, which is appropriate for discrete running variables such as ours, we focus on students around the cutoff, retaining six students below and six students above, which corresponds to the optimal bandwidth estimated by Calonico et al. (2019) for the ATT on the full sample. We then perform a simple comparison of means below and above the cutoff to test for the local average treatment effect (LATE). The underlying assumption is that these students are essentially locally randomly assigned to positions just below or just above the cutoff.

classes. We establish this using two complementary methodologies: (i) a difference-in-differences analysis of exam performance by gender and preparation class type, and (ii) a regression discontinuity design at the threshold of star class enrollment, where admission substantially increases men's likelihood of entering top graduate programs but has no significant effect for women. To support the validity of our strategy, we verify that admission to star classes is based solely on ability, with no gender differences in acceptance.

Our paper contributes to the literature on gender gaps in high-stakes competitive exams in three ways. First, we show that these gaps persist even in a highly selected and well-prepared student population. Second, we provide novel causal evidence on the role of the exam preparation environment, showing that more competitive settings amplify gender disparities. This is important because the selectivity of learning environments is likely to vary across a wide range of contexts. Third, we show that the gap arises primarily because men benefit disproportionately from such environments, rather than because women underperform in them. Our results have implications for explaining the gender pay gap at the top of the income distribution among STEM workers, and they document gender disparities that emerge long before fertility decisions, as individuals in our setting are 18–20 years old.

Our findings open avenues for future research. With our data, we cannot disentangle which specific aspects of competitive learning environments drive the widening of the gender gap—whether it is men's greater willingness to engage in competitive behavior, differential teacher behavior, and peer dynamics such as "boys' club" mechanisms (as described in Cullen and Perez-Truglia (2023)). Moreover, our results raise equity concerns about entrance examinations that rely heavily on extremely competitive preparatory processes. Improving the representation of women in the most selective STEM institutions, which is an explicit goal of many of these programs, may require rethinking the design of admission procedures. At the same time, our evidence does not address the efficiency of such selection systems: whether competitive traits are essential for these students' subsequent success in education and careers remains an open question. We leave these avenues for future work.

References

Ahimbisibwe, Isaac, Adam Altmejd, Georgy Artemov, Andrés Barrios Fernández, Aspasia Bizopoulou, Martti Kaila, Jin-Tan Liu, Rigissa Megalokonomou, José Montalbán, Christopher Neilson, Jintao Sun, Sebastian Otero, and Xiaoyang Ye, "Pipeline vs. Choice: The Global Gender Gap in STEM Applications," 2025. JSTOR Working Paper.

Arenas, Andreu and Caterina Calsamiglia, "Gender Differences in High-Stakes Performance and College Admission Policies," *Management Science*, 2025.

Avery, Christopher N., Mark E. Glickman, Caroline M. Hoxby, and Andrew Metrick, "A Revealed Preference Ranking of U.S. Colleges and Universities," *The Quarterly Journal of Economics*, 2013, *128* (1), 425–467.

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

Bechichi, Nagui, Julien Grenet, and Georgia Thebault, "D'Admission post-bac à Parcoursup: quels effets sur la répartition des néo-bacheliers dans les formations d'enseignement supérieur?," 2021. *France, portrait social - édition 2021*, collection « Insee Références », p. 105-121.

Beede, David N., Tiffany A. Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E. Doms, "Women in STEM: A Gender Gap to Innovation," 2011.

Betts, Julian R., "The Economics of Tracking in Education," *Handbook of the Economics of Education*, 2011, *3* (11), 341–381.

Bielby, Rob, Julie Renee Posselt, Ozan Jaquette, and Michael N. Bastedo, "Why Are Women Underrepresented in Elite Colleges and Universities? A Non-Linear Decomposition Analysis," *Research in Higher Education*, 2014, *55* (8), 735–760.

Blau, Francine D. and Lawrence M. Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

Bonneau, Cécile, Pauline Charousset, Julien Grenet, and Georgia Thebault, *Quelle démocratisation des grandes écoles depuis le milieu des années 2000 ?*, Vol. 30, Institut des Politiques Publiques, 2021.

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek, "Gender, Competitiveness, and Career Choices," *The Quarterly Journal of Economics*, 2014, *129* (3), 1409–1447.

Bütikofer, Aline, Rita Ginja, Katrine V. Løken, and Fanny Landaud, "Higher-Achievement Schools, Peers and Mental Health," *The Economic Journal*, 2023, *133* (655), 2580–2613.

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

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , _ , Max H. Farrell, and Rocio Titiunik, "Rdrobust: Software for Regression-Discontinuity Designs," *The Stata Journal*, 2017, *17* (2), 372–404.

_ , _ , _ , and _ , "Regression Discontinuity Designs Using Covariates," *Review of Economics and Statistics*, 2019, *101* (3), 442–451.

Card, David and Laura Giuliano, "Can Tracking Raise the Test Scores of High-Ability Minority Students?," *American Economic Review*, 2016, *106* (10), 2783–2816.

Chetty, Raj, David J. Deming, and John N. Friedman, "Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges," *The Quarterly Journal of Economics*, 2025, p. qjaf050.

Cullen, Zoe and Ricardo Perez-Truglia, "The Old Boys' Club: Schmoozing and the Gender Gap," *American Economic Review*, 2023, *113* (7), 1703–1740.

De Chaisemartin, Clément and Luc Behaghel, "Estimating the Effect of Treatments Allocated by Randomized Waiting Lists," *Econometrica*, 2020, *88* (4), 1453–1477.

De Sousa, José and Guillaume Hollard, "From Micro to Macro Gender Differences: Evidence from Field Tournaments," *Management Science*, 2023, *69* (6), 3358–3399.

Delaney, Judith M. and Paul J. Devereux, "Gender Differences in STEM Persistence After Graduation," *Economica*, 2022, *89* (356), 862–883.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence From a Randomized Evaluation in Kenya," *American Economic Review*, 2011, *101* (5), 1739–1774.

Ellis, Jessica, Bailey K. Fosdick, and Chris Rasmussen, "Women 1.5 Times More Likely To Leave Stem Pipeline After Calculus Compared to Men: Lack of Mathematical Confidence a Potential Culprit," *PLoS ONE*, 2016, *11* (7), 1–14.

Fack, Gabrielle, Julien Grenet, and Yinghua He, "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, 2019, *109* (4), 1486–1529.

Flory, Jeffrey A., Uri Gneezy, Kenneth L. Leonard, and John A. List, "Gender, Age, and Competition: A Disappearing Gap?," *Journal of Economic Behavior and Organization*, 2018, *150*, 256–276.

Franco, Catalina and Ingvild Skarpeid, "Stress, Exam Performance and Gender," 2025. Working paper.

— **and Marcela Gomez-Ruiz**, "Bridging the Gender Gap in Exam Performance through Nudges and Stress Reframing," 2024. Working Paper.

Gale, D. and Lloyd S. Shapley, "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 1962, *69* (1), 9–15.

Gneezy, Uri, Kenneth L. Leonard, and John A. List, "Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society," *Econometrica*, 2009, *77* (5), 1637–1664.

— **, Muriel Niederle, and Aldo Rustichini**, "Performance in Competitive Environments: Gender Differences," *The Quarterly Journal of Economics*, 2003, *118* (3), 1049–1074.

Han, Yi, Dorothea Kübler, Yiming Liu, Xinye Zheng, and Yibo Zong, "The Gender Gap in Gender-Blind University Admissions," 2025. Berlin School of Economics Discussion Papers 0062.

Hansen, Bruce E., "Sample Splitting and Threshold Estimation," *Econometrica*, 2000, *68* (3), 575–603.

Hoekstra, Mark, "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *The Review of Economics and Statistics*, 2009, *91* (4), 717–724.

Jurajda, Štěpán and Daniel Münich, "Gender Gap in Performance Under Competitive Pressure: Admissions to Czech Universities," *American Economic Review, Papers and Proceedings*, 2011, *101* (3), 514–518.

Kahn, Shulamit and Donna Ginther, "Women and Science, Technology, Engineering, and Mathematics (STEM): Are Differences in Education and Careers Due to Stereotypes, Interests, or Family?," in "The Oxford Handbook of Women and the Economy," Oxford University Press, 2018.

Kugler, Adriana D., Catherine H. Tinsley, and Olga Ukhaneva, "Choice of Majors: Are Women Really Different From Men?," *Economics of Education Review*, 2021, *81*, 102079.

Landaud, Fanny and Éric Maurin, "Tracking When Ranking Matters," 2022. THEMA Working Paper n°2023-12.

— **and** — , "The Choice of Retaking Competitive Exams," 2025. THEMA Working Paper n°2025-06.

— **, Son Thierry Ly, and Éric Maurin**, "Competitive Schools and the Gender Gap in the Choice of Field of Study," *Journal of Human Resources*, 2020, *55* (1), 278–308.

Montalbán, José and Almundena Sevilla, "Gender Differences in Performance: The Role of External Testing Environments," 2023. Working Paper.

Montolio, Daniel and Pere A. Taberner, "Gender Differences Under Test Pressure and Their Impact on Academic Performance: A Quasi-Experimental Design," *Journal of Economic Behavior & Organization*, 2021, *191*, 1065–1090.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away From Competition? Do Men Compete Too Much?," *Quarterly Journal of Economics*, 2007, *122* (3), 1067–1101.

