Peer Influence and College Major Choices in Male-Dominated Fields*

Rocío Valdebenito†

University of Illinois at Urbana-Champaign

[Link to latest version]

December 16, 2023

Abstract

This paper investigates the causal impact of high school peers' choices on individuals' college major choices, while also exploring whether the gender of both individuals and their peers plays a mediating role in these effects. Utilizing Chilean data spanning from 2006 to 2019, this study addresses key challenges in the peer effects literature and proposes two distinct approaches. The first approach explores idiosyncratic within-school variations in the proportion of classmates enrolled in male-dominated post-secondary programs. The second approach employs a regression discontinuity design, taking advantage of the unpredictable admission cutoffs within the centralized admission system for Chilean universities. The primary findings indicate that both classroom peers and older high school peers significantly shape students' choices of college majors in male-dominated fields. Furthermore, the results suggest the greater influence of female peers enrolled in male-dominated areas compared to male peers, particularly in influencing female applicants towards similar majors.

Keywords: Gender, Higher Education, Peer effects. **JEL:** J16, I23, I24, J24.

*I thank Mary Arends-Kuenning and Catalina Herrera-Almanza for their support and guidance. I am also grateful for all the comments and suggestions from Yilan Xu, Elizabeth Powers, Sarah Janzen, Andrew Hultgren, Marin Skidmore, Jared Hutchins, Drew Hanks, Iván Flores, Cristhian Molina, and Carolina Concha-Arriagada. I am also thankful to all participants at seminars and conferences in the North East Universities Development Consortium Conference, Southern Economics Association, International Policy and Development workshop at the University of Illinois Urbana-Champaign, Association for Education Finance and Policy, Population Association of America, Midwest Economics Association, Agricultural and Applied Economics Association, Department of Human Sciences at the Ohio State University. I also thank the Chilean Ministry of Education and *Departamento de Evaluación, Medición y Registro Educacional* from the University of Chile for providing the data. I thank the support of Joaquín Pérez, from the Ministry of Education. I am also grateful for the support of the Chilean National Agency for Research and Development (ANID), Scholarship Program.

†PhD Candidate, e-mail address: riv2@illinois.edu

1 Introduction

One of the most significant transformations in developed and emerging economies is the increased educational attainment of women over the past few decades. These changes have had an impact on women's labor force participation and have partially narrowed the gender wage gap. However, despite improvements in college attendance rates, notable gender differences persist in choosing college majors. Recent empirical evidence highlights that these gender differences in majors explain a substantial portion of the gender wage gap, as jobs in male-dominated fields tend to offer higher salaries compared to those in female-dominated fields (Altonji et al., 2012; Sloane et al., 2021). Moreover, several studies suggest that while workers on STEM (Science, Technology, Engineering, and Mathematics) play a crucial role in driving economic growth, gender diversity within these fields can enhance innovation and overall performance (Bayer & Rouse, 2016; Peri et al., 2015). Despite some progress in reducing the gender gap in male-dominated majors over time, a persistent and significant gap continues to hinder labor market equality (Goldin et al., 2006).

College enrollment and major choices represent one of the most important decisions after high school graduation, with important implications for labor market outcomes. Unfortunately, it is difficult to understand what determines majors and field choices. Several studies have highlighted the significance of peers, individuals from classrooms, high schools, neighborhoods, and families, as they constitute a central aspect of teenagers' social environment (Aguirre & Matta, 2021; Altmejd, 2022; Anelli & Peri, 2019; Barrios-Fernández, 2022; Brenøe & Zölitz, 2020; Dustan, 2018; Lavy & Schlosser, 2011; Li, 2018). These studies, among others, have established that peers play a vital role in shaping individuals' decisions through various mechanisms, including social influence, role modeling, and aspirational effects. Recent experimental evidence further emphasizes the importance of role models for young female students. Given the gender imbalances in college majors, it becomes challenging for female students to interact with peers pursuing male-dominated fields (Porter & Serra, 2020).

This paper explores the case of Chile, where the educational system is characterized by significant gender disparities in college major enrollment and graduation. Within this context, students select specific majors (programs) within the university right from the beginning, a distinctive char-

acteristic that sets them apart from systems where major selection occurs at a later stage in the academic journey. The primary objective of this paper is to investigate the connection between high school peers and the choice of majors that have traditionally been dominated by men. Specifically, I analyze the influence of high school peers' enrollment in male-dominated fields on potential applicants' decisions to pursue similar majors. Furthermore, I examine the potential mediating roles played by the gender of the applicant and the gender of their peers in these effects.

The extensive body of literature on peer effects in education has primarily concentrated on the analysis of the role of peers within various educational contexts and across different educational levels, namely elementary, secondary, and post-secondary education, with a focus on diverse educational outcomes. More than 30 years of research have tried to address the empirical challenges associated with identifying peer effects. However, there exists a broad consensus that peers play a crucial role in shaping students' outcomes and choices (Epple & Romano, 2011; Sacerdote, 2011).

Numerous studies reveal notable gender disparities in how peers' characteristics and achievements impact students' outcomes (Balestra et al., 2021; Cools et al., 2022; Han & Li, 2009). Recent empirical research, which suggests that competitive environments with high-achieving peers may discourage female students from pursuing scientific fields (Brenøe & Zölitz, 2020; Fischer, 2017), aligns with another line of research that suggests that women and men respond differently in competitive environments (Gneezy et al., 2003; Niederle & Vesterlund, 2007, 2010). Specifically, women tend to avoid competitive environments more than men, resulting in an exacerbation of the gender performance gap in high-competitive settings.

This paper uses student-level administrative data spanning from 2006 to 2019, obtained from enrollment records in secondary and tertiary education in Chile, along with information on university applications from a centralized admission system. The empirical strategy in this paper addresses the primary challenges associated with studying peer effects by employing two approaches. Firstly, it investigates peer effects at the classroom level, particularly by analyzing within-school variations in the proportion of classmates enrolled in male-dominated majors. Second, it leverages the unpredictable cutoffs of programs from the centralized admission system and employs a regression discontinuity design analysis. This second analysis aims to provide causal evidence regarding the impact of high school peers' enrollment in male-dominated majors on applicants' choices within the

same field.

The primary findings reveal that, on average, applicants are more inclined to apply for and enroll in Technology and Engineering (TE) fields when they are exposed to high school peers who have also chosen this path. The classroom-level analysis demonstrates that a higher proportion of classmates enrolling in TE increases the likelihood of application and enrollment in TE programs. However, significant gender differences emerge when differentiating these effects by the gender of the applicant and their peers. Specifically, a higher proportion of female classmates enrolled in TE has a more pronounced impact than a higher proportion of male classmates. Moreover, male applicants tend to perceive greater benefits from a classroom with a higher number of students enrolled in TE than their female counterparts. It is important to note two main threats in the classroom-level analysis. First, this approach is unable to isolate whether these effects are coming from peers' characteristics or peers' outcomes, and second, the simultaneity issue as the outcomes of the peers and applicants are observed at the same time.

The regression discontinuity analysis approach, which tackles the potential threats in the classroom-level analysis, indicates that, on average, male applicants are not influenced in their decisions to apply for and enroll in TE programs when exposed to a one-cohort-older peer who has enrolled in TE through the centralized admission system. However, when considering the gender of the peers, the primary findings reveal that female peers have a significant and positive impact on the likelihood of female applicants to apply for and enroll in TE programs. These effects are substantial, especially when taking into account the baseline levels of female application and enrollment in TE programs. For instance, the presence of a female peer enrolled in TE increases the probability that a female applicant enrolls in TE by 2.2 percentage points, relative to an enrollment mean of 3.3%.

This paper aims to contribute to three important gaps in this literature. The first one corresponds to analyzing the transition from secondary to post-secondary education. Previous papers have explored, for instance, the influence of high-ability peers on individual test scores, or the impact of exposure to predominantly female or male cohorts on performance and major choice at the college or high school level (Briole, 2021a; Calkins et al., 2023; Landaud et al., 2020; Mouganie & Wang, 2020). Utilizing the educational context and data from Chile, this paper examines the influ-

ence of peers at the secondary level on subsequent post-secondary education choices, encompassing program and field selections made by students during their application and enrollment processes.

A second gap involves the estimation of endogenous peer effects, which focuses on the impact of peers' outcomes rather than their background, commonly referred to as contextual effects. Within this body of literature, a common challenge relates to the independent identification of both these effects, which causes significant econometric difficulties, and few studies have addressed this issue empirically (Aguirre & Matta, 2021; Barrios-Fernández, 2022; De Giorgi et al., 2010). This paper aims to contribute to the estimation of endogenous peer effects by investigating a quasi-random shock that exclusively impacts peers' major choices, as opposed to their test scores, parents' education, or any other background variable that could be correlated with major selection and college enrollment.

Lastly, this paper aims to contribute to a third gap related to understanding peer effects in the context of major choices, particularly in fields traditionally considered male-oriented. Previous empirical research has identified peers' gender and ability as important variables, but there is limited knowledge about the impact of peers' choices in specific fields of study on individuals' decision-making processes. Furthermore, the studies that have analyzed the effects on major choices typically concentrate on peer effects within the same educational level, usually secondary or post-secondary levels.

This paper is organized as follows. Section 2 explains the main challenges in the peer effects literature, Section 3 briefly describes the Chilean institutional context, Section 5 presents the steps require in the sample construction, Section 4 presents the empirical strategy while Section 6 exhibits the empirical tests used to validate the main assumptions from the empirical strategy. Finally, Sections 7 and 8 present the results and final remarks.

2 Peer effects in education

A growing body of literature has focused on explaining the gender gap in higher education choices. From score gaps in math and science to gender stereotypes and role models, there are multiple factors that help explain these gaps. Recent literature suggests that, among other factors, high-

school subject choices and achievement at earlier school stages are relevant for explaining gender gaps (Card & Payne, 2021). However, differences in ability, such as math achievement, do not fully explain these gaps (Cimpian et al., 2020; Riegle-Crumb et al., 2012), and beliefs that men are naturally more skilled in quantitative domains are empirically unfounded (Favara, 2012).

Peers, on the other hand, represent a central aspect of teenagers' social environment as they can impact a variety of important outcomes, including test scores, educational attainment, and trajectories (Balestra et al., 2021; Black et al., 2013; Carrell et al., 2018).[1] A key goal in this literature is to learn how the composition of peer groups influences different educational outcomes. However, several identification challenges exist in the peer effects literature, as noted by the contributions of Manski (1993) and Moffitt (2001).

Peer effects can be driven by different social interactions: contextual (or exogenous) effects, corresponding to changes in an individual's behavior due to peers' characteristics, endogenous effects that correspond to changes in the individual's behavior due to the prevalence of that behavior among peers, and correlated effects that correspond to the similarity of outcomes between peers due to having similar individual characteristics or experiencing similar shocks or institutional environment. Under the presence of peer effects, the challenge is to distinguish between exogenous, endogenous or correlated effects. Finding any effect can be a signal of, for example, the "birds of a feather flock together" phenomenon rather than any actual peer effect. In addition, if outcomes of peers simultaneously affect each other, then it becomes even harder to separate contextual and endogenous effects, which is the so-called reflection problem (Manski, 1993).

The literature on peer effects has tried different empirical approaches to develop credible identification strategies. One of these approaches is to estimate peer effects by leveraging the random assignment of students at the classroom level (Anelli & Peri, 2019; Duflo et al., 2011; Goulas et al., 2022) or roommates and classrooms at the college level (Elsner et al., 2021; Sacerdote, 2001; Zimmerman, 2003). Another approach explores the natural within-school variation in peer composition and peer characteristics across time (Anelli & Peri, 2019; Brenøe & Zölitz, 2020; Briole, 2021b; Cools et al., 2022; Lavy & Schlosser, 2011). Several findings can be retrieved from these empirical approaches, for example, Anelli & Peri (2019) finds that male students exposed to high

[1] See Epple & Romano (2011) and Sacerdote (2011) for an extensive review.

school cohorts composed of more than 80% male peers are more likely to choose male-dominated majors in Italy. Brenøe & Zölitz (2020), using Danish data, finds evidence that having a large proportion of female peers in class decreases the likelihood of enrolling in STEM fields. Another concept that arises in this literature is the impact of the "peer quality," measured for example by either peers' performance on standardized tests or grades (Balestra et al., 2021; Card & Payne, 2021; Mouganie & Wang, 2020). The evidence on this matter is inconclusive, and two recent papers show somewhat different results. While Balestra et al. (2021) show that the presence of peers with high intellectual ability affects the likelihood of selecting STEM occupations for men only, Mouganie & Wang (2020) find that high-performing female peers in math increase the likelihood that women choose STEM tracks. Despite the useful conclusions from these analyses that inform us about the actual presence of peer effects, the identification strategy from these approaches is unable to separate contextual and endogenous effects.

