London School of Economics and Political Science

Essays on gender and higher education

Valentina Del Carmen Contreras Silva

A thesis submitted to the Department of Social Policy of the London School of Economics and Political Science for the Degree of Doctor of Philosophy.

London, September 2022

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 50298 words.

Statement of co-authored work

I confirm that Chapter 4 was jointly coauthored with Dr. Johann Koehler, Dr. Chiara Orsini, and Dr. Berkay Özcan, and I contributed 65% of this work. I present details on the extent of my contribution and that of my coauthors below:

- I designed the study and the survey with feedback and suggestions from coauthors. I collaborated in implementing the intervention for the second cohort of students. I processed the data, carried out the first set of analyses, and then a few more iterations with feedback from coauthors. I wrote the entire draft version of the paper, and revised it according to coauthors comments.
- Dr. Johann Koehler collaborated on the survey design by providing feedback and suggestions. Managed randomisation for the second cohort with Valentina's help and helped review and improve the paper. Overall contribution 7%.
- Dr. Chiara Orsini collaborated on the survey and study design by providing feedback and suggestions and contributed to obtaining administrative records from LSE. She collaborated with Valentina in analysing and discussing the results. Additionally, she assisted in careful reviewing of the paper and proposed various refinements to the draft proposal made by Valentina. Overall contribution 13%.
- Dr. Berkay Özcan collaborated on the survey and study design by providing feedback and suggestions. He managed the randomisation for the first cohort and obtained administrative records from LSE. Additionally, he assisted in careful reviewing of the paper and proposed various refinemnts to the draft proposal made by Valentina. Overall contribution 15%.

Declaration of Editorial Help

I can confirm that my thesis was copy edited for conventions of language, spelling and grammar by Clare Sandford.

Declaration of Funding and Support

This thesis was supported by ANID (*Agencia Nacional de Investigación y desarrollo*), Becas Chile Doctoral fellowship programme; Grant 72180240.

Acknowledgements

I would like to express my deepest gratitude to Prof. Stephen Jenkins and Dr Berkay Özcan for their invaluable supervision, support and unwavering patience. They have generously shared their knowledge and expertise throughout countless conversations along my PhD journey. Our discussions have deeply influenced this work and my identity as a researcher.

I am also thankful to Prof. Francisco Ferreira for his advice and mentorship. His insightful comments and suggestions have helped me improve this work, and his guidance has helped me gain clarity on the path ahead. Additionally, I would like to mention my coauthors, Dr Chiara Orsini and Dr Johann Koehler, to whom I am grateful for their collaboration, insightful suggestions and wise advice.

Finally, I would like to thank my colleagues and friends from the Social Policy department for their encouragement, support, and help. I would be remiss in not mentioning Magdalena Rossetti and Grace Chang; they have read my work numerous times and provided tremendously valuable contributions.

Abstract

I study horizontal inequalities in higher education, paying particular attention to gender imbalances. I evaluate policy interventions to understand how imbalances arise and how we can best correct them. Additionally, I study further consequences of the imbalances, primarily by analysing the interaction between advantaged and disadvantaged groups.

My first paper explains university admission systems' role in generating horizontal gender segregation in tertiary education. The paper exploits an admission system reform that improves application scores for students who graduate with a GPA higher than their high school average. I test whether the reform is neutral to gender and examine its effect on students' career choices. I find that (1) the reform improves women's application scores relative to men's, and (2) that students who benefit from the reform are more likely to apply and enrol in better-paid fields.

The second paper evaluates two gender affirmative action interventions implemented in engineering schools. First, I use a difference-in-difference approach to estimate the effectiveness of the policies. Then, a peer effects model to explore further consequences of the reform on students' academic performance and drop-out rates. I find (1) that both interventions successfully increased female participation in engineering school, (2) they led to lower first-year drop-out rates among women, and (3) indirectly improved performance in collaborative projects for men and women.