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

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

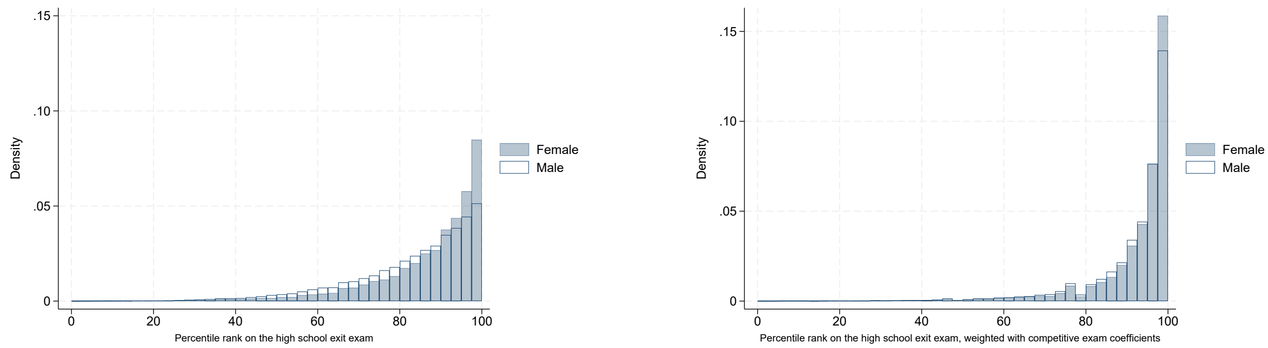
Porter, Jack and Ping Yu, "Regression Discontinuity Designs With Unknown Discontinuity Points: Testing and Estimation," *Journal of Econometrics*, 2015, *189* (1), 132–147.

Saccardo, Silvia, Aniela Pietrasz, and Uri Gneezy, "On the Size of the Gender Difference in Competitiveness," *Management Science*, 2018, *64* (4), 1541–1554.

Schlosser, Analia, Zvika Neeman, and Yigal Attali, "Differential Performance in High Versus Low Stakes Tests: Evidence From the GRE Test," *The Economic Journal*, 2019, *129* (623), 2916–2948.

Appendix

Figure 1.1: Percentile Rank at the High School Graduation Exam of STEM Prep Program Students, by Gender



(a) Raw Percentile Ranks

(b) Percentile Ranks with Re-Weighted GPA

Notes: These histograms illustrate the distribution of percentile ranks of prep program students on the high school graduation exam, by gender. Percentiles are computed among all high school students. In Figure (a), the percentile ranks are derived from the average GPA achieved on the high school graduation exam. In Figure (b), scores are re-weighted with the coefficients of the most selective competitive exams. The coefficients used for this re-weighting can be found in Online Appendix Table 1.A1.

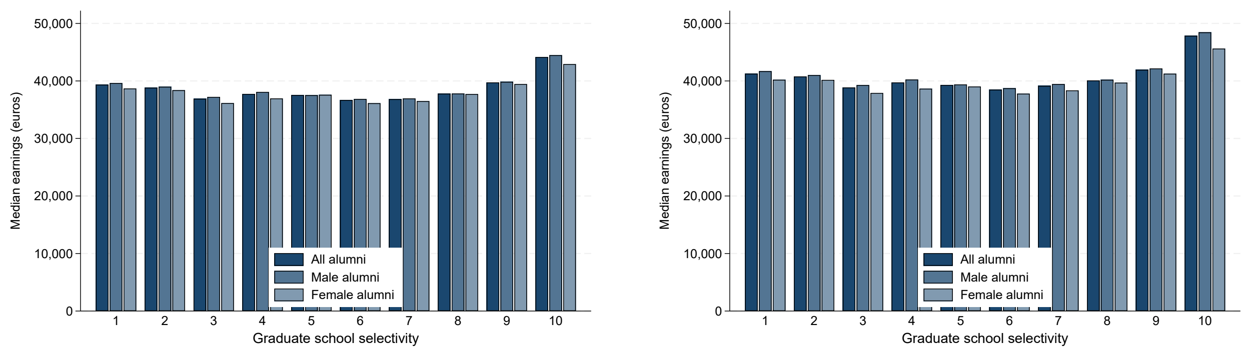
Table 1.1: Access to Star Class, by Gender

	(1)	(2)
	Access to star class	Access to star class
Baseline proba. of access	.52	.52
Female student	-0.083***	0.0013
	(0.014)	(0.0078)
Rank in the class at the end of first year		-0.030***
		(0.00054)
Year FE	✓	✓
Subfield FE	✓	✓
Program FE	✓	✓
N	8,857	8,857

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the effect of gender on access to the star class. Class rank corresponds to a student's GPA rank within their class at the end of the first year of the prep program, just prior to assignment to either a star class or a standard class. GPA is computed using the science coefficients of the most selective entrance exams (see Online Appendix Table 1.A1). We include cohort, subfield and prep program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Figure 1.2: Average Earnings of STEM Graduate School Graduates, by Selectivity Decile



(a) Without Bonuses

(b) With Bonuses

Notes: This figure presents median earnings one year after graduation by decile of STEM graduate school selectivity, separately for male, female, and overall alumni. Panel (a) reports earnings excluding bonuses, and Panel (b) reports earnings including bonuses. Earnings (including bonuses) are based on CTI-reported STEM graduate school medians (2018–2022), based on a compulsory student survey of alumni within each graduate school, and are matched to SCEI cohorts (2015–2023). Missing or unavailable gender-specific data are imputed using adjacent cohorts. Earnings data are missing for 4.6% of schools and 3% of admitted students.

Table 1.2: Comparison of Samples (2015-2023)

	STEM Prep Program Sample (2015-2023)				All students (2016-2017)
	All	Study	Survey (unweighted)	Survey (weighted)	
	(1)	(2)	(3)	(4)	(5)
A. Students					
Female	0.25	0.26	0.27	0.26	0.54
Age	19.6	19.5	19.3	19.3	19.2
	(0.8)	(0.8)	(0.6)	(0.6)	(1.4)
Need-based scholarship holder	0.28	0.26	0.24	0.26	0.38
Mother is high SES	0.48	0.52	0.57	0.54	0.21
Father is high SES	0.62	0.67	0.72	0.68	0.32
High School Graduation Exam					
Highest honors	0.50	0.64	0.77	0.68	0.10
High honors	0.32	0.27	0.17	0.24	0.18
Honors	0.14	0.08	0.05	0.07	0.31
From Paris	0.08	0.09	0.11	0.09	0.04
From Parisian area - outside Paris	0.18	0.16	0.15	0.17	0.15
In a prep program in Paris	0.18	0.23	0.26	0.26	0.10
In a prep program in Parisian area - outside Paris	0.13	0.10	0.21	0.16	0.11
Star class	0.36	0.47	0.52	0.47	–
Repeater	0.19	0.18	0.00	0.00	–
Year of Exams					
2015	0.11	0.11	0.03	–	–
2016	0.11	0.11	0.07	–	–
2017	0.11	0.11	0.10	–	–
2018	0.11	0.11	0.10	–	–
2019	0.11	0.11	0.12	–	–
2020	0.11	0.11	0.13	–	–
2021	0.11	0.11	0.18	–	–
2022	0.11	0.11	0.21	–	–
2023	0.10	0.10	0.07	–	–
B. Prep Programs					
Number of prep programs	200	66	17	–	–
in Paris	23	13	5	–	–
in Parisian area (outside Paris)	31	7	3	–	–
Number of classes	534	262	72	–	–
Star classes	182	130	37	–	–
Number of students	165,450	89,079	9,689	–	1,090,356

Notes: This table compares students across our three samples. Column (1) reports the universe of STEM prep program students from the SCEI administrative data (2015–2023, excluding biology). Column (2) restricts to program–subfield combinations offering both star and standard classes in the second year, our main sample of analysis in this paper. Column (3) presents the subset of students' for which we collected school records from prep-programs in 2022 and 2023 (unweighted), while Column (4) shows the same sample weighted to match the study sample. Out of the school record sample, we extract a balanced sample of students (N=8,857) and a sample used for RDD analysis (N=6,585). Column (5) compares these statistics with all first- and second-year higher-education students in France in 2016–2017, using figures from Bonneau et al. (2021).

Table 1.3: Effect on Expected Earnings

	(1) Aggregated earnings (without bonuses)	(2) Aggregated earnings (without bonuses)	(3) Aggregated earnings (with bonuses)	(4) Aggregated earnings (with bonuses)
Baseline	39,572	39,572	42,056	42,056
Female student × Star class	-545.3***	-463.9***	-938.4***	-773.4***
	(60.6)	(62.1)	(76.6)	(75.0)
Star class	2242.0***	2217.6***	2926.3***	2870.6***
	(57.0)	(56.5)	(80.7)	(78.5)
Female student	-772.1***		-1295.6***	
	(36.9)		(40.9)	
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
Demographic Controls × Female student		✓		✓
MS & HS Exam Score × Female student		✓		✓
Year Fixed-Effects × Female student		✓		✓
Subfield Fixed-Effects × Female student		✓		✓
Program Fixed-Effects × Female student		✓		✓
N	76,616	76,616	76,616	76,616

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in expected earnings, one year after graduation. It corresponds to the median gross annual salary of the most recent cohort of alumni of the school. Earnings are defined at the *school* × *cohort* × *gender* level and are sourced from the CTI website, from a compulsory student survey of alumni conducted in each STEM graduate school. When the salary was missing for a specific cohort, we inferred it from the adjacent cohort. For the most recent cohorts that haven't graduated yet, we inferred the salary based on the previous cohort and the average salary increase over year for this specific graduate school. Columns (1) and (2) correspond to salary without bonuses, while Columns (3) and (4) correspond to salary with bonuses. In Columns (2) and (4), we include interactions of all controls and fixed-effects with a gender dummy variable, allowing observable characteristics to have gender-specific effects. Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Table 1.4: Placebo Test: High School Graduation Exam and First Year Prep Program Percentile Ranks