A different empirical approach explores the variations generated by random shocks that affect peers' outcomes. A recent example of this approach can be found in Barrios-Fernández (2022), in which the author explores the quasi-random experiment generated by loan eligibility in the Chilean centralized admission system to analyze the influence of neighbors' enrollment on potential applicants' college choices. Other examples include Altmejd et al. (2021) and Aguirre & Matta (2021), where the authors explore the admission score cutoffs to identify the effects of older siblings' trajectories on younger siblings' college choices. All of these studies have the empirical advantage of independently identifying endogenous peer effects by isolating them from correlated effects. As Section 4 explains, the identification strategy in this paper is closely aligned with the papers above.

3 Chilean institutional context

The postsecondary education system has different types of higher education institutions based on the degree type they offer. Figure 1 shows the total freshmen enrollment based on the four types of institutions. Vocational Formation Centers are those that offer two-year vocational degrees, while professional institutes are those that offer four-year Professional degrees. The next type of institution are universities that offer five-year bachelor's degrees. We classify universities into two

types. The first group consists of universities that are part of the centralized admission system (CS), while the rest are private universities outside the centralized system (non-CS, or private system). There are 60 universities that offer bachelor's degrees, and 43 that participate in a centralized admission system. Universities that do not participate in this admission system are predominantly private and typically serve lower-scoring students.[2] The participating universities are all non-profit and can be public, private, or private-parochial. Although the institutions of the centralized admission system span a wide range of selectivity levels, it also includes the country's most prestigious and traditional universities.

Secondary education spans four years, and schools can be categorized as public, private subsidized (voucher schools), and private. It is widely acknowledged that students enrolled in private schools typically come from high-income backgrounds, voucher schools tend to educate middle-income students, and public schools primarily serve students from lower income levels.[3] As of 2019, there are 2,610 high schools in the country, with an average of 59 students per high school in their senior year. Within each cohort, students are typically divided into classrooms, with an average of 26 students per classroom.

During their senior year of high school, students sign up to take a series of standardized tests to apply to any of the academic programs offered in the centralized admission. The series of tests called PSU (*Prueba de Selección Universitaria*, in Spanish) consists of two mandatory tests, mathematics and language, and one of two optional tests, science and history. Besides PSU scores, the students' performance measures of high school GPA and GPA ranking,[4] are the only other components of the weighted average score considered in the system. Each program–institution has specific weights that apply to each component of the weighted score, and the information about such weights is public and available to students before sending their applications.

After being informed of their scores, students submit a list from most to least preferred of up to 10 program-institution combinations, referred to in this paper as choices or alternatives.[5] After

[2]The centralized admission system is called *Sistema Único de Admisíon*, or Unified System of Admission.

[3]Students at public, voucher, and private schools represent 36%, 53%, and 11% of the total enrollment, respectively.

[4]Starting in the admission year of 2012, the GPA ranking is an average measure of relative performance–in terms of GPA, with respect to the current and previous cohorts.

[5]An alternative can be, for example, Civil Engineering at the University of Chile. If the student is also interested in Economics at the same institution, she could include that as a second-best alternative in her

receiving the application list from the students alongside the specific weighted scores computed for each program-institution alternative,[6] the system's algorithm implements a *deferred acceptance* assignment mechanism, to determine which students are offered admission to each program.[7]

Chile is similar to many other countries with trends indicating that a greater number of female students have been enrolling in tertiary education compared to their male counterparts. Figure 2 depicts the trend where, starting from 2016, more women than men are participating in university enrollments through the centralized admission system. However, striking differences occur when analyzing patterns at the field level. Figure 3 shows that the lack of female graduates in the fields such as "Engineering, Manufacturing, and Construction", and "Information and Communication Technologies" is an issue in most countries, including Chile. The OECD average share of female graduates in these two fields are around 7% and 2%, respectively. In Chile, these shares reach 7.6% and 0.7%, respectively (OECD, 2017).

Enrollment data at universities from the centralized admission system in Chile show significant gender gaps by field. Figure 4 shows the average freshmen enrollment by men and women across different fields of study. On average, there are almost 10,000 more men than women enrolled in the field of "Technology and Engineering" (TE) per year, and women represent around 25% of the freshmen enrollment in such field. In contrast, the fields of "Humanities, Social Sciences, Arts, and Education" (HASSE), and "Health" show opposite patterns, with almost 5,000 more women than men per year enrolled in such fields, respectively.

Enrollment patterns indicate that fields dominated by males have experienced an even greater male dominance compared to the past, while the same trend is observed in female-dominated fields. Figure 5 illustrates the annual evolution of the total freshmen enrollment by gender in four of the fields of study presented in Figure 4, specifically the ones with the highest and lowest difference between male and female enrollment. The most male-dominated field–TE–shows that the gender

preference list.

[6]For example, Civil Engineering at the University of Chile in 2021 assigned 10% to the GPA score, 20% to GPA ranking, 10% to language PSU score, 45% to math PSU score, and 15% to science PSU score. This particular program requires that applicants have to take the science test instead of the history test. Other programs, in contrast, allow students to decide which test they want to take, and therefore the highest score between these two components is the one used to calculate the weighted score. In the end, regardless of the program, only one test (either science or history) is used in the final calculation of the weighted score.

[7]See Rios et al. (2021) for further details on the admission system.

gap has increased over time. In contrast, the field of "Business" exhibits a gender-balance pattern without significant changes across time. Regarding the female-dominated fields, both HASSE and "Health" have shown an increase in the gap between female and male enrollment over time.

4 Identification strategy

Many unobserved factors usually confound the effect of peers' choices on potential applicants' outcomes. For instance, one might suspect that correlated factors could arise between peers and applicants when a school excels in encouraging students to enroll in engineering programs, and there is self-selection and sorting of students across schools due to specific school characteristics. A common approach to account for these confounding factors is to rely on within-school variations in the proportion of students enrolling in male-dominated fields.[8] This method examines whether cohort-to-cohort changes in male and female choices within the same school are systematically associated with cohort-to-cohort changes in the proportion of students enrolled in a male-dominated field. An important assumption is that, while parents may make decisions based on school characteristics, they do not do so based on specific characteristics of their child's relevant cohort. Therefore, the variations in cohorts within schools can be treated as quasi-random.

The approach that leverages within-school variations allows for distinguishing between correlated and social interaction effects. However, within the social interaction effects, there are contextual and endogenous peer effects that are difficult to separate (Manski, 1993). I divide the empirical strategy into two approaches. The first one distinguishes between correlated and social interaction effects, while the second one expands the analysis by leveraging programs' cutoff and separating contextual and endogenous peer effects.

First empirical approach:

Using repeated cross-section data, I estimate the following reduced-form equation separately for male and female potential applicants:

$$\text{Applicant chooses TE}_{icst} = \alpha + \beta \text{Share Peers TE}_{-i,cst} + X'_{icst}\gamma + \delta_s + \lambda_t + \nu_s \text{year}_{st} + \varepsilon_{icst}, \quad (1)$$

[8] Similar empirical strategies can be found in Brenøe & Zölitz (2020); Briole (2021b); Cools et al. (2022); Lavy & Schlosser (2011).

where i denotes individuals, c denotes classroom, s denotes schools, and t denotes time. The outcome Applicant chooses TE_{icst} captures whether the applicants follow, either by enrollment or application, the field of technology and engineering. In practice, I primarily analyze three outcomes. First, "apply to TE – CAS" is a binary variable that takes value one when the potential applicant submits a program in a TE field from the centralized admission system. The variable takes the value of zero if either the application was made to another field or there was no application at all. Second, "enroll TE–CAS" is a binary outcome that takes the value one when the potential applicant enrolls in a TE program within the centralized admission system. The variable takes the value of zero if no enrollment is observed, or the applicant enrolled in a different field. Finally, "enroll TE non–CAS" is a binary outcome that takes the value one when the potential applicant enrolls in a TE program in any institution outside the centralized admission system. The variable takes the value of zero if no enrollment is observed, the applicant enrolled in a different field, or in an institution from the centralized admission system.

The term δ_s captures school fixed-effects and the term λ_t time fixed-effects. The former controls for unobserved average students and school characteristics that are constant across time, while the latter controls for time-variant unobserved shocks common across schools. I also include the term $\nu_s \text{year}_{st}$ that captures school-specific linear time trends to control for time-varying unobserved factors at the school level. The vector X_{icst} controls for individual characteristics: PSU math scores, GPA, high school attendance (%), family income, and parents' education.

The variable Share Peers $\text{TE-CAS}_{-i,cst}$ represents the share of students in the same classroom as i that enroll—the year after high-school graduation—in a TE program from the centralized admission system. Importantly, I eliminate student i from the classroom-school calculation. The parameter β captures the effect of the proportion of students enrolling in TE in the same classroom as i on i's college choices concerning TE fields.

I expand equation 1 by calculating the share of peers at the classroom level by gender:

$$\begin{aligned}\text{Applicant chooses TE}_{icst} = \alpha + \beta_1 \text{Share Male TE-CAS}_{-i,cst} + \beta_2 \text{Share Female TE-CAS}_{-i,cst} \\ + X'_{icst}\gamma + \delta_s + \lambda_t + \nu_s \text{year}_{st} + \varepsilon_{icst},\end{aligned} \tag{2}$$

where Share Male $\text{TE-CAS}_{-i,cst}$ (Share Female $\text{TE-CAS}_{-i,cst}$) is the share of male (female) peers in

the same classroom as i that enroll in TE the year after graduation, in a program at the centralized admission system. Both shares are the sample means of TE enrollment at the classroom level of the leave-one-out distribution of the specific gender.

Both reduced-form specifications—equations 1 and 2—exploit the idiosyncratic deviation from the school's long-term trend in the proportion of classroom peers enrolled in TE. In this approach, we compare the outcomes of potential applicants in cohorts exposed to the same school environments and characteristics, except for the fact that in some cohorts, the classroom has a higher proportion of peers enrolled in TE due to random factors. Using this approach, one can identify the presence of social interactions that are not confounded by correlated factors. However, two issues are important to note. First, these social interactions can be due to changes in peers' outcomes (endogenous peer effects) or changes in peers' characteristics (contextual peer effects), and this approach does not separate them. Second, as potential applicants and their peers are in the same classroom and, therefore, the same cohort, there is the potential for a simultaneity issue as they can influence each other's choices. The second empirical approach aims to address these two issues.

Second empirical approach:

The second empirical strategy in this paper employs a regression discontinuity (RD) approach that leverages the centralized admission system's unpredictable cutoffs, for which a subset of applicants, the cutoff effectively randomizes admission offers to a male-dominated field (TE). In the admission to universities in the centralized admission system, each admission cutoff can be used as a separate natural experiment that provides the exogenous variation needed to isolate endogenous from contextual peer effects. In this approach, marginally admitted students are comparable in observable characteristics to those marginally rejected, except for admission to a TE program.

Furthermore, I define peers as students who are one cohort older than potential applicants but are attending the same high school. This approach helps address the simultaneity issue because older peers can influence younger potential applicants, but not vice versa. For each potential applicant i, I identify an older peer p in the same school (s) as i that applies to universities in the admission year $t-1$, to an alternative j. In this case, the alternative j is a program classified as a TE program.

I estimate a Fuzzy RD design that captures the effect of having a marginal peer enrolled in a

TE program. I start by defining the running variable for all peers p as follows:

$$r_{pjt} = s_{pjt} - c_{jt}, \tag{3}$$

which measures the distance between the peer's weighted score applying to field j and its cutoff. If $r_{pjt} \geq 0$, then the peer is admitted in his preferred field j–TE, and if $r_{pjt} < 0$, then he is rejected.

In this setting, an indicator for being above the cutoff of a TE program is used as an instrument for the actual enrollment of the peer in the same field. Therefore, the first stage is represented as:

$$\text{Peer enrolls in TE-CAS}_{pj,t-1} = \pi_1 Z_{pj,t-1} + h(r_{pj,t-1}) + \beta_s + \psi_t + \gamma_j + \nu_{pj,t-1}, \tag{4}$$

where $Z_{pj,t-1}$ is an indicator variable that captures whether the peer p crossed the admission cutoff of the TE program, then $Z_{pj,t-1} = \mathbb{1}[r_{pj,t-1} \geq 0]$. The outcome variable from this stage, Peer enrolls in TE-CAS$_{pj,t-1}$, is a binary indicator that equals one when the peer is enrolled in TE in the centralized admission system and zero otherwise. The function $h(\cdot)$ represents a polynomial of the running variable $r_{pj,t-1}$, that it can be a first-order or higher-order function. The terms β_s, ψ_t, and γ_j, are school and year fixed effects, respectively. And ν_{ipjt} is an error term.

The second-stage of this procedure includes, as a regressor, the predicted outcome of equation (4). Equation (5) represents the second-stage as follows,

$$\text{Applicant chooses TE}_{ipjt} = \tau \text{Peer enrolls in TE-CAS}_{pj,t-1} + h(r_{pj,t-1}) + \mu_s + \alpha_t + \delta_j + \varepsilon_{ipjt}, \tag{5}$$

the outcome variable **Applicant chooses TE**$_{ipjt}$ captures whether the applicants follows, either by enrollment or application, the field of the peer that the applicant is exposed to at high school. I use the same outcomes defined earlier: "apply to TE–CAS","enroll TE–CAS", and "enroll TE non–CAS."