Finally, the third paper expands on the effects of peers on students' outcomes by examining interactions in small teams. The chapter tests how gender composition and the share of non-native English speakers affect students' academic performance, and perceptions of their team dynamics. We find (1) that an increased percentage of women improves students' grades and perceptions of whether they felt heard during discussions, and (2) the achievement gap between native and non-native speakers diminishes when more women and non-native peers are in the team.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Attending higher education has significant social and individual benefits, and it is often seen as one of the main engines of social mobility, and a driver for equality. University graduates tend to attain greater earnings than those who only completed high school, and they have better health outcomes (Lleras-Muney, 2005), better access to public services, and overall higher wellbeing (Powdthavee et al., 2015). Moreover, arguably, the opportunity of accessing university as a right on its own as it allows individuals to flourish. Although the benefits of attending higher education are clear, the relationships between education, social mobility and inequality are complex.

In the last few decades, we have seen a mass expansion of higher education in developed and developing countries worldwide (Machin and McNally, 2007). While it is reasonable to think that this would help reduce inequalities, higher education expansion has not generally led to greater income equality. Firstly, it is not necessarily the case that every group benefits equally from increasing access to higher education. Family background and other individual characteristics such as race and gender can often be important determinants of a person's educational opportunities and outcomes. Secondly, even if a person can access tertiary education, there are still important differences in terms of the type of education they receive.

Returns on higher education depend on several factors, such as years of schooling, quality of the institution, and field of study.

In this introductory chapter, I lay out how inequality in tertiary education materializes both as inequality in access, and as inequality in returns on higher education among different groups of students who do attend. In regard to the latter, I describe in detail the issue of segregation across fields of study, and how this contributes to perpetuating gender inequalities in labour market outcomes. Then, I outline policy responses to gender and other forms of inequality in higher education and summarize the approaches that the economic literature has taken to analyse and inform these policies. This overview sets my research into context, as in the following chapters I present three studies on gender and higher education. The chapter ends with an outline of the thesis.

1.1 The benefits of higher education

Higher education brings numerous benefits to those who go through it. The clearest and arguably most important of these is the earnings premium. On average, university graduates obtain substantially higher salaries than those with upper secondary education as their highest level of attainment. The average figures for OECD countries show that individuals who attain a tertiary degree obtain earnings that are on average 55% higher than those with only upper secondary attainment. The premium is even greater in countries with low shares of adults with tertiary attainment. For example, in Chile, Colombia and Costa Rica,[1] workers with a bachelor's degree earn more than twice as much as their peers for whom secondary education is their highest level of attainment (OECD, 2021b).

Moreover, earnings also increase within levels of tertiary attainment. On average, workers with a postgraduate degree earn more than those with a bachelor's or comparable degree, while workers with a bachelor's degree earn more than those

[1] In all of these countries, the share of adults with tertiary attainment is among the lowest across the OECD countries (about 25%).

with a short-cycle tertiary degree. Across OECD countries, adults who attain a master's or doctoral degree benefit from earnings that are on average 90% higher than those with upper secondary education. The differences in earning across these two groups can get significantly higher and, again, Chile is a case in point. Workers with a postgraduate degree earn on average 4.7 times more than workers with only a high school diploma (OECD, 2021b).

There are also employment status advantages associated with completing tertiary education. OECD (2021b) shows that in 2019, while 13% of adults with tertiary education were not employed, 21% of those with only secondary education were in the same situation. Thus, those without tertiary education not only earn less but are also more likely to not participate in the labour market at all.

In addition to the pecuniary benefits of higher education, graduates enjoy other, sometimes less expected, advantages. These include better health outcomes, enhanced social participation, and increased life satisfaction. For instance, an extensive body of research has documented a solid positive relationship between educational attainment and health outcomes (often referred to as the education-health gradient). See Grossman (2006), Lance (2011), and Clark and Royer (2013). More importantly, evidence suggests that higher education graduates live longer. For instance, on average, across OECD countries, at age 30, people with tertiary education have a life expectancy of five years more than those with less than upper secondary attainment (54 years versus 49 years).[2]

There is also strong evidence that higher levels of education are associated with greater civic engagement (Nie et al. (1996); Milligan et al. (2004); Glaeser et al. (2007), and Lance (2011).); higher likelihood of successful marriages (Aughinbaugh et al. (2013); and better mental wellbeing (Oswald and Wu (2011), and Powdthavee et al. (2015)).