	(1) Pct. rank raw GPA HS grad. exam	(2) Pct. rank raw GPA HS grad. exam	(3) Pct. rank re-weighted GPA HS grad. exam	(4) Pct. rank re-weighted GPA HS grad. exam	(5) Pct. Rank GPA prep prog. 1st yr – 1st sem	(6) Pct. Rank GPA prep prog. 1st yr – 1st sem
Baseline percentile rank	0.84	0.84	0.92	0.92	0.54	0.54
Female student × Star class	0.00030	0.00011	0.0019	0.0019	-0.00096	-0.0064
	(0.0017)	(0.0018)	(0.0015)	(0.0016)	(0.010)	(0.0094)
Star class	0.077***	0.078***	0.057***	0.058***	0.41***	0.41***
	(0.0017)	(0.0017)	(0.0016)	(0.0016)	(0.0056)	(0.0053)
Female student	0.041***		0.0097***		-0.016**	
	(0.0013)		(0.0012)		(0.0074)	
Demographic Controls	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓	✓	✓
Demographic Controls × Female student		✓		✓		✓
Year FE × Female student		✓		✓		✓
Track FE × Female student		✓		✓		✓
Program FE × Female student		✓		✓		✓
N	82,352	82,352	82,356	82,356	8,857	8,857

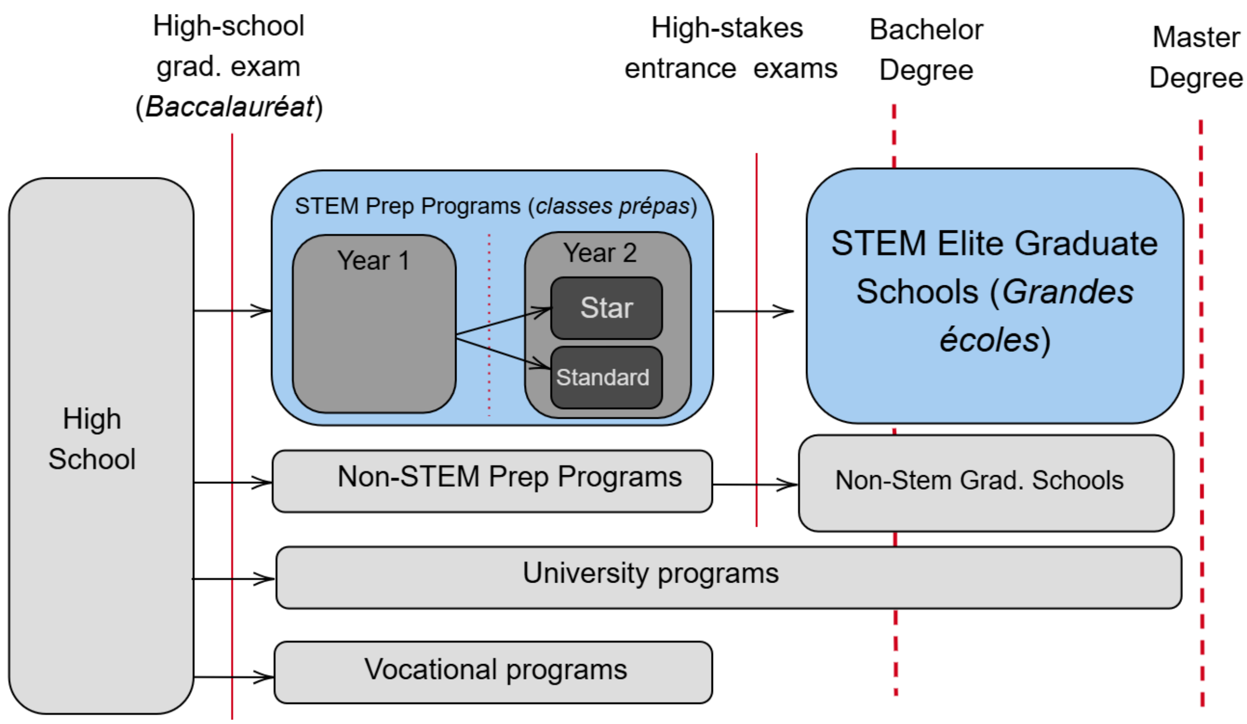
Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports changes in percentile rank at the high school graduation exam (Columns 1–4) and during the first year in the preparatory program (Columns 5–6) for female students in star classes. In Columns (1) and (2), ranks are computed using raw GPA among all high school graduates; in Columns (3) and (4), ranks are based on a re-weighted GPA, with weights given by the coefficients of the most selective competitive exams (see Table 1.A1 in the online appendix). Columns (5) and (6) report within-class percentile ranks for first-year prep program students during the first semester. Control variables include geographic origin (Paris/Paris area), low-income status, parental socio-economic status (four categories), French nationality, repeater status, disability status, high school academic track (science vs. other), specialization (engineering science vs. computer science), and the gender composition of the *subfield* × *program* × *cohort*. All specifications include cohort, subfield, and program fixed effects. In Columns (2), (4), and (6), all controls and fixed effects are interacted with a gender dummy, allowing observable characteristics to have gender-specific effects on performance. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Online Appendix

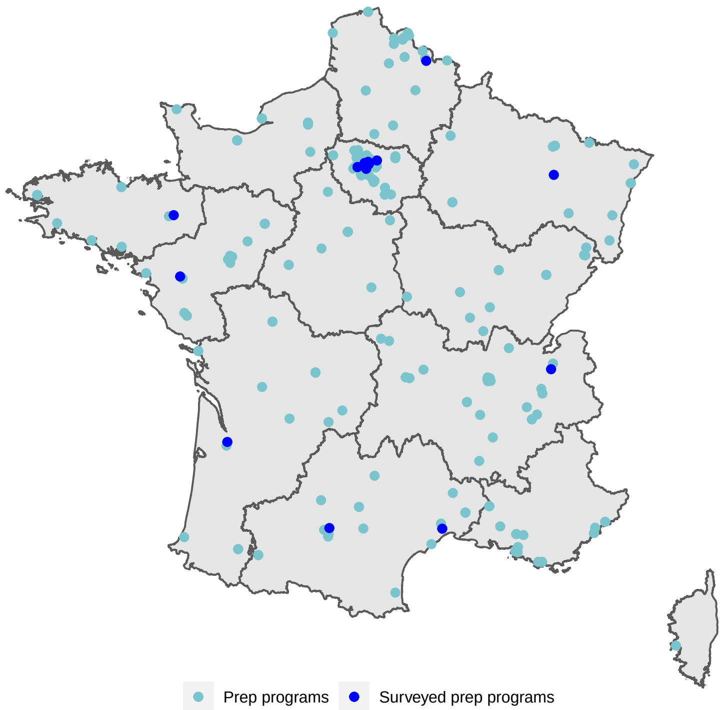
1.A Detailed Institutional Background

Figure 1.A1: Simplified Diagram of French Higher Education

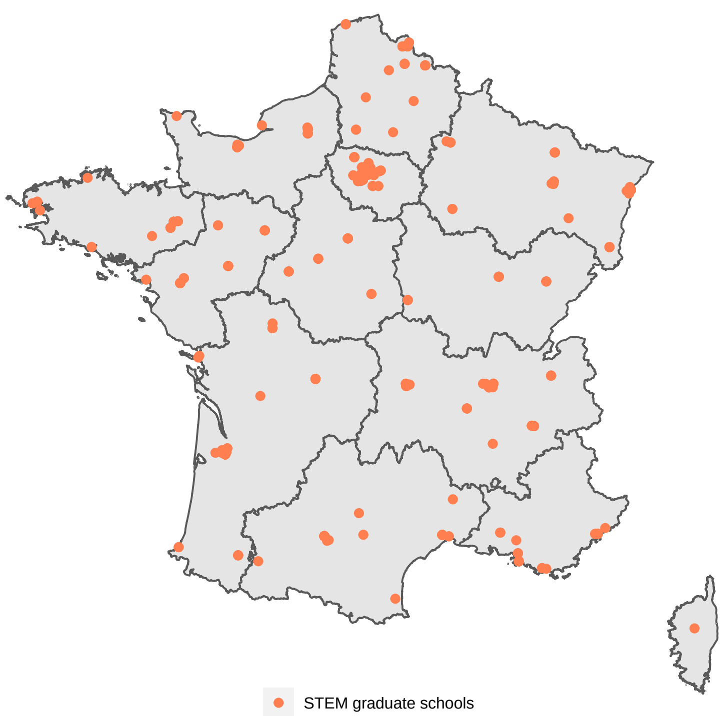


Notes: This diagram shows the organization of France's main higher education tracks. The programs we study here are STEM prep programs and STEM elite graduate schools, highlighted in blue.

Figure 1.A2: Maps of Preparatory Programs and STEM Graduate Schools



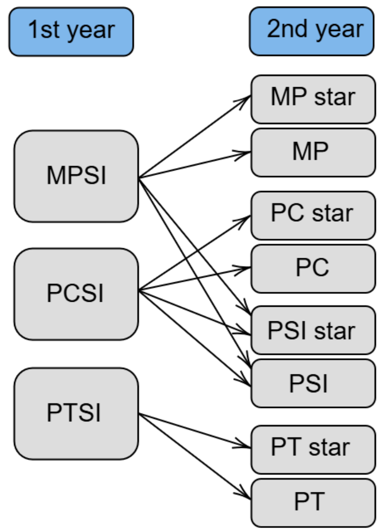
(a) STEM Preparatory Programs



(b) STEM Graduate Schools

Notes: These maps illustrate the geographical distribution of STEM prep programs across France (Panel a) and of STEM graduate schools (Panel b). Prep programs shown in dark blue in Panel a are those from which we collected school-record data.

Figure 1.A3: Organization Chart of STEM Prep Programs



Notes: This diagram illustrates the various subfields within STEM prep programs and the links between them from the first to the second year of the program. The boxes on the left represent the first year, while those on the right correspond to the second year of prep program. Although additional STEM preparatory programs exist (such as *BCPST*, *TB*, *TPC*, and *TSI*), they are excluded from the analysis due to the absence of tracking between standard and star classes in their second year.