The parameter of interest τ, recovers the effect of peer's enrollment into a TE program on potential applicants' choices in the same field. Similarly to the first stage, μ_s, α_t, and δ_j are terms capturing school, admission year, and preferred alternative fixed effects, respectively. The term $h(\cdot)$ represents the polynomial function of the running variable and ε_{ipjt} an error term.

I estimate equation (5) using the RD robust approach proposed by Calonico et al. (2014a,b), which is a non-parametric approach for RD design that does not impose strong assumptions on the shape of the relationship between the running variable and the outcome. In this paper, I present local linear polynomial estimation to both sides of the threshold using uniform kernels and optimal bandwidths selected by the *rdrobust* package (Calonico et al., 2020), that are chosen to minimize the mean squared error.

5 Data and descriptive statistics

This paper focuses on the transition from high school to universities over the period 2006–2019, using individual-level data where the information has been previously anonymized with a unique student *id* that allows for the identification of educational trajectories. The main sources of information are the Department of Evaluation, Measurement, and Educational Registry (DEMRE, in Spanish) and the Ministry of Education (*Centro de Estudios Mineduc*). DEMRE is the agency responsible for standardized tests and the entire process of the centralized admission system to the universities from the Council of Rectors of Chilean Universities. The datasets provided by DEMRE include the scores of each of the subjects included in the college admission test, the ranking of preferences submitted by the applicants, self-reported socioeconomic information, admissions offered from the centralized system, and enrollment at the universities involved in the centralized system. From the Ministry of Education, the primary datasets include student enrollment records for secondary education, which encompass schools and classrooms' *id*, as well as academic performance with consolidated GPAs at the end of each academic year. The Ministry of Education also provides student enrollment records for higher education, including enrollment at institutions both inside and outside the centralized admission system.[9]

The primary population of interest comprises senior high school students between 2006 and 2019 who must decide whether or not to pursue higher education and have taken the standardized test, PSU. This population is referred to as potential applicants. This population also corresponds to the analytic sample used in the first empirical strategy. Descriptive statistics for this sample

[9]See Online Appendix A.1 for further details.

are presented in Table 1, with the statistics further categorized by gender. Approximately 47% of the applicants opt to apply to (CS) programs, while roughly 28% ultimately enroll. Male and female students exhibit similar application rates for CS programs, but a slightly higher proportion of male students go on to enroll in these programs.As expected, there are significant gender-based disparities in the fields of study students pursue. When considering both CS and non-CS options, approximately 26% of male applicants enroll in a TE program, while this proportion is less than 7% for female applicants.

In the second empirical strategy—the regression discontinuity design–the main objective is to identify the exogenous variation around the cutoff that affects peers' admission to a male-dominated field. Peers and potential applicants are students who attended the same high school, but peers were exposed to the choice of enrollment one year before the potential applicants.

The construction of the peers data follows three important steps. The identification of the undersubscribed programs, the elimination of dominated alternatives, and the construction of target and counterfactual alternatives.[10]. In this case, the counterfactual alternative is that the peer is not admitted in any other choice.

First, undersubscribed programs are programs that could not fill completely its seats, and therefore cutoffs cannot be identified. I define the admission cutoff c_{jt} to a program-institution j at a given year t, as the minimum weighted score among students who were offered admission, in programs with at least one not-admitted applicant:

$$c_{jt} = \min\{s_{ijt}\} \quad \text{s.t. } i \text{ is offered admission to } j \text{ in the admission year t}, \tag{6}$$

where s_{ijt} is the average weighted score of applicant i obtained when applying to the alternative j at year t, calculated as:

$$s_{ijt} = \sum_{l} s_i^l \alpha_{jt}^l, \tag{7}$$

where α_{jt}^l is the weight that the program–institution (j) assigns to component l in the academic year t, and s_i^l is the score student i obtained in the specific component l–math, language, history/science

[10]See Online Appendix A.2 for further details and examples on the construction of the peers data.

or GPA scores.

The second aspect of data construction is the elimination of dominated alternatives (Abdulkadiroğlu et al., 2014; Aguirre & Matta, 2021; Aguirre et al., 2022). A dominated alternative happens when an applicant submits a highly selective alternative (i.e., relatively higher cutoffs) in a lower-ranked position. For example, assume the case of a student who ranks in the first place a program j with very low selectivity, followed by a program k with very high selectivity. If the student is above the cutoff of k, and by consequence, he is also above the cutoff of j, he would be admitted to program j because it is a preferred choice. So, being above the cutoff of k would have no effect on the assignment to k. Thus, in this stage, I identify and eliminate dominated alternatives from the sample, because keeping dominated alternatives in the data reduces the statistical power of the first stage. In other words, for a given applicant, the resulting sample after this cleaning procedure contains ordered preferences in which any lower-ranked choices are alternatives where the applicant could in fact be admitted, if she is below the cutoff in a higher-ranked choice.

An important aspect of the Chilean centralized admission system is that weights of each component of the final weighted score and cutoffs are program-specific. This feature adds another layer of complexity to the admission system in which simply comparing programs' cutoffs is not enough to define high/low selectivity programs.[11] Aguirre & Matta (2021) and Aguirre et al. (2022) explain the concept of *relative selectivity*, which helps to identify when a lower-ranked program is relatively more selective than a higher-ranked program, from the applicant's perspective.[12] If that is true, as explained before, the relatively more selective program would not survive the elimination procedure.[13]

The third step consists of merging fields of study classifications to the program-institution alternatives. Since each preference consists of a specific program–institution, it is possible that students apply to the same field across their preference list. In this step, I collapse consecutive

[11] Note that in the specific case when two programs assign identical components' weights, then the simple comparison between their cutoffs is enough to define which program is more selective.

[12] Following Aguirre & Matta (2021) and Aguirre et al. (2022), *Relative Selectivity* is calculated as $\phi_{ij} = \frac{s_{ij} - c_j}{\sqrt{\sum_l (\alpha_j^l)^2}}$, which represents the euclidean distance from applicant's i scores (components) to the admission frontier defined by the cutoff at program j.

[13] See Table A.1 with the results of the elimination procedure and the resulting number of observations after each iteration. According to this table, around 59% of the observations survived the elimination procedure.

alternatives classified in the same field. For example, if an applicant submits five consecutive preferences in the same field,[14] then I keep the alternative in which the applicant is closer to the cutoff.

The fourth step in the data construction consists of creating observation pairs of a preferred field (j) and a counterfactual or also called fall-back (k) field. The main objective in this step is constructing a pair in which the counterfactual alternative serves as a plausible scenario in the case the student is not admitted to his target choice. Given that the main objective is to "randomize" admission to Technology and Engineering[15], the counterfactual alternative is defined as "nothing", meaning that no other alternative in the centralized admission system was awarding admission.

Finally, I connect peers and potential applicants. The merging process is performed using the same school *id* between potential applicants and peers, but combining lagged cohorts of peers being one year ahead of the applicants. In other words, I perform a merging process by conditioning students at the same high school, where peers' admission process is observed at $t-1$, and potential applicants' admission process is observed at time t.[16] Thus, the merging process includes only one observation per potential applicant, and in a given year, all students from the same high school, are connected with only one peer at the admission margin of TE.

It is unsurprising to observe that high schools with peers who were marginally admitted to a TE program demonstrate slightly different characteristics compared to schools without peers at that margin. Table 2 presents the average characteristics of these two samples and underscores the distinctions between them. For instance, on average, schools with peers admitted to TE programs tend to have students with higher PSU scores in both math and verbal tests, and their students exhibit higher enrollment rates in programs offered through the centralized system.

Table 3 shows descriptive statistics of the analytical sample for the applicants and their peers. There are in total 397,817 potential applicants with an older peer at the admission margin TE which represents in total 6,617 unique peers. Given that peers are students who self-select themselves to

[14] Applying to engineering programs at different universities or different engineering programs at the same university, or a combination of both.

[15] Examples of programs classified as TE include Civil Engineering, Engineering in Informatics and Computing, Civil Engineering in Construction, among others.

[16] In order to avoid duplicate potential applicants, if there are in the same high school multiple peers on the margin of applying to j and k that survived the elimination procedure, then I only keep the peer that is closer to the cutoff of j.

participate in the centralized admission system after observing their scores,[17] it is expected that peers present higher PSU scores and GPA than potential applicants. Moreover, in this sample, peers are students who have at least one alternative submitted to a TE program; it is thus also expected that the proportion of female students is lower among the peers than among the potential applicants. As Table 3 indicates, the share of female students is 54% versus 26% at the sample of peers and potential applicants, respectively. Finally, peers also present better socio-economic status, as their parents' education and family income are higher than the potential applicants' parents' status.

6 Validation of RD assumptions

RD designs require that students whose weighted score is near the threshold are comparable in terms of observable and unobservable characteristics regardless of their actual admission status, and that the treatment assignment is not manipulable (Lee & Lemieux, 2010). This section aims to explore empirically these assumptions. A first falsification test explores the manipulation of the running variable, which in practice translates into testing whether the number of observations below the cutoff is substantially different from the number of observations above the cutoff. Figure 6 presents the density test suggested by Cattaneo et al. (2020), where the robust bias-corrected test of the null hypothesis of "no manipulation" shows a *p-value* of 0.7206, providing evidence in favor of the continuity of the running variable. Aside from the graphical and statistical evidence, as explained in section 3, the national entry exam is part of a complete centralized admission system, in which students are unable to manipulate programs' cutoff and thus, manipulation of the running variable is an unlikely scenario.

A second test examines whether, around the cutoff, treated and control individuals are similar in terms of baseline observable characteristics. If baseline covariates, that are expected to be correlated with the outcome, are not continuous at the cutoff, the continuity assumption of the potential outcome functions is likely to fail. Figure 7 shows the 95% confidence intervals of the treatment

[17] Students with extremely low-score are less likely to submit their applications because weights and the previous year's cutoffs are public information; thus, they can expect that admission would not be awarded.

effect[18] on a set of socioeconomic and individual characteristics, such as parents' education, family income, gender, and GPA. Figures 7a and 7b show the continuity test for potential applicants and peers, respectively. As it is shown, all estimates are not statistically significant, providing evidence that baseline covariates are continuous at the cutoff.

Finally, an additional assumption applicable to Fuzzy RD designs, is that the threshold-crossing indicator is a good instrument of the treatment assignment. A visual representation of that relationship is shown in Figure 8, which exhibits the first stage presented in Equation 4. The discontinuity of the peers' enrollment in TE at the cutoff supports the strong association between being above the admission cutoff and peers' enrollment.

7 Results

This section presents the main results of the paper. I begin by presenting the results for the empirical strategy that leverages within-school variation in the proportion of classmates enrolling in TE. Table 4 reports the estimation results and standard errors clustered at the school-classroom level for the model described in equation 1. The variable % Classmates TE is divided by 10, so the coefficient interpretation corresponds to a 10% change, which could be seen as an increase of 3 peers enrolled in TE when the classroom has 30 students, close to the average number of students per classroom (26 students). Column (1) shows that, for male applicants, a 10% increase in the proportion of classmates enrolled in TE is associated with a 4.1 percentage point increase in the probability of applying to a TE program. In the case of female applicants, as shown in column (2), a 10% increase in the proportion of peers enrolled in TE increases the likelihood of application to TE by 1.9 percentage points. It is important to note, however, the significant gender differences in TE application and enrollment. A 4.1 percentage point increase is associated with a 27% change for males, while a 1.9 percentage point increase is associated with a 43% change for females.

Table 5 presents the main results for the model described in equation 2. Throughout columns (2) to (4), a 10% increase in the proportion of female peers has a larger impact than a 10% increase in the proportion of male peers. However, male applicants tend to perceive a greater impact when

[18] The coefficient associated with the indicator variable that captures whether the running variable is above zero.

exposed to a larger proportion of male or female peers enrolled in TE. Columns (5) and (6) show null effects of male peers on enrollment to programs outside the centralized admission system, and a small but statistically significant effect from female peers on the same outcome. These results, as described in the empirical strategies discussion, are not able to distinguish whether these social effects are coming from the characteristics of peers or the outcomes of peers.

Table 6 presents the results from the Fuzzy RD model described in equation 5. First-stage estimates show that being above the cutoff of admission to a given TE program effectively creates a discontinuity in the probability of enrollment in that program. All of these estimates show the impact of having one high school peer, one year older than the applicant, who has enrolled in TE in a program from the centralized admission system. Column (1), for example, shows a positive effect on the probability that the applicant is applying to a TE–CAS program, but it is not statistically significant at conventional levels. However, column (2) exhibits a positive and statistically significant effect, with an increase of 1.9 percentage points in the probability of enrollment in a CAS program. Interestingly, there is a negative impact on the probability of enrollment in a TE program outside CAS, which are programs usually considered less selective.