[2]For a review of nonpecuniary benefits of schooling see Oreopoulos and Salvanes (2011).

1.2 The problem of access to higher education

Because of the crucial role of education in determining a broad set of economic and life outcomes, the question of who accesses higher education has far reaching consequences. On the one hand, if access is equally distributed among different groups in society, then tertiary education will inevitably help reduce inequalities. On the other hand, if disadvantaged groups systematically have lower levels of access to tertiary education, the effect will be the opposite. Unfortunately, the evidence suggests that the latter is true, as in all industrialised countries and in emerging economies there are differential transition rates into higher education for different groups (Björklund and Salvanes, 2011).

Family background is an important factor in determining educational attainment.[3] In a global overview, UNESCO (2022b) shows vast within countries inequalities in access to higher education between the rich and the poor. Similarly, children from families with high cultural capital are more likely to continue their studies after obtaining a high school diploma. For instance, Blanden et al. (2022) shows that for children of highly educated parents the probability of attending university at age 20 is 18 percentage points higher in Australia, and 28 points higher in the United States (compared to children from families in which neither of the parents have obtained education beyond high school).

Inequalities in educational attainment arise not only in terms of differential transition rates from secondary to tertiary education. Among those who access tertiary education, socioeconomic background is also linked to completion rates (see Blandin and Herrington (2022) for evidence from the USA). Additionally, there are differences in the length of the programmes that students from different groups pursue (short-cycle tertiary education versus bachelor's degrees), and differential rates of transition into postgraduate education (Boneva et al., 2022).

[3]See Björklund and Salvanes (2011) for a survey on empirical research on family background and education.

Why do inequalities in access to higher education exist?

One of the most critical barriers to access to higher education for some social groups is the unequal distribution of opportunities to either complete or attain highly in secondary school, and university admission exams. Differences in educational achievements among advantaged and disadvantaged groups appear early in life. Indeed, it is well documented that there is a connection between family background and student achievement in primary and secondary school.[4] In a cross-country comparison of international differences in educational attainment, Hanushek and Woessmann (2011) show that children from families at the bottom of the income distribution are at a disadvantage in all studied countries. Moreover, Hanushek et al. (2019) show that socioeconomic achievement gaps have not shrunk in the past 50 years.

Programme for International Student Assessment (PISA) scores provide additional insight into these differences. PISA measures reading and mathematics skills at age 15. The OECD (2019) report shows that for all countries with available data for 2018, there are significant gaps for reading and mathematics scores among students from families in the bottom and top quarters of the socioeconomic distribution (as defined by the PISA index of socio-economic status (ESCS)). Moreover, the size of the gaps is large even when compared to the overall differences in achievement between countries. Even in the best performing countries, the average achievement of students from the bottom quarter of socioeconomic status is below the OECD average. Migration status is also a determinant of achievement. In the same assessment, the reading performance of immigrant students was, on average, 41 points lower than their non-immigrant peers.

As access to higher education institutions is usually contingent on previous attainment, advantage and disadvantage accumulate. Those with worse opportunities in primary and secondary education are less likely to access higher education. However, differences in achievement in secondary education are not the only aspects that impede access to tertiary education for disadvantaged groups. Empirical evi-

[4] See Blanden et al. (2022) for a recent survey and a review of potential mechanisms.

dence shows that there are pronounced socioeconomic differences in access to tertiary education even conditional on intermediate measures of achievement such as test scores during high school (Belley and Lochner (2007), Blanden et al. (2022)). Other barriers are financial constraints (see Dynarski (2003), and Belley et al. (2014) for evidence on financial aid, and Caucutt and Lochner (2020) for evidence on parental investment); geographic location (Chetty and Hendren (2018), Celhay and Gallegos (2015)); differences in student's beliefs about the potential returns of further education (Boneva et al., 2022); differences in access to information and advice (McGuigan et al., 2016), as well as embedded familiar, cultural, and social expectations. Thus, the different circumstances that young people face after the end of secondary education are also an important driver of unequal educational opportunities.