Table 1.A1: Most Selective Competitive Entrance Exam Coefficients

	Subfield 1 Math-Physics	Subfield 2 Physics-Chemistry	Subfield 3 Physics-Engineering Sciences	Subfield 4 Engineering Sciences
Math.	0.36	0.24	0.28	0.18
Physics & Chemistry	0.31	–	0.32	0.18
Physics	–	0.31	–	–
Chemistry	–	0.17	–	–
Engineering Science	0.03	–	0.13	0.31
Computer Science	0.06	0.03	0.04	0.06
French	0.13	0.13	0.11	0.15
Foreign Language	0.12	0.12	0.13	0.13

Notes: This table reports the coefficients from competitive entrance exams used to construct weighted GPAs for the high school graduation exam (all subjects) and the first-year preparatory program GPA (all scientific subjects). The coefficients are based on mean values from the three most selective exam consortia (*Banque X-ENS*, *Banque Mines-Ponts*, and *Banque Centrale*). When particular subjects, such as computer science or engineering science, are missing, the coefficients are adjusted to account for their absence.

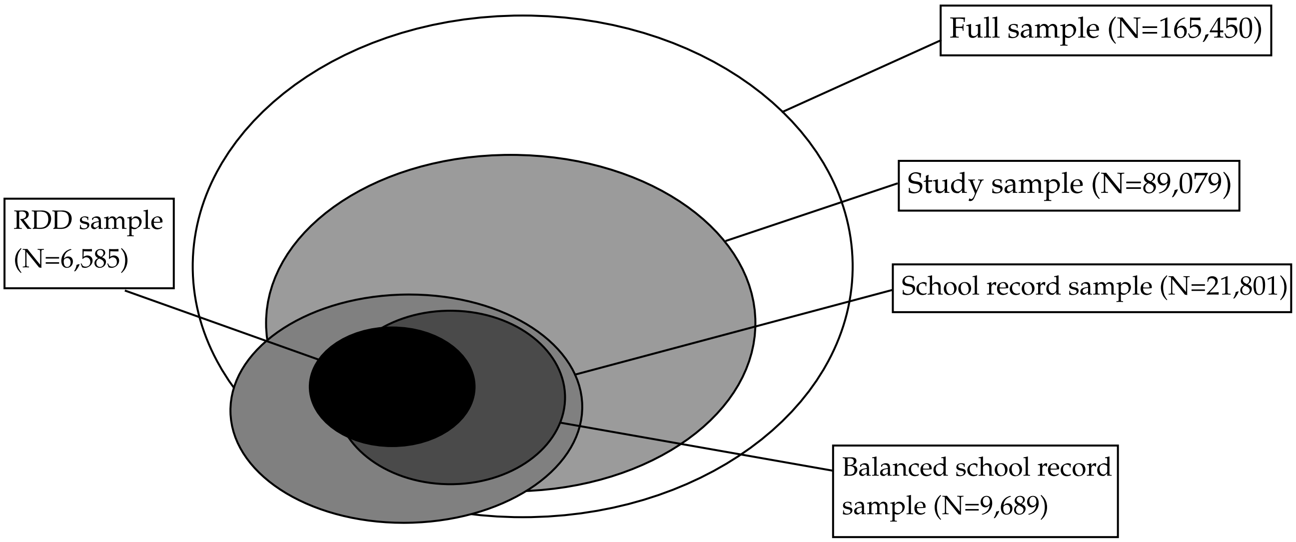
Table 1.A2: STEM Graduate Schools in the Top 10% of Selectivity or Desirability

	Selectivity				Desirability (all applicants)				Desirability (female applicants)			
	MP	PC	PSI	PT	MP	PC	PSI	PT	MP	PC	PSI	PT
Centrale Lyon	X	X	X	X	X	X	X	X	X	X	X	X
CentraleSupélec	X	X	X	X	X	X	X	X	X	X	X	X
ISAE - SUPAERO Toulouse	X	X	X	X	X	X	X	X	X	X	X	X
Mines de Paris	X	X	X	X	X	X	X	X	X	X	X	X
Polytechnique	X	X	X	X	X	X	X	X	X	X	X	X
Ponts ParisTech	X	X	X	X	X	X	X	X	X	X	X	X
ENSTA ParisTech	X	X	X	X	X	X	X	X	X	X	X	
ENS Ulm	X	X	X		X	X	X		X	X		
ENSAE Paris	X	X	X									
ENS de Lyon	X	X			X	X			X	X		
Télécom ParisTech	X		X		X	X	X		X	X	X	
Mines de Nancy	X			X			X				X	
Ecole Météorologie	X											
Centrale Nantes		X	X	X	X	X	X	X	X	X	X	X
ESPCI Paris		X				X				X		
Chimie ParisTech		X										
ENS Cachan Paris-Saclay			X		X	X			X	X		
Centrale Lille					X		X	X	X		X	X
Mines de Saint-Etienne											X	
Arts et Métiers												X
Total number of schools	130	134	128	96	130	134	128	96	130	134	128	96

Notes: This table enumerates the STEM graduate schools which are in the top 10% of selectivity, desirability and desirability for female students. In the first four columns, selectivity of STEM graduate schools is measured using the average percentile rank of admitted students at the high school graduation exam. The subsequent eight columns use the revealed preferences of applicants as a metric for desirability, following the approach of (Avery et al. (2013)). This method captures the desirability of STEM graduate schools, where more sought-after institutions garner more applications and more often appear higher on students' rank-ordered lists. The final four columns consider only female applicants when calculating school desirability. For all the definitions, they are consistently defined on a subfield-by-subfield basis, acknowledging that not all subfields lead to the exact same STEM graduate schools, even if many overlap.

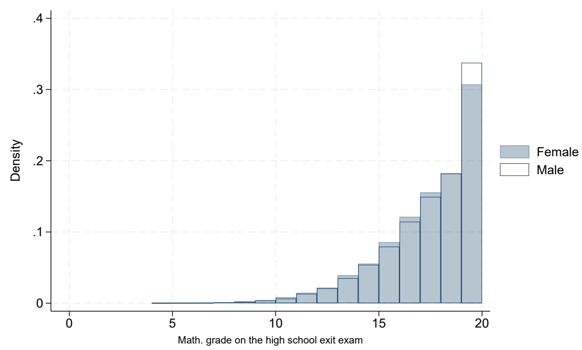
1.B Complementary Descriptive Statistics

Figure 1.B1: Diagram of the Different Samples of Analysis

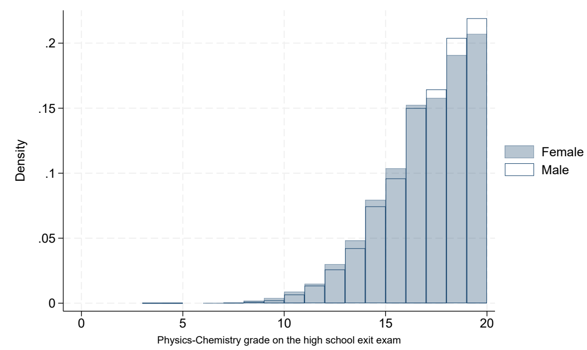


Notes: This diagram depicts the various subsets of data used in our study. The 'full sample' corresponds to the universe of applicants to STEM graduate schools from 2015 to 2023, excluding the biology subfield. The 'study sample' corresponds to the restriction of the full sample to combinations of *program* $\times$ *subfield* that have both a standard and a star class in the second year of the program, which we use in our main analysis. The 'school record sample' corresponds to the sample for which we collected school records. The location of these programs can be found on the map in Figure 1.A2. The 'matched and balanced school record sample' corresponds to students for which we have balanced school record (from begining to end of prep program) and we were able to statistically match with the administrative data from applications to STEM graduate schools. The 'RDD sample' corresponds to the sample used for the regression discontinuity analysis.

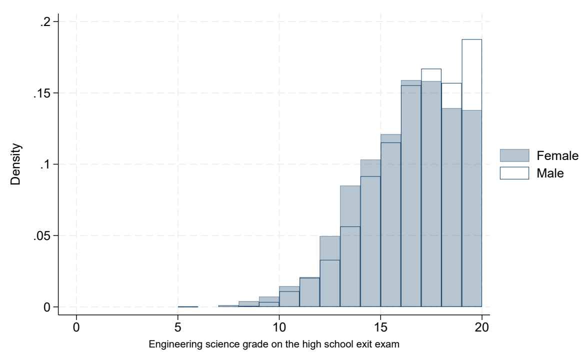
Figure 1.B2: Density of Raw Grades at the High School Graduation Exam of STEM Prep Program Students, by Gender



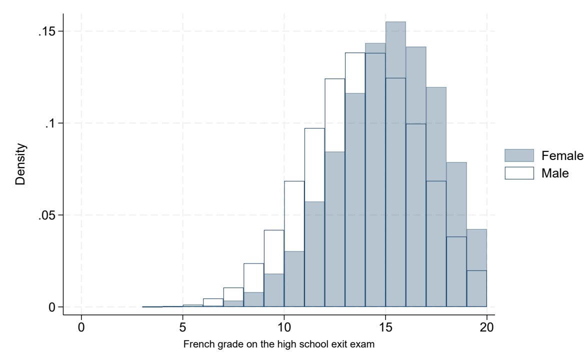
(a) Math.