Table 7 exhibits the results when analyzing the differentiated role of the applicants and peer's gender in the transmission of peer effects. For each outcome, Table 7 shows two columns, the first column representing female potential applicants and the second column representing male potential applicants. Panel A and B exhibit the results of female and male peers, respectively. Panel A of Column (1) shows that when female applicants are exposed to a female peer enrolled in TE, there is an increase of 6.6 percentage points in the probability of application to a TE program from the centralized admission system. Panel A Column (3) shows an increase of 2.2 percentage points in the female probability of TE–CAS enrollment when exposed to a female high school peer enrolled in the same type of program. Columns (5) and (6) show that, female peers are significantly impacting both male and female potential applicants in their enrollment probability in TE programs outside the centralized admission system.

Panel B of Table 7 presents the impact of male peers. As shown in columns (2) and (4), there is no effect on the application and enrollment in TE–CAS programs for male applicants. However, male peers enrolled in TE–CAS discourage male applicants from enrolling in TE programs outside

CAS.

Heterogeneous effects

Heterogeneous effects by school/classroom size are presented in Figure 9 for the within-school variation approach and Figure 10 for the fuzzy RD approach. It is important to note that, while the first specification measures the impact of the proportion of peers enrolled in TE, the second specification analyzes the impact of one marginal peer enrolled in TE. Figure 9 shows a larger effect when classrooms are larger, represented in the highest quartiles of the classroom size distribution. As expected, a given proportion of classmates enrolled in TE translates into more students when the classrooms are bigger. In contrast, Figure 10 shows that the impact of one marginal peer enrolled in TE decreases and is not statistically significant when schools are in the top terciles of the school size distribution.

Finally, heterogeneous effects based on math scores are presented in Figure 11 for the within-school variation approach and Figure 12 for the fuzzy RD approach. In Figure 11, the top panel illustrates the impact of the proportion of male classmates enrolled in TE, while the bottom panel illustrates the impact of the proportion of female classmates enrolled in TE. For both application and enrollment in TE–CAS programs, students in the top quartiles of the math distribution are significantly more affected than students in the bottom quartile of the distribution. In contrast, concerning enrollment in TE non–CAS programs, students in the two lowest quartiles increase their probability of enrollment when exposed to a higher proportion of classmates enrolled in TE–CAS.

Figure 12 demonstrates that the point estimates of the effect of having a high school peer enrolled in TE increase with higher math scores for female potential applicants, and it is statistically significant in the application to TE–CAS in the top tercile. Although a similar trend is found for enrollment in TE–CAS for both male and female applicants, the estimates are not significant at conventional levels.

Potential Mechanisms

The administrative records used in this analysis do not allow the identification of whether female students, for example, are more like to enroll and apply to TE programs because female peers are impacting their aspirations. In this section, I present two approaches that shed light on potential explanations for the results presented previously.

First, in high schools where enrollment at programs from the centralized admission system is rare, marginal peers could become more relevant as they could bring information about these institutions and majors. Table 8 explores the results for two types of high schools. Panel A exhibits the estimates for schools in which the number of students enrolled in programs at the centralized admission system is less than 15, Panel B shows the results for schools in which there are more than 15 students enrolled in the system. The results suggest that the impact of a high school peer enrolled in TE increases the enrollment and application to TE–CAS programs of applicants when schools are less connected with the system as measured by the lack of students usually enrolled in programs from these institutions. Panel B shows that the estimates are smaller in magnitude or not statistically significant, suggesting that the relative importance of a high school peer marginally enrolled in TE is not significant when schools are already having an important number of students enrolled in CAS–programs.

Secondly, one could expect that a one-year-older high school peer enrolled in TE–CAS could influence younger students by encouraging them to study harder in their senior year. Thus, the increase in enrollment probability could be driven by an increase in their test scores and GPA. However, Figure 13 shows that high school peers enrolled in TE are not impacting the 12th-grade GPA or test score in math or verbal for female students. They do, however, impact the GPA score of male students and their verbal score.

Although more research is needed to identify potential mechanisms, these preliminary results suggest that high school peers enrolled in TE seem to become relevant within school environments where these college trajectories are scarce, and that–at least within female students–high school peers enrolled in TE are not impacting the effort or investment in better grades or scores during their senior year.

8 Conclusions

This paper resides at the intersection of two significant areas within the economics of education literature: gender gaps in college major selection and peer effects. While extensive research has been conducted in these domains across various educational contexts, there are still notable limitations

and opportunities for further advancements. Recent empirical studies have aimed to explore the causal impact of peers' outcomes and characteristics on college major choice. However, isolating the endogenous and exogenous effects that emerge due to the simultaneity problem, as described in the seminal works by Manski (1993) and Manski (1993), remains challenging.

The growing body of literature on college choice and gender disparities in fields like STEM has provided valuable insights into understanding the reasons behind differing choices made by women and men. However, there still exists a substantial knowledge gap to be addressed. This paper aims to contribute to the current understanding in this area by investigating the relationship between high school peers and the selection of college majors that have traditionally been male-dominated. Furthermore, it examines whether the gender of the applicant and their peers mediate these influences.

Given the numerous empirical challenges present in the peer effects literature, this study initially explores peer effects at the classroom level, focusing on the analysis of within-school variations in the proportion of classmates enrolled in male-dominated majors. Additionally, it builds upon the contributions of studies like Altmejd et al. (2021), Barrios-Fernández (2022), and Aguirre et al. (2022), which take advantage of Chile's centralized admission system. This system provides a useful setting for isolating endogenous peer effects from contextual peer effects. In this second approach, the study compares students whose peers were marginally admitted to male-dominated programs with students whose peers were marginally rejected from such programs.

Through both of the empirical approaches, I found that classroom and older high school peers are important predictors of students college major choices related to a male-dominated field. However, both of the empirical results are consistent by showing that female peers enrolled in male-dominated areas are relatively more important than male peers on potential applicants' decisions on majoring in similar programs. Particularly, the regression discontinuity approach, that addressed the main limitations discussed in the classroom-level approach, reveals that male applicants aren't influenced by one-cohort-older peers enrolled TE. But, considering peer gender, female peers significantly and positively impact female applicants. For instance, the presence of a female peer enrolled in TE increases the probability that a female applicant enrolls in TE by 2.2 percentage points.

References

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1), 137–196.

Aguirre, J. & Matta, J. (2021). Walking in your footsteps: Sibling spillovers in higher education choices. *Economics of Education Review*, 80, 102062.

Aguirre, J., Matta, J., & Montoya, A. M. (2022). Joining the Old Boys' Club: Women's Returns to Majoring in Technology and Engineering. unpublished.

Altmejd, A. (2022). Inheritance of fields of study. unpublished.

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., & Smith, J. (2021). O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries. *The Quarterly Journal of Economics*, 136(3), 1831–1886.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4(1), 185–223.

Anelli, M. & Peri, G. (2019). The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, 129(618), 553–602.

Balestra, S., Sallin, A., & Wolter, S. C. (2021). High-Ability Influencers? The Heterogeneous Effects of Gifted Classmates. *Journal of Human Resources*, (pp. 0920–11170R1).

Barrios-Fernández, A. (2022). Neighbors' Effects on University Enrollment. *American Economic Journal: Applied Economics*, 14(3), 30–60.

Bayer, A. & Rouse, C. E. (2016). Diversity in the Economics Profession: A New Attack on an Old Problem. *Journal of Economic Perspectives*, 30(4), 221–242.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics*, 31(1), 119–153. Publisher: The University of Chicago Press.

Brenøe, A. A. & Zölitz, U. (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics*, 38(4), 1009–1054.

Briole, S. (2021a). Are girls always good for boys? Short and long term effects of school peers' gender. *Economics of Education Review*, 84, 102150.

Briole, S. (2021b). Are Girls Always Good For Boys? Short and Long Term Effects of School Peers' Gender. *Economics of Education Review*, 84, 102150.

Calkins, A., Binder, A. J., Shaat, D., & Timpe, B. (2023). When Sarah Meets Lawrence: The Effects of Coeducation on Women's College Major Choices. *American Economic Journal: Applied Economics*, 15(3), 1–34.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2), 192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for Regression-discontinuity Designs. *The Stata Journal: Promoting communications on statistics and Stata*, 17(2), 372–404.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014a). Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal*, 14(4), 909–946.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6), 2295–2326.

Card, D. & Payne, A. A. (2021). High School Choices and the Gender Gap In STEM. *Economic Inquiry*, 59(1), 9–28.

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The Long-Run Effects of Disruptive Peers. *American Economic Review*, 108(11), 3377–3415.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Elements in Quantitative and Computational Methods for the Social Sciences*.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531), 1449–1455.

Cimpian, J. R., Kim, T. H., & McDermott, Z. T. (2020). Understanding persistent gender gaps in STEM. *Science*, 368(6497), 1317–1319.

Cools, A., Fernández, R., & Patacchini, E. (2022). The asymmetric gender effects of high flyers. *Labour Economics*, 79, 102287.

De Giorgi, G., Pellizzari, M., & Redaelli, S. (2010). Identification of Social Interactions Through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics*, 2(2), 241–275.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), 1739–1774.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics*, 134, 372–391.

Elsner, B., Isphording, I. E., & Zölitz, U. (2021). Achievement Rank Affects Performance and Major Choices in College. *The Economic Journal*, 131(640), 3182–3206.

Epple, D. & Romano, R. E. (2011). Peer Effects in Education. In *Handbook of Social Economics*, volume 1 (pp. 1053–1163). Elsevier.

Favara, M. (2012). The Cost of Acting 'Girly': Gender Stereotypes and Educational Choices. unpublished.

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, 46, 211–226.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *The Quarterly Journal of Economics*, 118(3), 1049–1074.

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives*, 20(4), 133–156.

Goulas, S., Griselda, S., & Megalokonomou, R. (2022). Comparative Advantage and Gender Gap in STEM. *Journal of Human Resources*.

Han, L. & Li, T. (2009). The gender difference of peer influence in higher education. *Economics of Education Review*, 28(1), 129–134.

Landaud, F., Ly, S. T., & Maurin, . (2020). Competitive Schools and the Gender Gap in the Choice of Field of Study. *Journal of Human Resources*, 55(1), 278–308.

Lavy, V. & Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2), 1–33.

Lee, D. S. & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281–355.

Li, H.-H. (2018). Do mentoring, information, and nudge reduce the gender gap in economics majors? *Economics of Education Review*, 64, 165–183.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3), 531–542.

Moffitt, R. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions. In *Social Dynamics* (pp. 45–82). Booking Institutions Press.

Mouganie, P. & Wang, Y. (2020). High-Performing Peers and Female STEM Choices in School. *Journal of Labor Economics*, 38(3), 805–841.

Niederle, M. & Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3), 1067–1101.

Niederle, M. & Vesterlund, L. (2010). Explaining the Gender Gap in Math Test Scores: The Role of Competition. *Journal of Economic Perspectives*, 24(2), 129–144.

OECD (2017). OECD Gender Data.

Peri, G., Shih, K., & Sparber, C. (2015). STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics*, 33(S1), S225–S255. Publisher: The University of Chicago Press.

Porter, C. & Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*, 12(3), 226–254.

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The More Things Change, the More They Stay the Same? Prior Achievement Fails to Explain Gender Inequality in Entry Into STEM College Majors Over Time. *American Educational Research Journal*, 49(6), 1048–1073.

Rios, I., Larroucau, T., Parra, G., & Cominetti, R. (2021). Improving the Chilean College Admissions System. *Operations Research*, 69(4), 1186–1205.

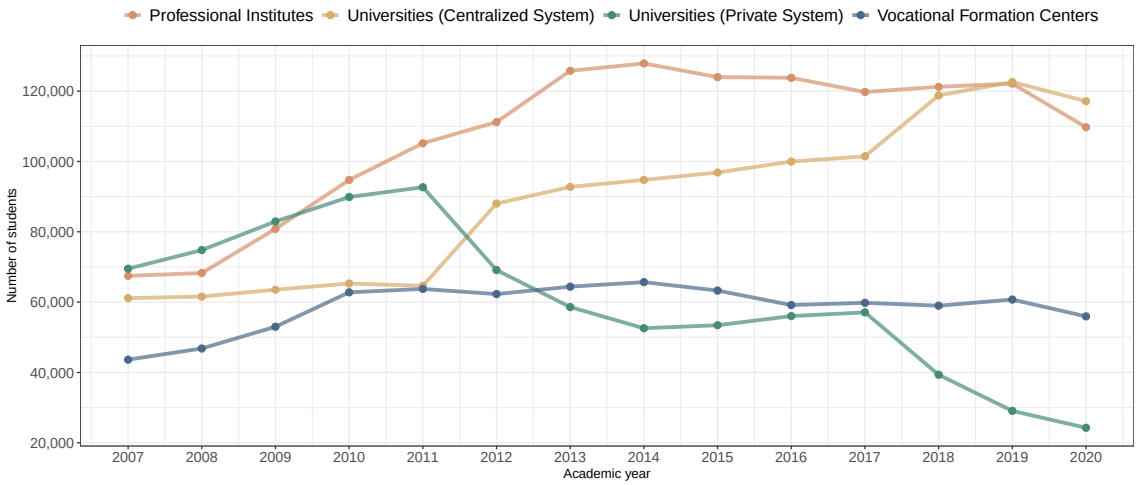
Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates*. *The Quarterly Journal of Economics*, 116(2), 681–704.