1.3 The problem of heterogenous quality, and fields segregation in higher education

In the last few decades, we have observed a rapid expansion of higher education that has led to a widening of access for different social groups. In most industrialized countries, women now constitute either half or the majority of undergraduate students. At the same time, the proportion of first-generation university graduates is at an all-time high (UNESCO, 2022a). Moreover, there is a longstanding agreement that a better distribution of educational opportunities leads to a better distribution of the benefits associated with education, including income (see Mincer (1958); Chiswick (2003)). Why is it then that inequality is not falling as steeply as access to higher education is increasing? Research points to two main explanations: heterogeneity in institutional quality, and uneven occupational returns across different fields of study.

Firstly, on the issue of institutional quality, Hearn (1991) and Davies and Guppy (1997) have shown that individuals with higher socio-economic background have higher probabilities of entering selective and prestigious universities. Using data from the UK, Campbell et al. (2022) show that, at equivalent level of achievement, disadvantaged students enter high-quality universities at higher rates than students from more privileged backgrounds; Chetty et al. (2020) find similar results for the USA. Elite universities tend to have access to more funding and spend more per student. As a result, students who cannot gain admission to selective universities end up receiving a lower-cost education. More importantly, Chetty et al. (2017) shows that institutional quality affects graduates' occupational outcomes, and that higher quality universities often grant better employment opportunities. Thus, when access to high quality institutions is disproportionally granted to privileged groups, higher education aids the reproduction of inequality rather than its reduction.

Secondly, socioeconomic and gender segregation in fields of study, and consequently, in occupations is also a contributing factor in the reproduction of inequality. In a cross-national comparative study Reimer et al. (2008) shows that with higher education expansion, the differences in labour market outcomes of university graduates from different fields of study increase.[5] Again, students from higher socioeconomic backgrounds are slightly more likely to choose more prestigious and remunerative fields of study. But arguably, the group that is affected the most by occupational segregation is women, who are particularly underrepresented in the most lucrative subjects (Chavatzia (2017), Gerber and Cheung (2008)).

Although education has the potential to reduce inequalities across different groups of society, widening access is not a foolproof method to correct all disparities. Inequalities can persist when occupational returns depend on types of degrees, quality of institutions, and fields of study. The heterogeneity in returns across different types of universities and degrees, combined with the inequalities in opportunities to access the type of education that provides the highest compensation, are inevitably contributing to the reproduction of social inequalities.

[5] See also Van de Werfhorst (2008) for an overview of literature on fields of study and European labour markets.

1.4 Gender Inequalities in higher education: An overview

The case of gender inequality in higher education is a prime example of how widening access is not enough to lessen disparities in occupational outcomes. In most countries, women now constitute half or more than half of higher education graduates. According to Chavatzia (2017), women's participation rate in tertiary education almost doubled from 2000 and 2014, and by the end of that period, women already constituted the majority of students undertaking Bachelor's and Master's degree levels globally. However, the increase in women's participation has not been evenly distributed across disciplines. Consequently, despite the impressive changes observed in recent years, women remain a minority in certain fields of study.

While women constitute a majority in education, humanities, arts, and health-related degree programmes, men are more likely to enrol in, and graduate from degrees in engineering, mathematics, science, and business. According to OECD (2018), at a global level, men represent at least 71% of the students enrolled in the fields of information and communication technologies, and engineering, and 61% in construction and manufacturing. Meanwhile, in the fields of education and health and welfare, men represent at most 38% of new students.