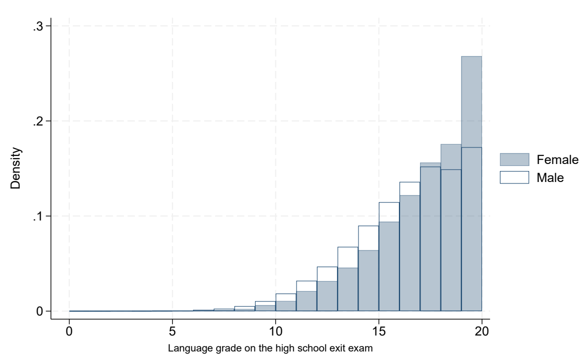
(b) Physics and Chemistry



(c) Engineering Science



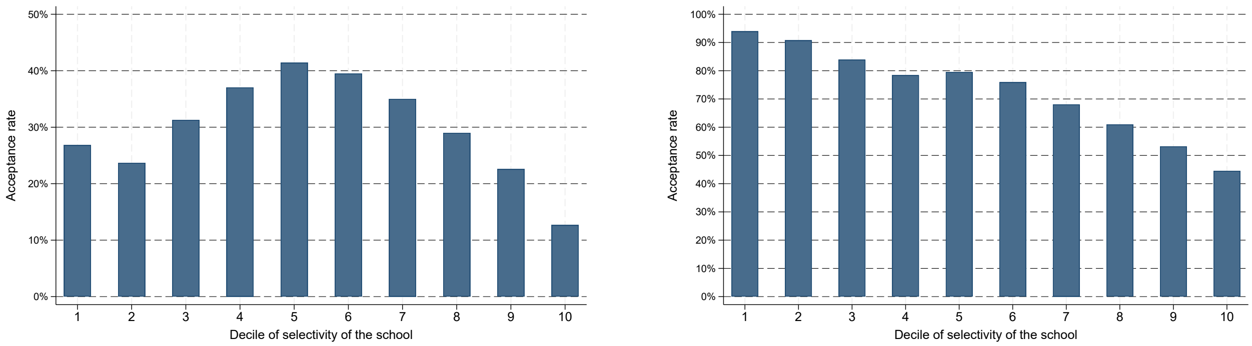
(d) French



(e) Foreign Language

Notes: These histograms depict the distribution of raw grades achieved at the high school graduation exam by prep program students and by gender. Grades in engineering science are available for only approximately 20 percent of the entire sample; this is because the majority of students do not take up engineering science in high school.

Figure 1.B3: Acceptance Rate, by STEM Graduate School Selectivity Decile

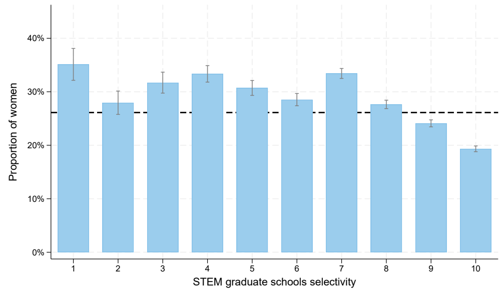


(a) All applications

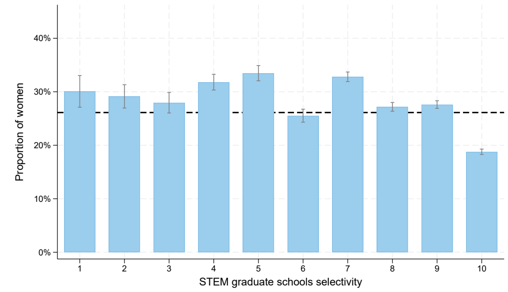
(b) Students ranked at the entrance exam

Notes: These figures illustrates the acceptance rate of graduate schools, by graduate schools selectivity decile. The acceptance rate is computed as the ratio of candidates who met or exceeded the admission criteria — meaning their rank at the entrance exam was below the admission threshold — over the overall number of applicants to the school (Panel a) or the number of applicants ranked at one of the entrance exam to the school (Panel b). Acceptance rate captures offers of admission and not actual enrollment. Consequently, while a student could theoretically receive offers from multiple schools, they will ultimately register at only one institution.

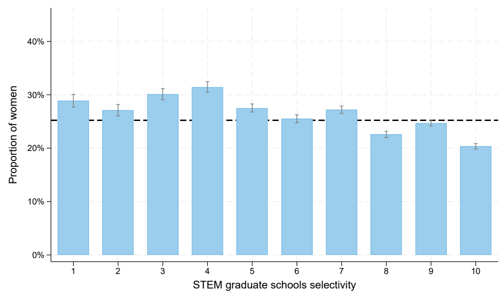
Figure 1.B4: Proportion of Female Students, by Decile of Selectivity or Desirability of STEM Graduate Schools and Subfields



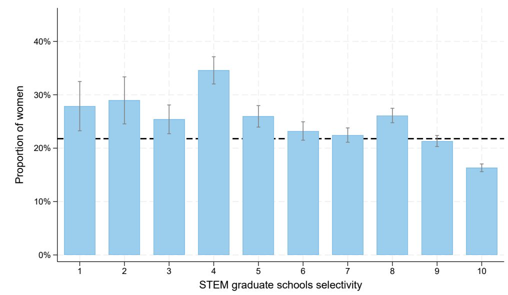
(a) Desirability



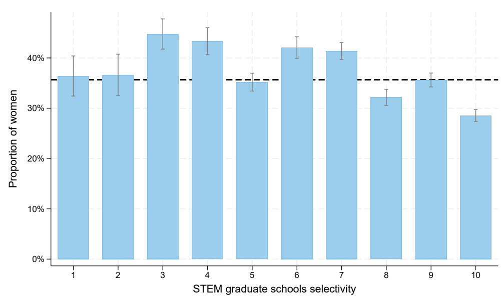
(b) Desirability for female students



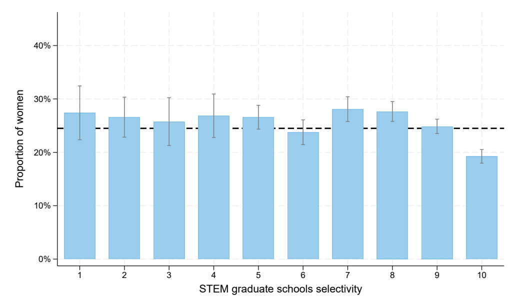
(c) Full Sample



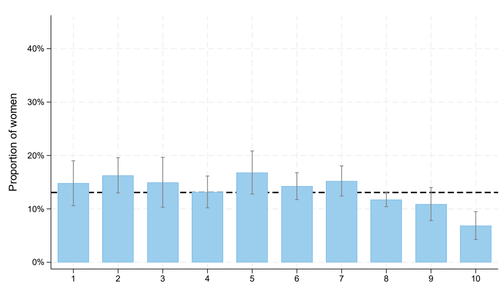
(d) Subfield 1 (Math-Physics)



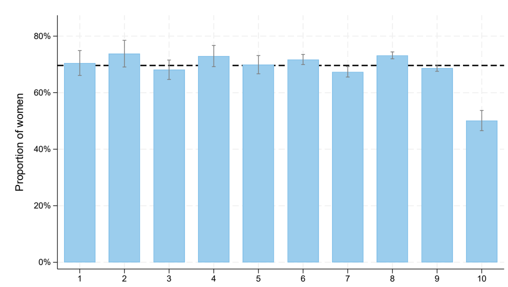
(e) Subfield 2 (Physics-Chemistry)



(f) Subfield 3 (Physics-Engineering Sciences)



(g) Subfield 4 (Engineering Sciences)



(h) Subfield 5 (Biology)

Notes: This figure shows the proportion of female students admitted to STEM graduate schools by decile of graduate school selectivity: (i) using school desirability and female-specific desirability based on revealed preferences (Avery et al. (2013)); (ii) for the full applicant sample without restricting to program–field combinations with star classes; (iii) for our study sample across the four main subfields; and (iv) for the biology subfield, which is otherwise excluded from our analysis because it does not feature tracking between star and standard classes. The dashed line indicates the average proportion of female students. Selectivity is measured in panels (a)–(b) using school desirability (Avery et al. (2013)) and in panels (c)–(h) using the average percentile rank of admitted students on the high school graduation exam, defined separatly by subfield for both definitions.

1.C Complementary Descriptive Analysis

1.C.1 Complementary Decomposition Results

Table 1.C1: Decomposition of the Gender Gap in Admission to the Top 10% Most Selective STEM Graduate Schools

	Receive an offer from top 10% most selective graduate schools						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female student	-0.057***	-0.068***	-0.040***	-0.021***	-0.013*	-0.006	-0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.005)	(0.005)
Decomposition % explained by add. control		-19%	49%	33%	14%	12%	-2%
Controls:							
Demographics and Fixed-effects Demographic characteristics Program, track, and year FE	✓	✓	✓	✓	✓	✓	✓
Previous ability HS exam honors		✓	✓	✓	✓	✓	✓
Beginning of program 1st year/1st sem. grades FE			✓	✓	✓	✓	✓
End of program 2nd year/2nd sem. grades FE Star class				✓	✓	✓	✓
Applications Apply to top school					✓	✓	✓
D-Day Effect Percentile rank at top school exams						✓	✓
Preferences Top schools in ROLs							✓
N	8,857	8,857	8,857	8,857	8,857	8,857	8,857
Adj-R^2	0.345	0.402	0.495	0.554	0.581	0.757	0.760