Sacerdote, B. (2011). Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education*, volume 3 (pp. 249–277). Elsevier.

Sloane, C. M., Hurst, E. G., & Black, D. A. (2021). College Majors, Occupations, and the Gender Wage Gap. *Journal of Economic Perspectives*, 35(4), 223–248.

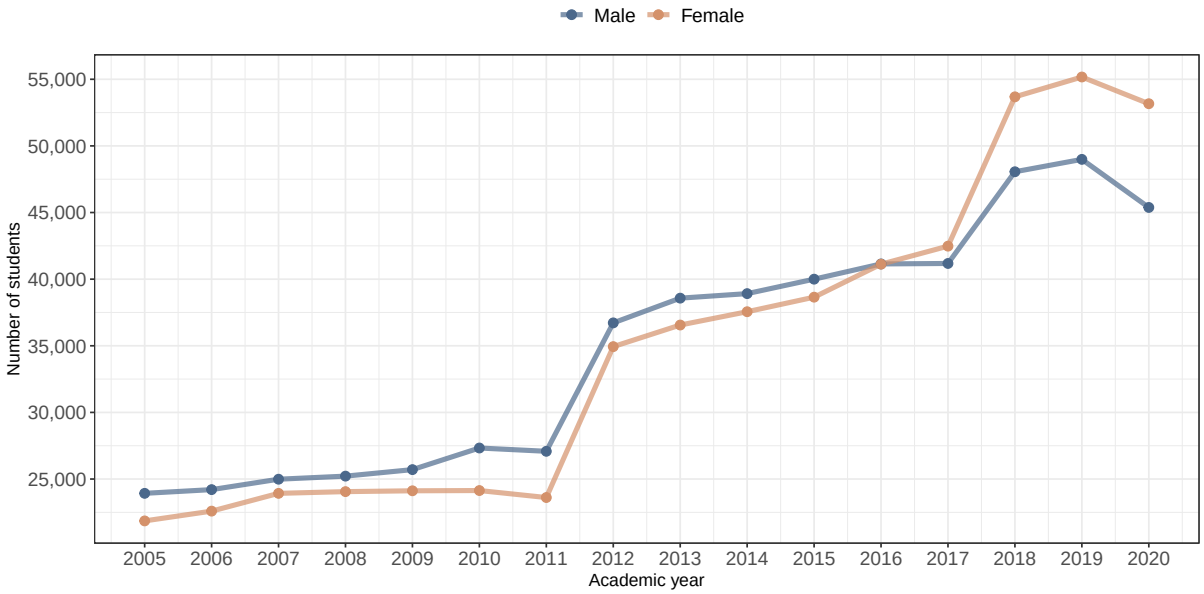
Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics*, 85(1), 9–23.

Figure 1: Total freshmen enrollment by type of higher education institution



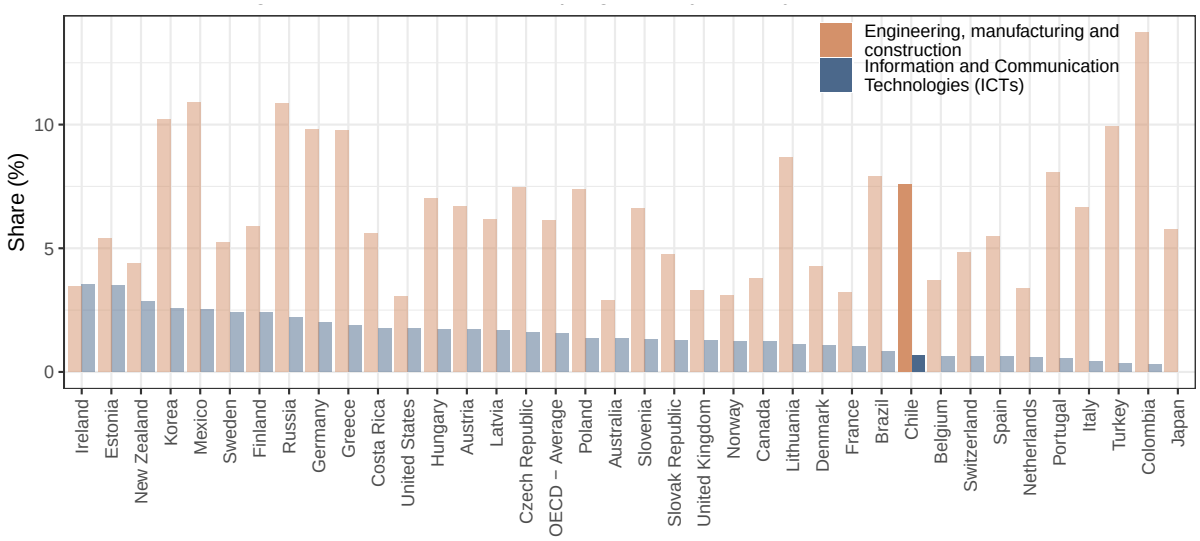
Note: This figure exhibits the total freshmen enrollment by type of higher education institution. Source: Ministry of Education

Figure 2: Freshmen enrollment by gender at universities from the centralized admission system



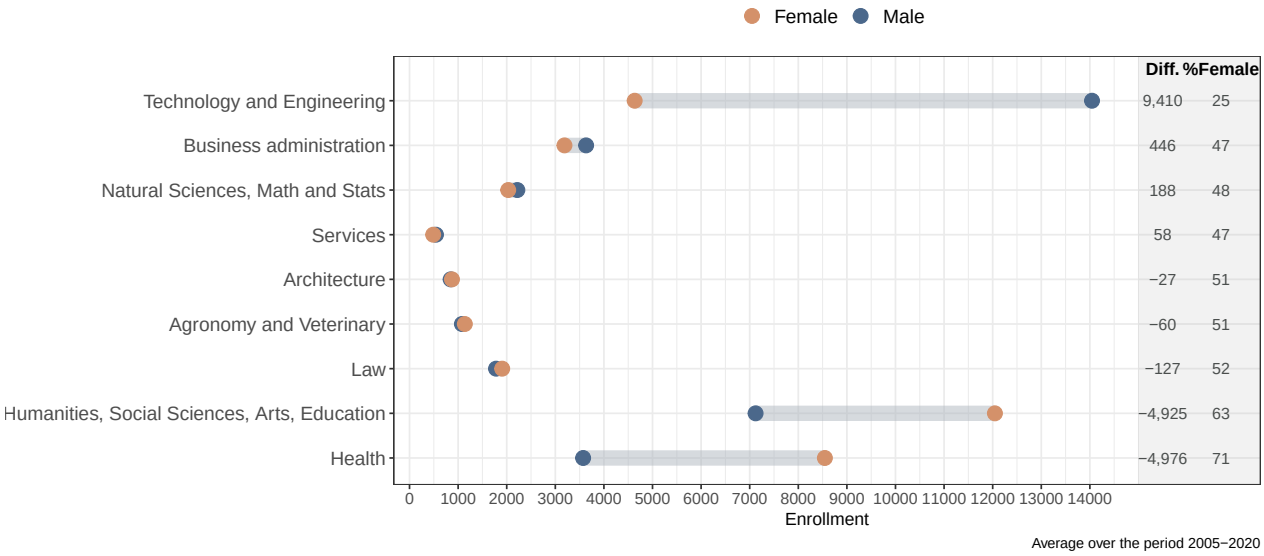
Note: This figure exhibits the total freshmen enrollment by gender at universities that are participants of the centralized admission system. Source: DEMRE administrative records.

Figure 3: Share of female graduates from bachelor's programs, by country and field



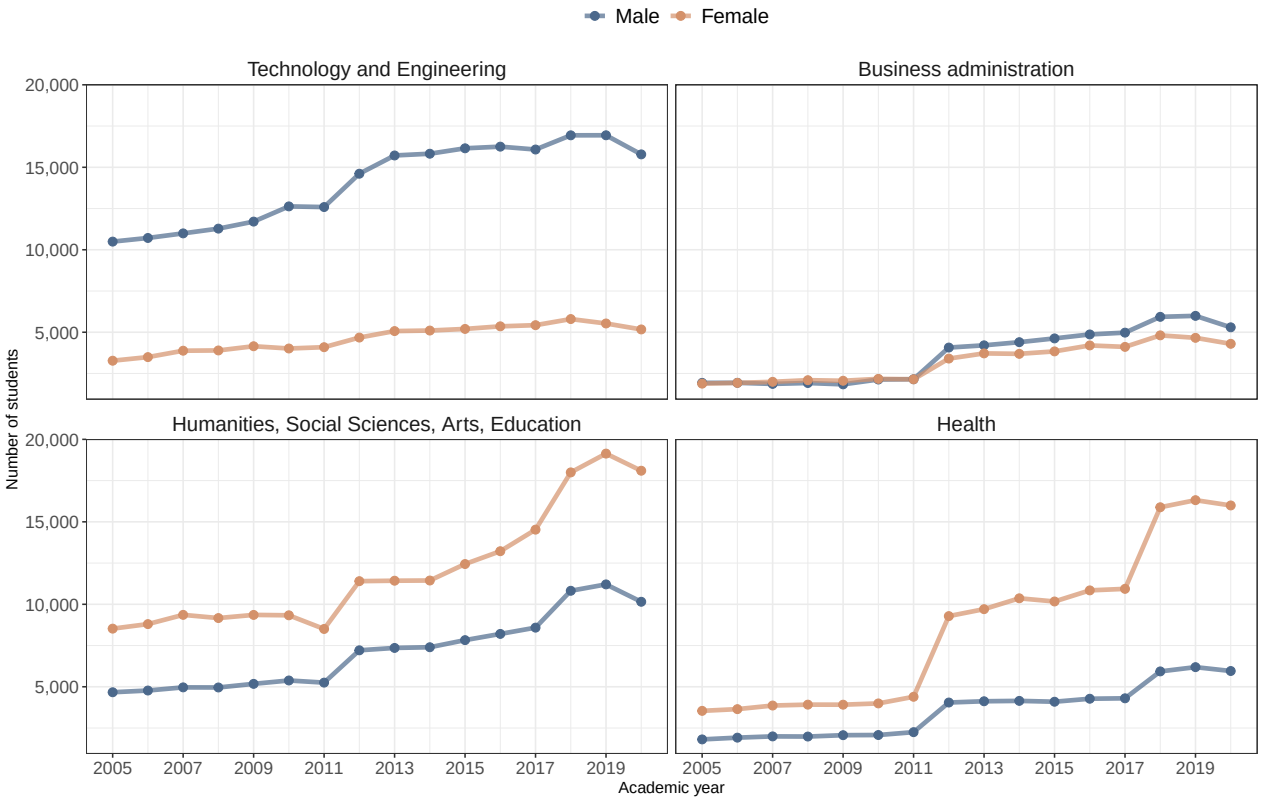
Note: This figure shows the share of female graduates by field across OECD countries. Source: Gender Data, OECD (2017).

Figure 4: Average freshmen enrollment by field and gender



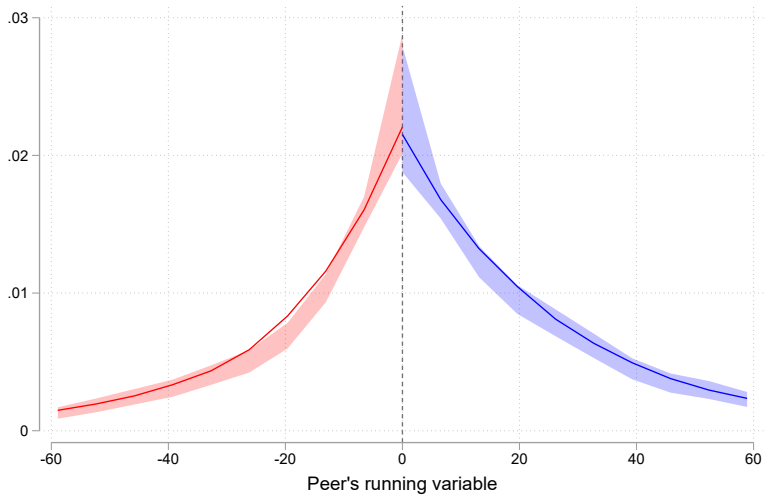
Note: This figure shows the average freshmen enrollment by gender at different fields of study. Grey bars represent the distance between men and female enrollment. The difference is depicted under the "Diff." column. The share of the female enrollment by field is depicted under the "%Female" column. For example, on average, there are 9,410 more men enrolled in Technology and Engineering than women. And women represent 25% of the enrollment at that field.