These markedly different career paths between men and women have significant individual and social consequences. There are substantial differences in labour market returns across disciplines, and careers in the male dominated fields are also the ones with the highest economic returns (Arcidiacono, 2004). Thus, although women have achieved majority participation in tertiary education, segregation in terms of fields of study is still a major contributing factor to gender inequality in employment outcomes. In an empirical analysis of extent and trends of the gender wage gap in the USA, and a survey of international studies on the topic, Blau and Kahn (2017) concludes that a significant proportion of the sex gap is still attributable to occupational sex segregation and the concentration of women in relatively low-paying occupations.[6]

Among adults in full-time employment, women with higher education earn only a fraction of men's average income. For instance, in Chile, on average, women make about 68% of men's salary, while this number increases to 71% in the US, 76% in the UK, and 80% in Sweden (OECD, 2021a). Therefore, although on average women now have higher levels of education than men, men still enjoy better employment and earning outcomes from education.

Economic loss is also a consequence of gender inequalities. If talent is evenly distributed across the population, women systematically not pursuing careers in certain fields causes a talent misallocation. The latter generates an efficiency loss as employers end up with a smaller pool of talent from which they can recruit in male-dominated fields. Given the critical role of science, technology, engineering, and mathematics fields (STEM) in improving technology and production, the misallocation of talent is particularly hurtful for economic growth.[7]

An important question then is, why do women disproportionately pursue lower-paying occupations? There is no simple or single explanation that can be provided for the gender differences in terms of choice of field of study. However, a broad body of research has enhanced our understanding of these issues.[8] Several studies have focused on gender gaps in levels of academic preparedness to pursue degrees in the STEM fields. The existence of gender gaps in mathematics achievement at the school level is potentially linked to the lack of women in these fields. Although the gaps have closed in some countries, some differences favouring boys remain (Schleicher, 2019). Moreover, decisions of what courses or academic track to take at late stages of secondary education can have a direct impact on the likelihood of enrolling in a STEM degree-programme. Studying a Canadian university system, Card and Payne (2021) show that an important part of the gender differences

[6] See also Machin and Puhani (2003) for evidence on the role of gender segregation in field of studies on wage differentials in the UK and Germany.

[7] See Kolovich et al. (2020) for a survey of the literature on gender inequality and macroeconomics outcomes.

[8] For a comprehensive survey of this literature see McNally (2020).

in STEM enrolment can be traced to a gap in the proportion of girls and boys who fill the prerequisites to enter. However, although the gender gap in academic preparation is important, differences in achievement cannot fully explain the higher relative likelihood of majoring in sciences and engineering for men. In a study of subject choice in England, Cavaglia et al. (2021) show that at similar mathematics achievement levels, the different choices across genders persist. Correll (2001), Rapoport and Thibout (2018) and Justman and Méndez (2018) argue that the persistence of the gap at the same level of achievement is partially explained by female students requiring stronger prior signals of mathematical ability to choose male-dominated subjects.

An interesting branch of the literature on gender differences in mathematics achievement has focused on the social determinants of these gaps. Guiso et al. (2008) links cross-country measures of gender inequality with mathematic achievement gaps, and find a positive correlation between the size of the gap and the country's level of gender inequality. Similarly, Nollenberger et al. (2016) study achievement of second generation immigrants from 35 countries to provide causal evidence of the role of culture on determining the mathematics achievement gap. The study finds that the math gender gap decreases for children whose parents come from countries with higher levels of gender equality. Other social influences come directly from teachers, and role models. Studying Israeli schools, Lavy and Sand (2018) find that teachers' gender biases affect academic achievement and course-taking choices for secondary school students. Additionally, using data from the USA, Carrell et al. (2010) find that female students' performance in math and science classes and their likelihood of graduating with a STEM degree improves when they have a female.