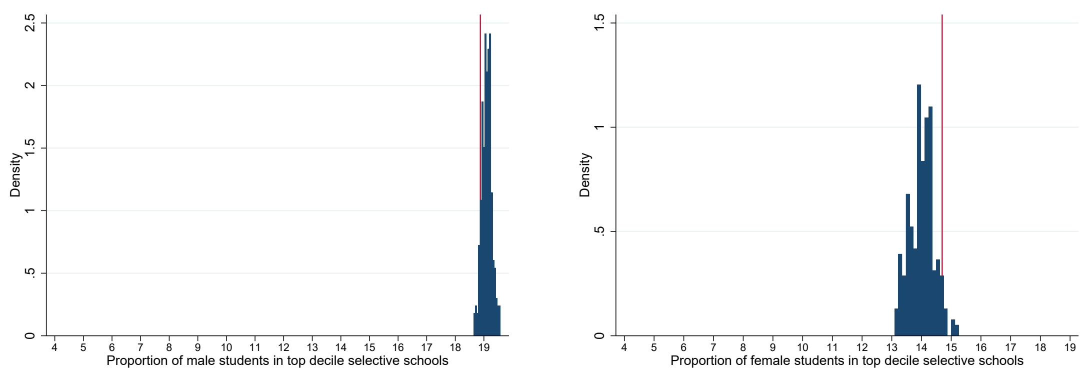
Standard errors in parentheses
$^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Notes: This table shows the evolution of the gender gap in admission to the top 10% most selective STEM graduate schools when adding various controls. We use the reweighted subsample of students for whom we have collected prep program scohol records. The decomposition is computed by observing how much the gender gap is reduced when an additional control is added, compared to the raw gender gap observed in the first column. Selectivity of graduate schools is defined by the average percentile rank of admitted students at the high school graduation exam. Demographic controls include geographic origin (Paris or the Parisian area), low-income status, the socio-economic status of each parent (categorized into four groups for each parent), French nationality, disability status, whether the student was in a science academic track during high school. The baseline specification already includes those demographic characteristics as well as cohort, subfield, and prep program fixed effects. Previous ability is measured using decile indicators of GPA, as well as quintile indicators of grades in each core subject in the high school and middle school graduation exams. Prep-program ability is measured using decile indicators of GPA and quintile indicators of grades in each core subject in the first and last semesters, along with a dummy indicator for star-class status. Applications to top graduate schools are controlled by dummies for application to each of the exams leading to top graduate schools, defined separately by subfield. D-Day effect refers to performance during competitive entrance exams and is controlled for by (i) dummies for whether or not the applicant is ranked at competitive exams leading to top graduate schools and (ii) the percentile rank of their rank in those exams, controlled linearly for each entrance exam. Preferences are controlled for by adding dummy variables for whether or not each of the top graduate schools is ranked among students' ranked-ordered lists of schools and the percentile rank of the top graduate schools in their rankings. Controls follow the chronological order of students' decisions.

1.C.2 Complementary Results on Students Preferences

To gauge gender differences in preferences more precisely, we estimate male students' preferences using a rank-order logit model (Equation 1) and predict valuations for both genders. We then run 300 simulations of the matching algorithm using different error vectors (with $\epsilon_{i,j}$ being independent and identically distributed (i.i.d.) according to a Type I extreme value (Gumbel) distribution), observing the gender distribution in the top 10% most selective schools. Results suggest that if women had the same preferences as men, their representation in the top 10% most selective graduate schools might actually slightly decrease (Figure 1.C5). This could be due to intensified competition for the same STEM graduate schools, considering that women tend to underperform in high-stakes exams. Current minimal gender differences in preferences might benefit women's representation in top STEM graduate schools.

Figure 1.C5: Admission to the 10% Most Selective Graduate Schools When Estimating Student Preferences From Male Ones



(a) Male students

(b) Female students

Notes: These figures depict results from 300 counterfactual simulations of the graduate school-student matching algorithm, using preferences of male students to predict the preferences of all students. The red bars display the current proportion of male and female students accessing the top 10% most selective STEM graduate schools; the blue bars display this proportion under the counterfactual simulation scenario. Specifically, we estimate a rank-ordered logit model based on graduate schools ranked in the ROLs by male students and predict the school valuation for all students. We then generate 300 different error vectors, each distributed according to a Type I extreme value (Gumbel) distribution, and re-run the graduate school-student matching algorithm 300 times.

1.D Double Difference: Complementary Results

Table 1.D1: Admission to STEM Graduate Schools in the Top 10 Percent of Selectivity: Coefficients on Control Variables

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Desirability)
Baseline proba. of access	0.22	0.23
Female student $\times$ Star class	-0.042***	-0.058***
	(0.0061)	(0.0060)
Star class	0.27***	0.29***
	(0.0061)	(0.0061)
Female student	-0.027***	-0.028***
	(0.0031)	(0.0030)
From Paris	0.052***	0.049***
	(0.0065)	(0.0064)
From parisian area	0.019***	0.016***
	(0.0056)	(0.0054)
Need-based scholarship students	-0.0093***	-0.0087***
	(0.0028)	(0.0028)
Repeater	0.055***	0.061***
	(0.0036)	(0.0037)
Disabled student	-0.0076	-0.0016
	(0.0087)	(0.0087)
Scientific high-school background	-0.0070	0.0031
	(0.0080)	(0.0086)
Student with French nationality	0.0016	0.010
	(0.0063)	(0.0065)
Engineering science option	0.031**	0.050***
	(0.013)	(0.014)
Computer science option	0.026*	0.052***
	(0.013)	(0.015)
Proportion of female students in the track	-0.00014	-0.00043
	(0.00032)	(0.00033)
Father High SES	0.023***	0.031***
	(0.0056)	(0.0054)
Father Medium High SES	0.010	0.019***
	(0.0063)	(0.0064)
Father Medium Low SES	-0.0027	0.0027
	(0.0057)	(0.0057)
Father Low SES	0.0042	0.0074
	(0.0061)	(0.0061)
Mother High SES	0.022***	0.021***
	(0.0044)	(0.0043)
Mother Medium High SES	-0.00048	0.0010
	(0.0050)	(0.0050)
Mother Medium Low SES	-0.0033	-0.0064
	(0.0046)	(0.0045)
Mother Low SES	-0.0066	-0.0066
	(0.0052)	(0.0053)
MS & HS Exam Score	✓	✓
Year Fixed Effects	✓	✓
Subfield Fixed Effects	✓	✓
Program Fixed Effects	✓	✓
N	89,065	89,065

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table displays the change in probability of admission to top 10% STEM graduate schools for female students in star classes. Selectivity of graduate schools is measured in two ways: Column (1) uses the average percentile rank of admitted students at the high school graduation exam, while Column (2) is based on the revealed preferences of applicants (Avery et al. (2013)). Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* $\times$ *program* $\times$ *cohort* level.

1.D.1 Heterogeneity

Table 1.D2: Heterogeneity by Income Status: Admission to STEM Graduate Schools in the Top 10% of Selectivity

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Selectivity)	(3) Top 10% grad. schools (Desirability)	(4) Top 10% grad. schools (Desirability)
Income status	Low-income	High-income	Low-income	High-income
Baseline proba. of access	0.14	0.25	0.16	0.26
Female student × Star class	-0.054***	-0.038***	-0.072***	-0.053***
	(0.010)	(0.0070)	(0.011)	(0.0070)
Star class	0.23***	0.29***	0.24***	0.30***
	(0.0075)	(0.0064)	(0.0079)	(0.0063)
Female student	-0.022***	-0.030***	-0.020***	-0.032***
	(0.0040)	(0.0037)	(0.0041)	(0.0037)
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
N	23,409	65,656	23,409	65,656

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in probability of admission to top 10% STEM graduate schools for female students in star classes, by parental income status. Low income students are need-based scholarships students and high-income students are non need-based scholarships students. Selectivity of graduate schools is measured in two ways: Columns (1) and (2) use the average percentile rank of admitted students at the high school graduation exam, while Columns (3) and (4) are based on the revealed preferences of applicants (Avery et al. (2013)). Demographic controls include geographic origin (Paris or Parisian area), the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Table 1.D3: Heterogeneity by Ability: Admission to STEM Graduate Schools in the Top 10% of Selectivity

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Desirability)	(3) Top 10% grad. schools (Selectivity)	(4) Top 10% grad. schools (Desirability)	(5) Top 10% grad. schools (Selectivity)	(6) Top 10% grad. schools (Desirability)	(7) Top 10% grad. schools (Selectivity)	(8) Top 10% grad. schools (Desirability)
Quartile of ability	Q1	Q1	Q2	Q2	Q3	Q3	Q4	Q4
Baseline proba. of access	0.02	0.03	0.11	0.12	0.25	0.27	0.49	0.51
Female student × Star class	-0.036***	-0.042***	-0.081***	-0.092***	-0.082***	-0.11***	-0.035**	-0.049***
	(0.0076)	(0.0081)	(0.0099)	(0.0100)	(0.012)	(0.012)	(0.013)	(0.013)
Star class	0.078***	0.092***	0.21***	0.22***	0.34***	0.35***	0.44***	0.45***
	(0.0051)	(0.0056)	(0.0071)	(0.0072)	(0.0077)	(0.0078)	(0.0087)	(0.0087)
Female student	-0.0048***	-0.0063***	-0.0070*	-0.013***	-0.016**	-0.016**	-0.057***	-0.051***
	(0.0016)	(0.0018)	(0.0040)	(0.0042)	(0.0062)	(0.0067)	(0.010)	(0.0099)
Demographic Controls	✓	✓	✓	✓	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
N	20,587	20,587	20,591	20,591	20,585	20,585	20,593	20,593

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in probability of admission to top 10% STEM graduate schools for female students in star classes, by prior ability status. We define ability based on re-weighted GPA at the high school graduation exam (with weights equal to competitive exams coefficients), and present results for each quartile. Selectivity of graduate schools is measured in two ways: Columns (1); (3), (5) and (7) use the average percentile rank of admitted students at the high school graduation exam, while Columns (2); (4), (6) and (8) are based on the revealed preferences of applicants (Avery et al. (2013)). Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Table 1.D4: Admission to STEM Graduate Schools in the Top 10% of Selectivity, by Repeater Status