Figure 5: Enrollment by gender and field at programs from the centralized admission system



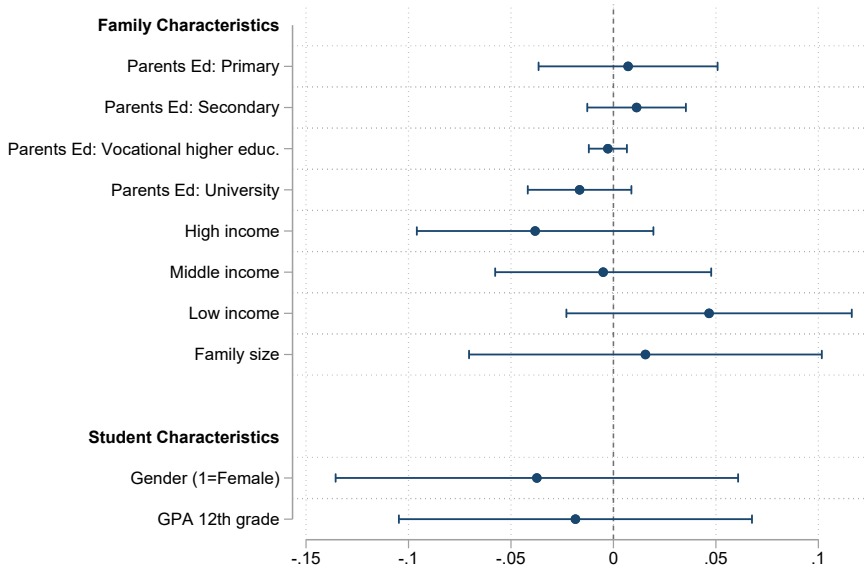
Note: This figure shows freshmen enrollment by year, gender and main fields of study.

Figure 6: Density plot of peer's running variable around the cutoff

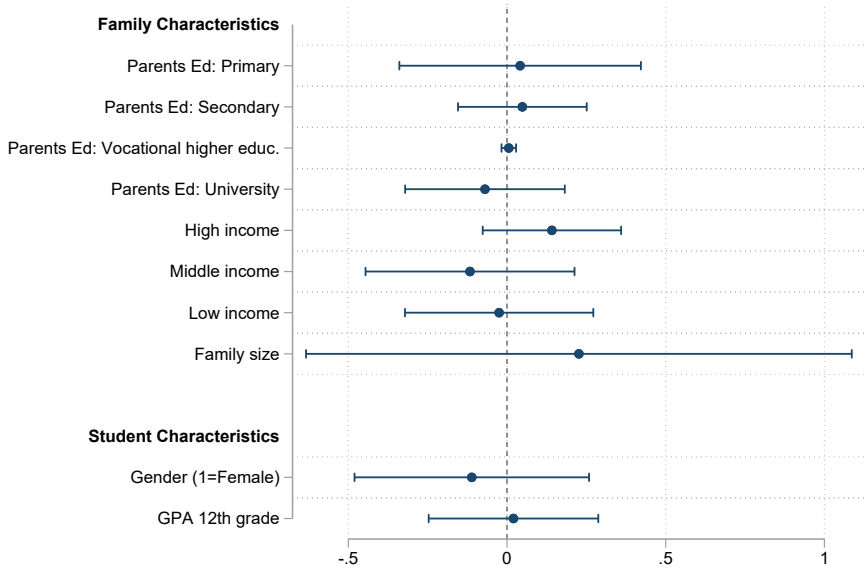


Note: This figure shows the density plot of the running variable around the cutoff, using the *rddensity* package that implements manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo et al. (2020). The robust bias-corrected test proposed by the authors have a *p-value* of 0.7206, which provides empirical evidence in favor of the continuity of the running variable.

Figure 7: Potential applicants and peers' continuity test around the cutoff



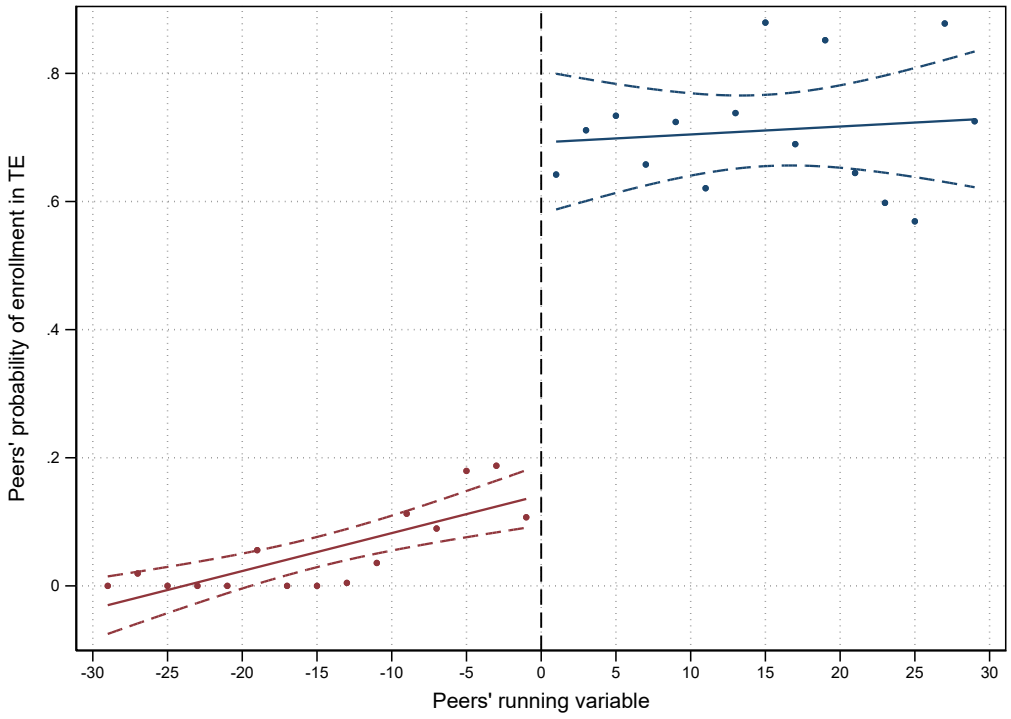
(a) Potential applicants' covariates



(b) Peers' covariates

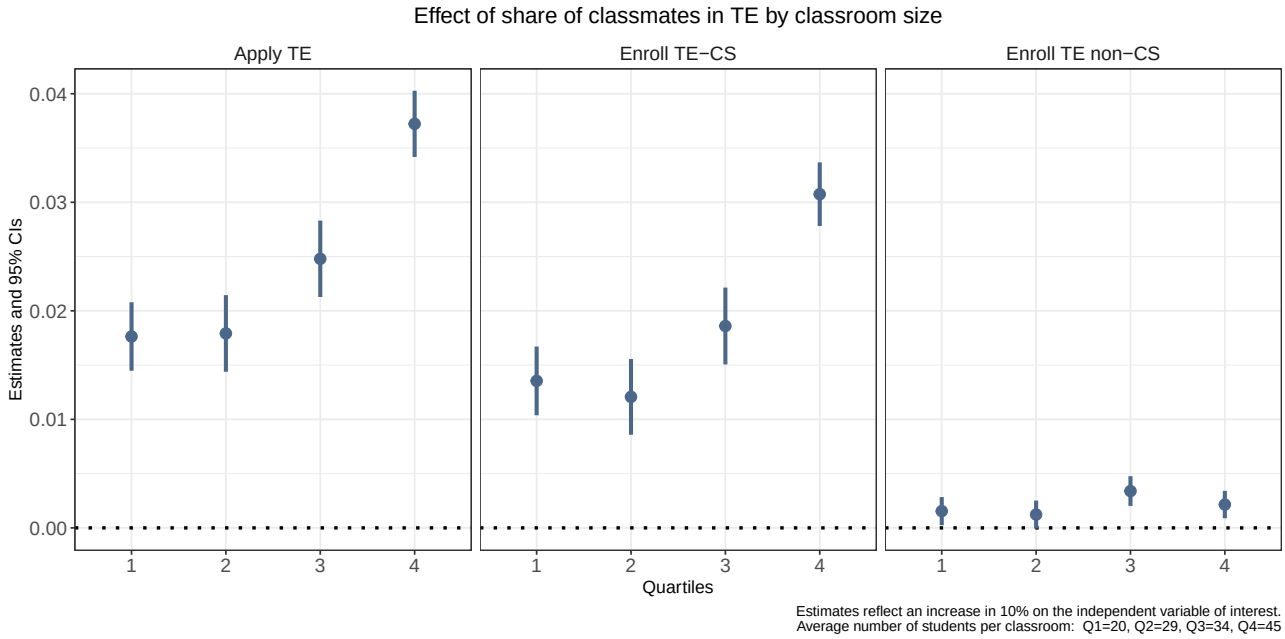
Note: This figure shows the bias-corrected treatment effects on baseline covariates, using confidence intervals obtained from robust standard errors following Calonico et al. (2014a,b). Each covariate was tested individually as the outcome, with optimal bandwidths chosen separately, as suggested in Cattaneo et al. (2019). Panel (a) and (b) present the point estimates and 95% CIs of potential applicant's covariates and peers, respectively.

Figure 8: First stage



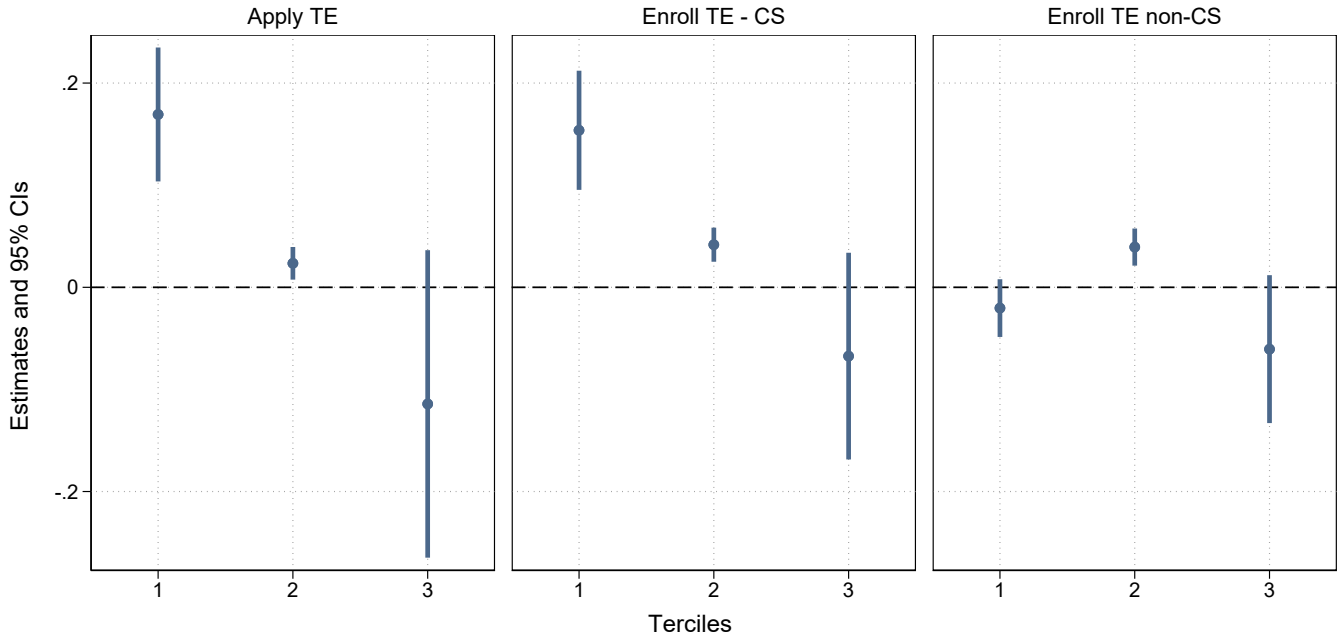
Note: This figure shows the discontinuity at the cutoff of the first stage represented in Equation 4.