Other than differences in mathematics achievement, one possible explanation for segregation in occupational placement is gender differences in preferences. Studies of high school and college students have found that young women attach greater importance than men to the intrinsic, altruistic, and social rewards associated with an occupation, whereas male students place a higher value on such extrinsic rewards as money and power (Wiswall and Zafar (2015), and Wiswall and Zafar

(2018)). Similarly, in a study of USA college students, Zafar (2013) find that men and women have similar expectations of their university experience, but they differ in terms of their tastes regarding the workplace. Women, more than men, value nonpecuniary outcomes in the labour market, such as enjoying tasks in their jobs, and being able to reconcile work and family. In a similar vein, Buser et al. (2014), and Buser et al. (2017) point to gender differences in competitiveness as an important factor in choosing mathematics and science intensive academic tracks in the Netherlands, and Switzerland respectively.

Wiswall and Zafar (2015) point out that boys and girls also have different expectations of future earnings, which (in addition to preferences) can explain the gender differences in career choices. In fact, according to OECD (2020a), labour markets do not reward men and women equally, even when they both have the same degree. The gender gaps in the employment rates within STEM fields were the largest across all fields of study on average across OECD countries in 2020. While 93% of men with a degree in information and communication technologies (ICT) were employed, only 81% of women were. Therefore, gender differences in perception about returns might be well founded.

1.4.1 How can we solve gender inequalities?

The countries that have achieved high levels of gender equality are also the countries with the smallest gaps in mathematics achievement between boys and girls. Moreover, studies have demonstrated that even small challenges to gender norms, like exposure to female role models with a background in science, increases the share of girls enrolling in STEM programmes. If gender norms are the main driver of differences between girls' and boys' mathematics achievement, and occupational preferences, then, change in societal norms is perhaps the ultimate solution to gender inequality in higher education and in the labour market.

Although shifting social norms is important, cultural changes are slow. Relying on modifying gender norms alone is a bet that will only start to pay off for future generations, as in most cases these types of changes take years to have significant effects. A shorter-term option, that can help reverse the effect of past discrimination for the current generation and that has immediate measurable results, is the implementation of affirmative action policies in university admission.

Affirmative action refers to a set of practices undertaken by employers, universities, and government agencies to provide preferential treatment to groups who have been treated unfairly in the past, and to promote equal access to the services and benefits of a society. Universities across the world use these programmes in their admissions processes to deal with inequalities in their student bodies by going beyond non-discrimination.

The affirmative action policies that have been implemented in universities usually aim to reduce disadvantages related to race, caste, or class. In contrast, affirmative action policies aimed at closing gender gaps are considerably less common. Because women do not constitute a minority in higher education, the lack of initiatives to increase female participation is reasonable. However, affirmative action policies can help reduce gender segregation if implemented in the specific fields where women are underrepresented.

There are many possible methods for universities to raise the share of minority students, including:

- Increasing the share of minority high school students who qualify for admission (e.g., by changing admissions criteria).
- Increasing the application rates of minorities who qualify for admission (perhaps through targeted recruiting).
- Accepting a higher share of minority students who apply (quota policies)
- Increasing the rates of minority students that accept an offer after it is made (perhaps through targeted financial aid).

- Providing support systems to help retain and graduate minority students (such as tutoring).

As there is a wide range of possible interventions, the level of effectiveness and the potential consequences vary across policies. The programmes that have proven to be most effective are the ones that set a quota for a specific group, establishing the minority's participation rate by construction. However, they are not always easy to implement.

In the USA, for instance, consideration of the applicant's race in admission decisions has motivated many court cases.[9] In some states, the intensity of the controversy has resulted in affirmative action bans. Besides effectiveness, policymakers care about the further consequences of affirmative actions. Are the beneficiaries academically prepared? Do the policies harm the intended beneficiaries? Thus, the type of intervention chosen depends on the context, possibilities, effectiveness, and potential further consequences.

What do we know about affirmative action?