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Selectivity)	(3) Top 10% grad. schools (Desirability)	(4) Top 10% grad. schools (Desirability)
Repeater status	Non-repeaters	Repeaters	Non-repeaters	Repeaters
Baseline proba. of access	0.22	0.23	0.23	0.25
Female student × Star class	-0.042***	-0.046***	-0.057***	-0.062***
	(0.0065)	(0.014)	(0.0063)	(0.014)
Star class	0.27***	0.28***	0.29***	0.29***
	(0.0064)	(0.0088)	(0.0064)	(0.0091)
Female student	-0.026***	-0.030***	-0.026***	-0.033***
	(0.0032)	(0.0065)	(0.0033)	(0.0071)
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
N	72,728	16,268	72,728	16,268

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in probability of admission to top 10% STEM graduate schools for female students in star classes, by repeater status. Selectivity of graduate schools is measured in two ways: Columns (1) and (2) use the average percentile rank of admitted students at the high school graduation exam, while Columns (3) and (4) are based on the revealed preferences of applicants (Avery et al. (2013)). Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Table 1.D5: Admission to STEM Graduate Schools in the Top 10% of Selectivity, by Proportion of Women

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Selectivity)	(3) Top 10% grad. schools (Desirability)	(4) Top 10% grad. schools (Desirability)
Proportion of women	Below median	Above median	Below median	Above median
Baseline proba. of access	0.22	0.22	0.24	0.22
Female student × Star class	-0.035***	-0.051***	-0.033***	-0.074***
	(0.0091)	(0.0081)	(0.0089)	(0.0081)
Star class	0.26***	0.29***	0.28***	0.29***
	(0.0088)	(0.0083)	(0.0090)	(0.0082)
Female student	-0.034***	-0.016***	-0.038***	-0.017***
	(0.0045)	(0.0043)	(0.0043)	(0.0043)
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
N	44,303	44,762	44,303	44,762

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in probability of admission to top 10% STEM graduate schools for female students in star classes, by proportion of women. Proportion of women is computed at the *subfield* × *program* × *cohort* level. Selectivity of graduate schools is measured in two ways: Columns (1) and (2) use the average percentile rank of admitted students at the high school graduation exam, while Columns (3) and (4) are based on the revealed preferences of applicants (Avery et al. (2013)). Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

1.D.2 Robustness Checks

Table 1.D6: Robustness Check: Other Measures of Graduate School Selectivity

	(1) Top 10% grad. schools (Desirability for female students)	(2) Top 20% grad. schools (Selectivity)	(3) Top 20% grad. schools (Desirability)	(4) Average selectivity	(5) Average desirability
Baseline proba. of access	0.26	0.39	0.42	83	0.95
Female student × Star class	-0.068***	-0.022***	-0.044***	-0.40***	-0.13***
	(0.0062)	(0.0064)	(0.0066)	(0.095)	(0.012)
Star class	0.31***	0.34***	0.35***	6.91***	0.92***
	(0.0057)	(0.0054)	(0.0050)	(0.073)	(0.011)
Female student	-0.022***	-0.033***	-0.037***	-0.69***	-0.11***
	(0.0031)	(0.0037)	(0.0039)	(0.076)	(0.0074)
Demographic Controls	✓	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓	✓
N	89,065	89,065	89,065	78,011	78,011

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in average selectivity of graduate school of admission and the probability of admission to top quintile STEM graduate schools for female students in star classes. Selectivity of graduate schools is measured in two ways: Columns (2) and (4) use the average percentile rank of admitted students at the high school graduation exam, while Columns (1), (3) and (5) are based on the revealed preferences of applicants (Avery et al. (2013)). Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Table 1.D7: Probit Model: Admission to STEM Graduate Schools in the Top 10 Percent of Selectivity

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Desirability)
Baseline proba. of access	0.22	0.23
Female student × Star class	-0.095***	-0.12***
	(0.034)	(0.032)
Star class	1.47***	1.47***
	(0.020)	(0.019)
Female student	-0.21***	-0.24***
	(0.030)	(0.027)
Demographic Controls	✓	✓
MS & HS Exam Score	✓	✓
Year Fixed Effects	✓	✓
Subfield Fixed Effects	✓	✓
Program Fixed Effects	✓	✓
N	88,424	88,424

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table displays the change in probability of admission to top 10% STEM graduate schools for female students in star classes. Selectivity of graduate schools is measured in two ways: Column (1) uses the average percentile rank of admitted students at the high school graduation exam, while Column (2) is based on the revealed preferences of applicants (Avery et al. (2013)). Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

Table 1.D8: Robustness Check: Interaction of Fixed Effects

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Selectivity)	(3) Top 10% grad. schools (Desirability)	(4) Top 10% grad. schools (Desirability)
Baseline proba. of access	0.22	0.22	0.23	0.23
Female student $\times$ Star class	-0.042***	-0.040***	-0.058***	-0.057***
	(0.0061)	(0.0061)	(0.0060)	(0.0061)
Star class	0.27***	0.28***	0.29***	0.29***
	(0.0061)	(0.0062)	(0.0061)	(0.0062)
Female student	-0.027***	-0.052***	-0.028***	-0.056***
	(0.0031)	(0.0087)	(0.0030)	(0.0087)
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
Year $\times$ Subfield FE		✓		✓
Year $\times$ Program FE		✓		✓
Subfield $\times$ Program FE		✓		✓
Year $\times$ Subfield $\times$ Program FE		✓		✓
N	89,065	89,065	89,065	89,065

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table reports the change in probability of admission to top 10% STEM graduate schools for female students in star classes. Selectivity of graduate schools is measured in two ways: Columns (1) and (2) use the average percentile rank of admitted students at the high school graduation exam, while Columns (3) and (4) are based on the revealed preferences of applicants (Avery et al. (2013)). Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* $\times$ *program* $\times$ *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. We include *cohort* $\times$ *subfield*, *cohort* $\times$ *program*, *subfield* $\times$ *program*, and *cohort* $\times$ *subfield* $\times$ *program* fixed effects in Columns (2) and (4). Standard errors are clustered at the *subfield* $\times$ *program* $\times$ *cohort* level.

Table 1.D9: Admission to STEM Graduate Schools in the Top 10 Percent of Selectivity, Controlling for Number of Selective Exams Taken

	(1) Top 10% grad. schools (Selectivity)	(2) Top 10% grad. schools (Selectivity)	(3) Top 10% grad. schools (Desirability)	(4) Top 10% grad. schools (Desirability)
Baseline proba. of access	0.22	0.22	0.23	0.23
Female student × Star class	-0.044***	-0.028***	-0.060***	-0.041***
	(0.0058)	(0.0067)	(0.0058)	(0.0066)
Star class	0.23***	0.23***	0.25***	0.24***
	(0.0062)	(0.0064)	(0.0063)	(0.0065)
Female student	-0.022***		-0.022***	
	(0.0030)		(0.0030)	
Demographic Controls	✓	✓	✓	✓
MS & HS Exam Score	✓	✓	✓	✓
FE Number of Selective Exams Taken	✓	✓	✓	✓
Year Fixed Effects	✓	✓	✓	✓
Subfield Fixed Effects	✓	✓	✓	✓
Program Fixed Effects	✓	✓	✓	✓
Demographic Controls × Female student		✓		✓
MS & HS Exam Score × Female student		✓		✓
FE # of Selective Exams Taken × Female student		✓		✓
Year Fixed-Effects × Female student		✓		✓
Track Fixed-Effects × Female student		✓		✓
Program Fixed-Effects × Female student		✓		✓
N	89,065	89,065	89,065	89,065

Standard errors in parentheses

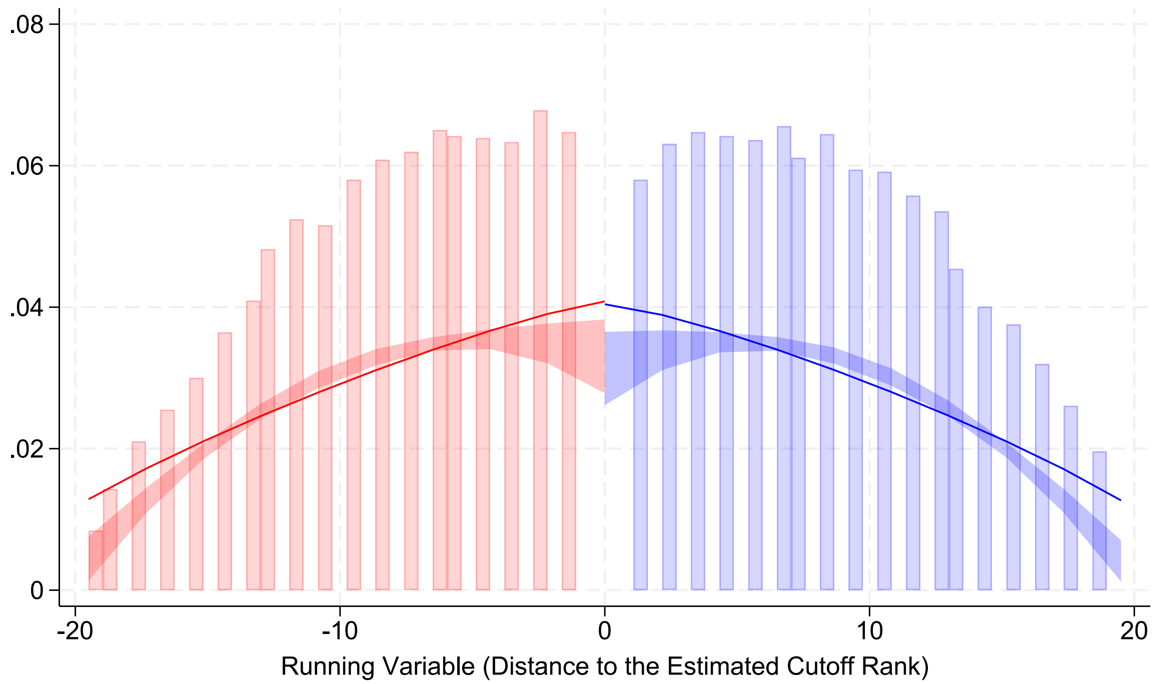
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: The table illustrates the change in the probability of admission to top 10% STEM graduate schools for female students in star classes, while controlling for the number of selective entrance exams taken. We include fixed effects for the number of (1) very selective exams taken (i.e., those leading to only top-tier schools) and (2) selective exams taken (i.e., those leading to at least one top-tier school). Selectivity of graduate schools is measured in two ways: Columns (1) and (2) use the average percentile rank of admitted students at the high school graduation exam, while Columns (3) and (4) are based on the revealed preferences of applicants (Avery et al. (2013)). In Columns (2) and (4), we include interactions of all controls and fixed-effects with a gender dummy variable, allowing observable characteristics to have gender-specific performance impacts. Demographic controls include geographic origin (Paris or Parisian area), low-income status, the socio-economic status of each parent (into four categories), French nationality, repeater status, disability status, whether the student was in a science academic track during high school, the student's option (either engineering science or computer science), and the gender composition in the *subfield* × *program* × *cohort*. Previous ability is controlled by (i) decile rank dummy variables from both high school and middle school graduation exams GPA, and (ii) quintile rank dummy variables for each grades at the high and middle school graduation exams at the subjects studied in prep programs — mathematics, physics and chemistry, engineering science, French (both written and oral), and foreign languages. We include cohort, subfield and program fixed effects. Standard errors are clustered at the *subfield* × *program* × *cohort* level.