Figure 9: Within-school variation specification, heterogeneous treatment effect by classroom size quartiles on main outcomes



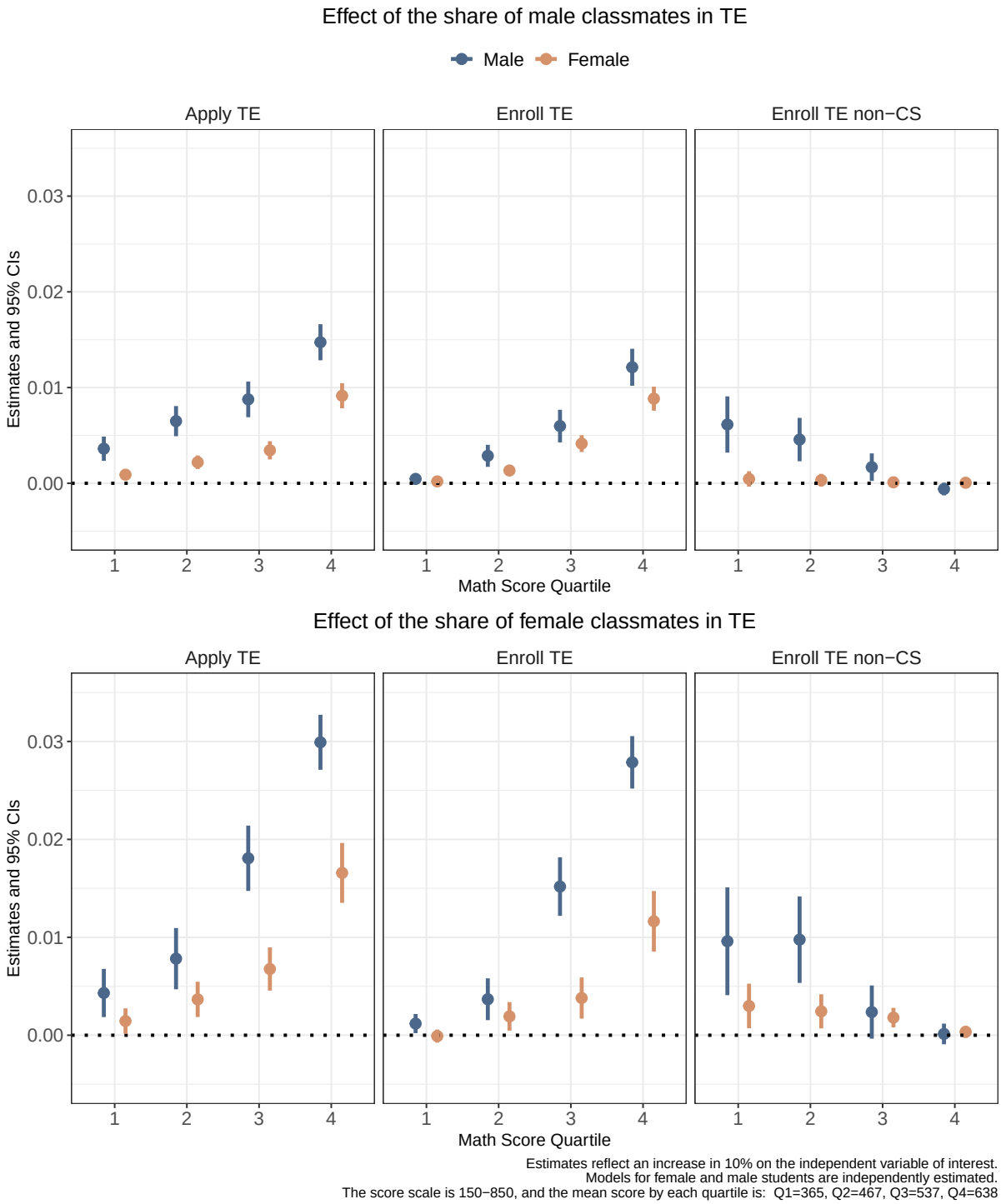
Note: This figure shows the effect of the proportion of classmates enrolled in Technology and Engineering on the main outcomes by terciles of the number of students in the classroom. Each coefficient and its 95% CI are obtained from separate regressions calculated for the corresponding quartile. All regressions include time fixed-effects, school fixed-effects, and school time trends. Standard errors are clustered at the school–classroom level.

Figure 10: Fuzzy RD specification, heterogeneous treatment effect by school size terciles on main outcomes



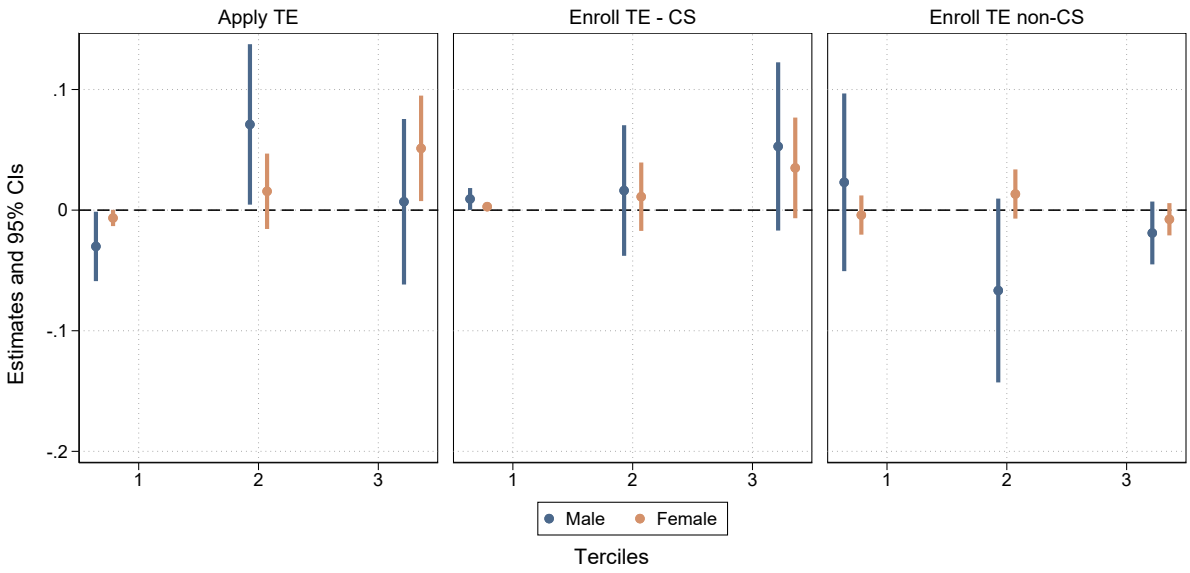
Note: This figure shows the effect of having an older high school peer marginally enrolled on Technology and Engineering on the main outcomes by terciles of the number of students in the school. Each coefficient and its 95% CI are obtained from separate regressions calculated for the corresponding tercile. All regressions include time fixed-effects and school fixed-effects. Standard errors are clustered at the school level.

Figure 11: Within-school variation specification, heterogeneous treatment effect by math score quartiles on main outcomes



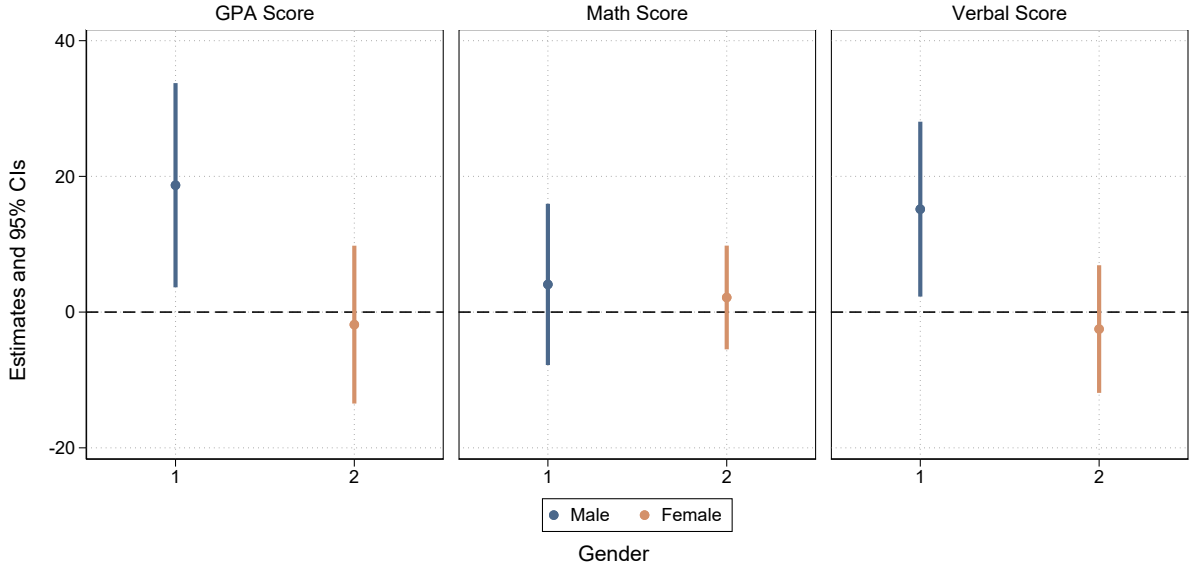
Note: This figure shows the effect of the proportion of classmates enrolled in Technology and Engineering on the main outcomes by terciles of the number of students in the classroom. Each coefficient and its 95% CI are obtained from separate regressions calculated for the corresponding quartile. All regressions include time fixed-effects, school fixed-effects, and school time trends. Standard errors are clustered at the school–classroom level.

Figure 12: Fuzzy RD specification, heterogeneous treatment effect by math score terciles and gender of the applicant



Note: This figure shows the effect of having an older high school peer marginally enrolled on Technology and Engineering on the main outcomes by terciles of math PSU score. Each coefficient and its 95% CI are obtained from separate regressions calculated for the corresponding tercile. All regressions include time fixed-effects and school fixed-effects. Standard errors are clustered at the school level.

Figure 13: Effect of marginal peer on outcomes related to GPA and test scores



Note: This figure shows the effect of having an older high school peer marginally enrolled in Technology and Engineering on GPA and scores in the college entry exam–PSU. Each coefficient and its 95% CI are obtained from separate regressions calculated for the corresponding tercile. All regressions include time fixed-effects and school fixed-effects. Standard errors are clustered at the school level.

Table 1: Descriptive statistics of senior high school students, total and by gender

	All sample				Male		Female	
Variable	Mean	SD	Min	Max	Mean	SD	Mean	SD
Female (1=Yes)	0.543	0.498	0	1				
Apply CS (1=Yes)	0.472	0.499	0	1	0.470	0.499	0.473	0.499
Enroll CS	0.284	0.451	0	1	0.307	0.461	0.265	0.441
Enroll CS - TE	0.083	0.276	0	1	0.136	0.343	0.039	0.193
Enroll CS - HASS	0.067	0.249	0	1	0.050	0.217	0.081	0.272
Enroll CS - Health	0.052	0.221	0	1	0.032	0.176	0.068	0.252
Enroll non-CS	0.490	0.500	0	1	0.469	0.499	0.508	0.500
Enroll non-CS - TE	0.066	0.248	0	1	0.125	0.331	0.016	0.127
Enroll non-CS - HASS	0.072	0.258	0	1	0.047	0.212	0.093	0.290
Enroll non-CS - Health	0.071	0.257	0	1	0.032	0.177	0.103	0.304
Apply 1st Choice TE	0.107	0.309	0	1	0.177	0.382	0.048	0.214
Apply 1st Choice HASS	0.118	0.322	0	1	0.088	0.284	0.142	0.349
Apply 1st Choice Health	0.127	0.333	0	1	0.076	0.265	0.170	0.376
PSU Math Score	502.143	107.297	150	850	516.545	109.702	490.023	103.692
PSU Verbal Score	498.761	106.216	150	850	500.334	107.464	497.436	105.136
GPA Score	549.770	101.099	208	835	537.262	101.289	560.297	99.730
Num. Observations	1574518							

Note: This table presents summary statistics of potential applicants that are in their senior year of high school between 2006 and 2019. PSU scores, are the score from the standardized test score with mean 500 and a standard deviation of 110. The minimum and maximum scores in the scale are 150 and 850 points, respectively.

Table 2: Descriptive comparison between schools with and without marginally admitted peers in Technology and Engineering

	School w/ marginal peers (N=6617)		School w/o marginal peers (N=24971)			
	Mean (1)	Std. Dev. (2)	Mean (3)	Std. Dev. (4)	Diff. (5)	Std. Error (6)
Female (1=Yes)	0.528	0.110	0.526	0.120	-0.002	0.002
PSU Test - Math	505.570	71.959	487.991	74.927	-17.579***	1.004
PSU Test - Verbal	501.604	68.758	483.997	73.139	-17.607***	0.964
Num. students	60.120	44.682	58.803	54.679	-1.317**	0.649
Num. students per classroom	28.704	9.496	26.954	10.117	-1.750***	0.133
Apply CS	0.463	0.261	0.398	0.270	-0.065***	0.004
Enroll CS	0.288	0.227	0.240	0.220	-0.048***	0.003
Enroll non-CS	0.466	0.228	0.485	0.220	0.018***	0.003
Apply TE - CS	0.162	0.112	0.141	0.116	-0.021***	0.002
Enroll TE - CS	0.083	0.078	0.069	0.076	-0.014***	0.001

Note: This table shows descriptive statistics and mean differences of the school characteristics for the sample of schools with marginally admitted peers to technology and engineering programs versus school without peers at that margin. The first sample corresponds to the analytic sample used in the fuzzy regression discontinuity approach. TE stands for Technology and Engineering and CS for Centralized System. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 3: Descriptive statistics for the analytic sample

	Mean (1)	SD (2)	Min (3)	Max (4)	N (5)
Potential applicants					
Parents Education - Primary	0.21	0.41	0.0	1.0	368,108
Parents Education - Secondary	0.45	0.50	0.0	1.0	368,108
Parents Education - Tertiary Vocational	0.11	0.32	0.0	1.0	368,108
Parents Education - Tertiary University	0.23	0.42	0.0	1.0	368,108
High Income	0.17	0.38	0.0	1.0	396,871
Mid Income	0.42	0.49	0.0	1.0	396,871
Low Income	0.41	0.49	0.0	1.0	396,871
Family size	4.30	1.76	0.0	57.0	396,871
Female	0.54	0.50	0.0	1.0	396,871
GPA Score	545.79	102.99	208.0	835.0	392,527
PSU math score	500.99	107.71	150.0	850.0	363,441
PSU verbal score	496.44	106.91	150.0	850.0	366,495
Peers					
Parents Education - Primary	0.16	0.36	0.0	1.0	370,166
Parents Education - Secondary	0.47	0.50	0.0	1.0	370,166
Parents Education - Tertiary Vocational	0.12	0.32	0.0	1.0	370,166
Parents Education - Tertiary University	0.26	0.44	0.0	1.0	370,166
High Income	0.18	0.38	0.0	1.0	284,927
Mid Income	0.38	0.49	0.0	1.0	284,927
Low Income	0.44	0.50	0.0	1.0	284,927
Family size	4.29	2.10	0.0	79.0	396,871
Female	0.26	0.44	0.0	1.0	396,871
GPA Score	584.85	92.54	315.0	830.0	396,871
PSU math score	569.08	72.02	355.0	850.0	396,871
PSU verbal score	536.41	69.21	353.0	829.0	396,871

Note: This table presents summary statistics for the estimation sample used in the fuzzy regression discontinuity approach, that corresponds 397,817 potential applicants and 6,617 unique peers. PSU scores are the score from the standardized test score with mean 500 and a standard deviation of 110. The minimum and maximum scores in the scale are 150 and 850 points, respectively.