While there is an extensive body of literature on the effectiveness of, and issues related to, affirmative action policies, most research has focussed on programmes that target race, class, or caste disparities. Although there is little evidence on the effectiveness of gender affirmative action policies, the academic debates surrounding other types of affirmative action policies is illustrative. They have centred on four main issues:

1. The targeting properties of the policies

   The issue of targeting relates mostly to how effective affirmative action policies are. There are two main concerns surrounding this issue. First, how

[9] See Long (2007) for an overview of these court cases and a summary of the history of affirmative action policies in the US.

well is the policy doing in improving the share of minority students in their institutions? This question is particularly relevant for policies that do not set a specific participation rate for the underrepresented group. The bans on affirmative action in the USA have provided a good scenario for researchers to measure their effectiveness once they are removed. In this strand of literature, Arcidiacono (2005), Howell (2010), and Hinrichs (2012) show that the affirmative action bans significantly decreased minority representation.

Second, some scholars have argued that targeting should be based on income (rather than race or ethnicity, for example) because family income is a strong predictor of performance. If admission systems allow only wealthier students within the minority group to traverse the (lower) hurdles required for admissions, then they may be displacing poor students from the non-minority or general group. In light of concerns around fairness, targeting analysis can be very useful. In the Indian and Brazilian contexts, Bertrand et al. (2010) and Francis and Tannuri-Pianto (2012a) have shown that affirmative action successfully targets the financially disadvantaged. In the Indian case, however, it may lead to the exclusion of other disadvantaged groups. For example, caste-based targeting reduces the overall number of women entering engineering colleges.

2. Possible harm to intended beneficiaries by placing them in academic programmes for which they are ill-prepared

   A second relevant issue surrounding affirmative action policies is the potential harm to intended beneficiaries. The main concern arises when the interventions involve setting lower academic requirements for minority applicants to be granted access. Under the assumption that the academic credentials accurately reflect knowledge and skills, the beneficiaries of the intervention will start further behind than those admitted under regular admissions criteria. Thus, the concern is how does the gap between these two groups change as they progress through tertiary education? Do the beneficiaries catch up or fall further behind?

   The academic literature is inconclusive. Frisancho and Krishna (2016) studied a caste quota policy in an elite engineering college in India. They found that minority students, especially those in more selective majors, fall behind their same-major peers in terms of grades as they progress through college. On the other hand, analysing a similar quota policy, Bagde et al. (2016) state that they did not find any evidence of harm to intended beneficiaries when they were placed in academic programmes for which they might have been ill-prepared. Fischer and Massey (2007) provided evidence from the USA, and found that affirmative action policies enhanced the academic achievement of minority students.

3. Implications for their labour market outcomes

   Even if the gap in academic performance persists, the intended beneficiaries can still gain from the intervention. A question that might better reflect the concern about harm to intended beneficiaries is: Would students who do not qualify for regular admission do better if they enrolled in schools and majors that are more aligned with their credentials? Then, if there is a severe mismatch, it would be fair to conclude that preferential treatment may do more harm than good. In order to answer this question, scholars have looked at the labour market outcomes of the beneficiaries of affirmative action. For instance, studying caste quotas in India, Bertrand et al. (2010) found positive earning effects for lower-caste students attending engineering colleges via quotas. Similarly, Bleemer (2021) analysed the effect of bans on race based affirmative action in the USA and found that the ban led to a decline on minority applicant's wages.

4. Pre university effort

   Finally, a less explored topic is the issue of pre university effort. This concern relates to the consequences that the implementation of affirmative action policies in higher education admissions can have for secondary education students. The argument is that high school students who expect to receive preferential treatment in the admission process to higher education might reduce their efforts to be accepted into an institution. Up to now, the evidence suggests that this is not happening. For instance, Francis and Tannuri-Pianto (2012b) looked at the preuniversity effort of applicants or students in regard to a Brazilian race affirmative action policy. Their findings suggest that the applicants did not reduce their effort after the implementation of the racial quotas.