1.E Regression Discontinuity: Complementary Results

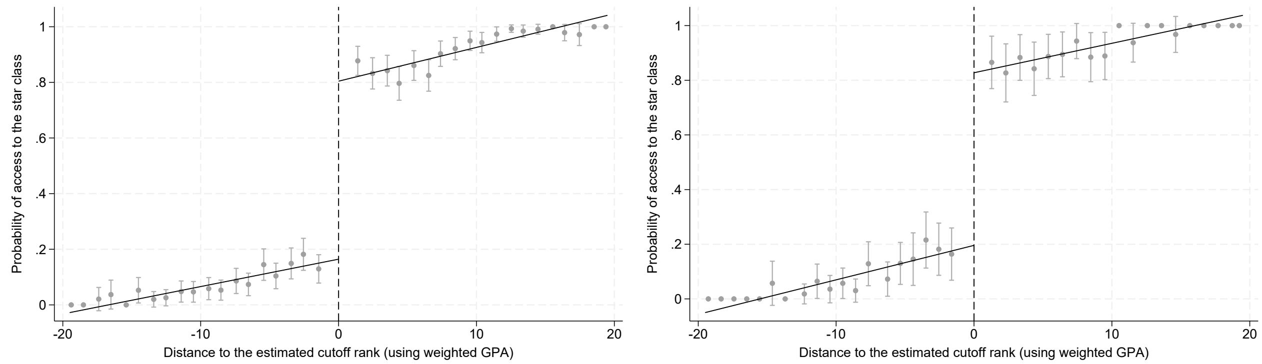
1.E.1 Validity of the Regression Discontinuity

Figure 1.E1: Density of the Running Variable Around the Threshold



Notes: The graph illustrates the density of the running variable (students' GPA rank in the 1st-year class) around the star class admission threshold, assessing potential manipulation at the cutoff via an RDDensity test. The running variable is defined as the distance between a student's GPA rank and the estimated threshold rank for star-class admission, computed at the *class* $\times$ *subfield* $\times$ *prep program* $\times$ *cohort* through an algorithm. We exclude individuals immediately below and above the threshold because, given the characteristics of the algorithm used to determine the cutoff, these individuals are outliers—specifically, always-takers or never-takers (De Chaisemartin and Behaghel (2020)).

Figure 1.E2: First Stage of the Regression Discontinuity, by Gender



(a) Male Students (b) Female Students

Notes: These figures depict the probability of admission probability to a star class, separately by gender. The running variable is defined as the distance between a student's GPA rank and the estimated threshold rank for star-class admission, computed at the *class* $\times$ *subfield* $\times$ *prep program* $\times$ *cohort* through an algorithm.

1.E.2 Complementary Results

Table 1.E1: Admission to Top 10% Most Selective STEM Graduate Schools, by Gender

	All Students (1)	Male Students (2)	Female Students (3)
Regression Discontinuity	Top 10 % of grad. schools (Desirability)		
Baseline mean (Below cutoff)	0.14	0.14	0.13
RD estimate (**ITT**)	0.074	0.103	-0.013
	(0.057)	(0.067)	(0.111)
RD estimate (**ATT**)	0.103	0.139	-0.020
	(0.079)	(0.090)	(0.166)
Regression Discontinuity	Top 10 % of grad. schools (Desirability for female students)		
Baseline mean (Below cutoff)	0.16	0.16	0.14
RD estimate (**ITT**)	0.110*	0.136**	0.027
	(0.059)	(0.069)	(0.115)
RD estimate (**ATT**)	0.152*	0.185**	0.040
	(0.081)	(0.093)	(0.170)
Obs. used in estimation	2,275	1,668	607
Total number of obs.	6,585	4,788	1,797

Notes: This table displays non-parametric regression discontinuity estimates of admission probability to the top 10% STEM graduate schools. These estimates are based on Calonico et al. (2017, 2019). STEM graduate school desirability is measured using the revealed preferences of applicants (Avery et al. (2013)). The running variable is defined as the distance between a student's GPA rank and the cutoff rank for star class admission, computed at the *class* $\times$ *subfield* $\times$ *prep program* $\times$ *cohort*. For results comparability, the bandwidth for the estimation is fixed at -6 and 6, the one selected by the optimal bandwidth algorithm for the ATT on the full sample in Table 5.

Table 1.E2: Robustness Checks for Regression Discontinuity Estimates of Admission to the Top 10% Most Selective STEM Graduate Schools, by Gender

	All Students (1)	Male Students (2)	Female Students (3)
RD estimate (ITT)	Top 10 % of grad. schools (Selectivity)		
1) Bandwidth = 4	0.152*	0.165	0.105
	(0.084)	(0.100)	(0.149)
Obs. used in estimation	1,362	1,006	356
2) Bandwidth = 8	0.084**	0.107**	0.008
	(0.043)	(0.051)	(0.078)
Obs. used in estimation	3,180	2,324	856
3) Bandwidth = 10	0.077**	0.091**	0.027
	(0.036)	(0.043)	(0.066)
Obs. used in estimation	4,046	2,948	1,098
4) Polynomial order 2	0.208	0.218	0.166
	(0.135)	(0.162)	(0.242)
Obs. used in estimation	2,275	1,668	607
5) Drop more students around cutoff	0.078	0.111	-0.034
	(0.086)	(0.101)	(0.163)
Obs. used in estimation	1,837	1,343	494
6) With sample weight	0.062	0.087*	-0.017
	(0.044)	(0.053)	(0.076)
Obs. used in estimation	2,275	1,668	607
7) Cluster at the running variable	0.114***	0.138***	0.036
	(0.018)	(0.016)	(0.045)
Obs. used in estimation	2,275	1,668	607
8) Local randomization	0.180***	0.188***	0.155***
95% CI	[0.150; 0.210]	[0.150; 0.230]	[0.090; 0.220]
Obs. used in estimation	2,275	1,668	607
Total number of obs.	6,585	4,788	1,797

Notes: This table reports nonparametric regression discontinuity estimates of the probability of admission to top 10% STEM graduate schools under a series of robustness checks. The estimates are computed following Calonico et al. (2017, 2019). School selectivity is measured by the percentile rank in the high school graduation examination of admitted students. The running variable is defined as the distance between a student's rank and the estimated threshold rank for star-class admission, calculated at the *class* $\times$ *subfield* $\times$ *preparatory program* $\times$ *cohort* level using an algorithmCompared to the baseline specification of Table 5, Panel 1 reduces the bandwidth from $[-6, 6]$ to $[-4, 4]$. Panel 2 increases the bandwidth to $[-8, 8]$, while Panel 3 further expands it to $[-10, 10]$. Panel 4 employs a second-order polynomial instead of the baseline first-order specification. Panel 5 extends the exclusion window around the estimated cutoff from $[-1, +1]$ students to $[-2, +2]$ students. Panel 6 uses a probability-weighted sample; estimates in this panel are not computed using the `rdrobust` package, which does not support probability weights. It displays the estimate of the coefficient of a variable *treat*, which takes value 1 if a student is ranked above the estimated cutoff, and 0 otherwise, in a linear regression. Panel 7 reports estimates with standard errors clustered at the running-variable level. Panel 8 presents local randomization estimates using the `rdrandinf` package, which reports confidence intervals rather than standard errors.

Online Appendix References

Avery, Christopher N., Mark E. Glickman, Caroline M. Hoxby, and Andrew Metrick, "A Revealed Preference Ranking of U.S. Colleges and Universities," *The Quarterly Journal of Economics*, 2013, *128* (1), 425–467.

Bonneau, Cécile, Pauline Charousset, Julien Grenet, and Georgia Thebault, *Quelle démocratisation des grandes écoles depuis le milieu des années 2000 ?*, Vol. 30, Institut des Politiques Publiques, 2021.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik, "Rdrobust: Software for Regression-Discontinuity Designs," *The Stata Journal*, 2017, *17* (2), 372–404.

_ , _ , _ , and _ , "Regression Discontinuity Designs Using Covariates," *Review of Economics and Statistics*, 2019, *101* (3), 442–451.

De Chaisemartin, Clément and Luc Behaghel, "Estimating the Effect of Treatments Allocated by Randomized Waiting Lists," *Econometrica*, 2020, *88* (4), 1453–1477.