Table 4: Estimates on Technology and Engineering (TE) enrollment

	Apply to TE		Enroll TE - CS		Enroll TE non-CS	
	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
% Classmates TE	0.041***	0.019***	0.035***	0.017***	0.002***	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
PSU Test - Math	0.154***	0.055***	0.130***	0.047***	0.008***	0.005***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
PSU Test - Verbal	-0.029***	-0.014***	-0.018***	-0.009***	-0.045***	-0.006***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
GPA Score	0.032***	0.017***	0.035***	0.015***	-0.022***	-0.001***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Num. Observations	719521	854997	719521	854997	719521	854997
Outcome Mean	0.1593	0.044	0.1225	0.0352	0.1198	0.0157
School and Year FEs	✓	✓	✓	✓	✓	✓
School time trend	✓	✓	✓	✓	✓	✓

Note: TE stands for Technology and Engineering and CS for Centralized System. All columns control for math score, verbal score, GPA score, high school annual attendance, family income, and parents' education. Standard Errors clustered at the School x Classroom level are in parentheses. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 5: Estimates on Technology and Engineering (TE) enrollment

	Apply to TE		Enroll TE - CS		Enroll TE non-CS	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)
% Male Classmates TE	0.015***	0.006***	0.013***	0.006***	0.000	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
% Female Classmates TE	0.027***	0.014***	0.025***	0.011***	0.001**	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
PSU Test - Math	0.155***	0.056***	0.131***	0.047***	0.008***	0.005***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
PSU Test - Verbal	-0.029***	-0.014***	-0.018***	-0.009***	-0.045***	-0.006***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
GPA Score	0.032***	0.017***	0.036***	0.015***	-0.022***	-0.001***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Num. Observations	719521	854997	719521	854997	719521	854997
Outcome Mean	0.1593	0.044	0.1225	0.0352	0.1198	0.0157
School and Year FEs	✓	✓	✓	✓	✓	✓
School time trend	✓	✓	✓	✓	✓	✓

Note: TE stands for Technology and Engineering and CS for Centralized System. All columns control for math score, verbal score, GPA score, high school annual attendance, family income, and parents' education. Standard Errors clustered at the School x Classroom level are in parentheses. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 6: Average effect of high school peer's TE enrollment on potential applicant's outcomes by gender

	Apply to TE-CAS (1)	Enroll TE-CAS (2)	Enroll TE non-CAS (3)
Peer enrolls in TE - CAS	0.014	0.015*	-0.026***
	(0.010)	(0.009)	(0.010)
First stage	0.294***	0.294***	0.294***
	(0.027)	(0.027)	(0.027)
BW Est. (h)	[16.2 ; 13.1]	[16.2 ; 13.1]	[16.2 ; 13.1]
Outcome mean	0.095	0.074	0.064
Number of applicants	108229	108229	108229

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and uniform kernel. Optimal bandwidths are chosen to be MSE optimal. All procedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering. All regressions include school and time-fixed effects.$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$

Table 7: Average effect of high school peer's TE enrollment on potential applicant's outcomes by peer and applicants' gender

	Apply to TE-CAS		Enroll to TE-CAS		Enroll to TE non-CAS	
	Female (1)	Male (2)	Female (3)	Male (4)	Female (5)	Male (6)
Panel A: Female Peer						
Peer enrolls in TE - CAS	0.066***	-0.085***	0.022***	0.012**	0.026***	0.036***
	(0.006)	(0.007)	(0.003)	(0.006)	(0.002)	(0.006)
First stage	0.559***	0.601***	0.559***	0.601***	0.559***	0.601***
	(0.017)	(0.014)	(0.017)	(0.014)	(0.017)	(0.014)
Outcome mean	0.041	0.158	0.033	0.120	0.015	0.120
Number of applicants	15954	12671	15954	12671	15954	12671
Panel B: Male Peer						
Peer enrolls in TE - CAS	-0.008	-0.023	-0.020**	0.011	-0.003	-0.063**
	(0.009)	(0.028)	(0.008)	(0.024)	(0.006)	(0.028)
First stage	0.297***	0.202***	0.297***	0.202***	0.297***	0.202***
	(0.027)	(0.026)	(0.027)	(0.026)	(0.027)	(0.026)
Outcome mean	0.043	0.155	0.035	0.122	0.016	0.121
Number of applicants	42623	36981	42623	36981	42623	36981

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and uniform kernel. All estimations use the average lower and upper bandwidths chosen optimally in Table 6. All procedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering, and CAS for centralized admission system. All regressions include school and time-fixed effects. Standard errors clustered at the school level are in parentheses. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 8: Average effect of high school peer's TE enrollment on potential applicant's outcomes by having high school peers enrolled in CAS

	Apply to TE-CAS (1)	Enroll to TE-CAS (2)	Enroll to TE non-CAS (3)
Panel A: Less than 15 high school students enrolled in CAS			
Peer enrolls in TE - CAS	0.040***	0.054***	-0.045**
	(0.012)	(0.012)	(0.018)
First stage	0.237***	0.237***	0.237***
	(0.024)	(0.024)	(0.024)
Outcome mean	0.058	0.037	0.079
Number of applicants	193679	193679	193679
Panel B: More than 15 high school students enrolled in CAS			
Peer enrolls in TE - CAS	0.014	0.016*	-0.005
	(0.010)	(0.010)	(0.005)
First stage	0.406***	0.406***	0.406***
	(0.036)	(0.036)	(0.036)
Outcome mean	0.134	0.113	0.050
Number of applicants	203192	203192	203192

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and uniform kernel. All procedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering, and CAS for centralized admission system. All regressions include school and time-fixed effects. Standard errors clustered at the school level.$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Online Appendix for

Peer Influence and College Major Choices in Male-Dominated Fields

Rocío Valdebenito

December 16, 2023

A Data construction

This section explains the data sources and steps needed to construct the final datasets used in the analysis. The section is divided in four parts, data sources, peers data, applicants data, and merging process.

A.1 Data sources

1. *Departamento de Evaluación, Medición y Registro Educacional* (DEMRE) 2004–2020. This organization is in charge of monitoring and implementing the admission to universities members of the centralized admission system. The datasets available from this agency contains student-level data of the entire centralized admission system. Particularly, the scores obtained from the whole universe of students taking the standardized college entry exam, self-reported socioeconomic information, applications to universities members of the system, the results of the submission process after the algorithm is implemented, and final enrollment.

2. DEMRE, programs' components weights, and number of seats per year and program over the period 2004–2020.

3. Student annual GPA, Ministry of Education. This datasets contains student-level variables of the annual GPA obtained by the students at the end of the academic year with the respective students and schools' identifiers.

A.2 Peers data

A.2.1 Identification of undersubscribed programs

A first step of the data cleaning consists of identifying programs that are undersubscribed, which are programs that were not able to fill their available seats completely in a given academic year. In these programs, I cannot identify their cutoff because any additional applicant would be admitted, regardless of the application score of the last admitted student.

Using data from the application process and the outcomes after the deferred acceptance mechanism, this step identifies, for each program-year, the list of admitted students and the list of non-admitted students. Therefore, after ordering the application weighted scores, I calculate the minimum score of the last admitted student, conditional on having at least one non-admitted student (i.e., below the admission cutoff) per program-year. Consequently, I can identify the cutoffs only for those programs where there is a 'waiting-list' of at least one student, as the available seats have been filled after the implementation of the algorithm.

A.2.2 Application data and cleaning procedures

First-time takers

It is important to note that students can take the college entry exam multiple times. The first step consists of combining all the years of the student-level datasets on scores and socioeconomic variables. If a student appears multiple times, this step keeps only the first time that the student took the test.

Eliminate undersubscribed programs from the list

A second step analyzes the datasets of the application list submitted by each student. This dataset is then merged with the information processed from the previous section where undersub-

scribed programs are identified. In this procedure, different cases can occur per student:

1. All the applications were made in oversubscribed programs: If that is the case, then all cutoffs across all alternatives are available.

2. If at least one alternative is made in an undersubscribed program:

   (a) And the admitted alternative is the undersubscribed program: In this case, the cutoff cannot be obtained for the alternative in which the student was admitted. Therefore, the running variable is impossible to calculate, and the entire list (student) is eliminated from the sample.

   (b) And the admitted alternative is not the undersubscribed program: In this case, the undersubscribed program does not affect the ability to calculate the running variable in the admitted choice. Therefore, if the undersubscribed program is just right below the admitted alternative, then the undersubscribed program would be used as the counterfactual for the admitted program. If, in contrast, the undersubscribed alternative is even further away (e.g., alternative 7th), then I eliminate the undersubscribed program from the list because its presence does not affect any target and counterfactual combination.

   (c) If the student was not admitted in any choice, then I eliminate the undersubscribed program from the list.

Identify relative selectivity and eliminate dominated alternatives

The main purpose of this step is to further clean the preference lists to have an ordered list where subsequent alternatives can be used as a counterfactual for a previous alternative. To accomplish this goal, this step borrows the concept of *relative selectivity* from Abdulkadiroğlu et al. (2014), Aguirre & Matta (2021), and Aguirre et al. (2022).

Relative Selectivity is calculated as $\phi_{ij} = \frac{s_{ij}-c_j}{\sqrt{\sum_l (\alpha_j^l)^2}}$, which represents the Euclidean distance between the applicant's scores (in math, language, science, history) and the admission frontier defined by the cutoff at program j. If the relative selectivity of a lower-ranked option is higher than that of a higher-ranked option, the relatively more selective option will be eliminated since

it does not serve as a proper counterfactual. In other words, higher selective programs submitted in lower-ranked positions will not represent a real scenario of what would have happened if the student is below an admission cutoff in the higher-ranked option.

The table below shows the resulting number of observations after running the iterative process of dropping dominated alternatives:

Table A.1: Number of observations at each iterative process of elimination

Step	Number of observations
iteration 0:	3,586,477
iteration 1:	2,471,055
iteration 2:	2,198,677
iteration 3:	2,140,679
iteration 4:	2,130,907
iteration 5:	2,129,686
iteration 6:	2,129,569
iteration 7:	2,129,552

Constructing target and counterfactual

After cleaning the preferences in the previous step, I construct pairs of target and fallback/counterfactual alternatives. Each alternative corresponds to an specific program-university combination. Table A.2 exhibits an example of a preference list after the relative dominated alternatives are eliminated. There are 6 preferences that survived the elimination procedure. The student is admitted in the third choice, because it is the first one where the score is above the program-specific cutoff.

Table A.2: Illustration of a preference list

Preference	Program	University	Score	Cutoff	Status
1	Civil Engineering	University of Chile	685	710	No admitted
2	Business	University of Chile	675	695	No admitted
3	Civil Engineering	University of Santiago	685	675	Admitted
4	Business	University of Santiago	667	650	Admitted above
5	Mechanical Engineering	Diego Portales University	680	620	Admitted above
6	Electrical Engineering	Andres Bello University	678	610	Admitted above

The main objective is to construct target and fall-back program pairs. The target program is the preferred program, while the fall-back or counterfactual represents the "what would have happened" scenario. However, if the student was not admitted to their target choice, then by construction, the fall-back choice is a representation of what actually occurred in practice. For example, in Table A.2, the first and second preferences are alternatives where the student was not admitted. Therefore, for each of those, the counterfactual scenario is the admitted choice (i.e., Civil Engineering at the University of Santiago). Moreover, for the admitted choice, the counterfactual would be the alternative just below, since the preference list is already cleaned from dominated choices and impossible choices are not present. In that case, the fourth preference represents what would have happened if the student's score was below the cutoff for their admitted choice.

Table A.3 presents the target and fall-back pairs obtained from the list presented in Table A.2.

Table A.3: Target and fall-back pairs

Preference	Target Program	Target University	Fall-back Program	Fall-back University
1	Civil Engineering	University of Chile	Civil Engineering	University of Santiago
2	Business	University of Chile	Civil Engineering	University of Santiago
3	Civil Engineering	University of Santiago	Business	University of Santiago