Understanding the issues of targeting, catch-up, mismatch, and pre-university effort is useful because it sheds light on how these policies affect their beneficiaries. However, it is also essential to understand the effects of affirmative action policies on the non-beneficiaries (students that are either not members of the minority group or that were admitted without the use of the policy). Successful affirmative action policies generate changes in the composition of the student body, as increasing diversity is often a fundamental part of their aim. In this process, all students, beneficiaries and non-beneficiaries are affected.

Students interact with and influence each other, and demographic characteristics such as students' gender are important in determining how that influence is manifested. And although the effects of interactions among diverse groups have not been studied in the light of the consequences of affirmative action policies, this is still a topic of public and academic interest. Consequently, various studies have explored the effects of peer and group composition in contexts where diversity occurs naturally, i.e., in the absence of interventions.

Compositional effects

The idea that peers matter, specially at a young age, is fairly intuitive. The premise is that the behaviour of one member of a group can affect or influence the behaviour of the rest. The economic literature has extensively studied the impact of interactions with peers on individual outcomes, especially in terms of the association between the academic performance of students and their classmates (see Sacerdote (2011) for a survey of education peer effects).[10]

[10] Peer effects have also been studied in contexts other than the classroom. For example, there are several studies that have investigated peer effects among citizens and neighbours on a range of dimensions, including health, body weight, work, and consumption. See Durlauf and Ioannides (2010) for a review of this literature.

The idea that peers' demographic characteristics such as gender (as opposed to performance) can influence outcomes too, is perhaps less intuitive. Still, research in primary and secondary school settings has shown that gender peer effects do arise in classrooms. Several studies suggest that an increased share of girls in the classroom is associated with better student performance. See for instance Hoxby (2000) for evidence from the USA, Lavy and Schlosser (2011) for Israel, Li et al. (2013) for Turkey and Anil et al. (2016) for evidence from China.

Although there is scarce research on gender peer effects in higher education, some evidence suggests that positive effects arise from increasing the share of females peers. For instance Booth et al. (2018) found that young women assigned to all-female classes in their first year of university were roughly 57% less likely to drop out and 61% more likely to get a top ranked degree under the UK system.[11] Therefore, gender segregation not only affects the women who do not manage to enter male dominated fields, but also the ones who do.

If demographic characteristics and diversity matter for student outcomes, this means that there are further consequences of the existing vertical and horizontal inequalities in tertiary education. This points to the importance of studying the dynamics of peers in the light of both segregation, and policy intervention that aim to increase diversity.

1.5 Aims and contributions of the thesis

This thesis contains three substantive papers, set out in Chapters 2 to 4, and a conclusions chapter. In each paper, I study issues related to horizontal inequalities in higher education, putting particular focus on inequalities related to gender.

I evaluate policy interventions to understand how we can best correct these inequalities. I also study further consequences of the inequalities, primarily by analysing

[11] See also Shan (2022), and Stoddard et al. (2020) for evidence on the effects of women's minority status in tertiary education settings.

the interaction between advantaged and disadvantaged groups to provide a more comprehensive view of what is at stake when deciding to intervene.

Chapter 2 explains admission systems' role in generating horizontal gender segregation in tertiary education and the Chapter 3 analyses two interventions that reduce gender inequalities in engineering schools. I show the effectiveness of these policies and provide evidence on how the policies might potentially affect beneficiaries and non-beneficiaries by investigating the effect of gender composition on student outcomes. Finally, Chapter 4 expands on the effect of the interactions between different groups on student performance by studying the effects of gender composition and the share of native speakers on various individual outcomes.

Overall, my research provides a better understanding of the causes and consequences of segregation in higher education and ways in which we can revert the inequalities and their adverse effects. My thesis contributes to the literature on inequality in higher education by helping explain how inequalities arise from an angle that other studies have not yet explored: the role of admission systems. Admission systems have been overlooked as a determinant of segregation in higher education. I argue that university admissions systems can be biased against certain groups, even when practitioners rely on objective factors.

The thesis also contributes to understanding how to reduce gender segregation in higher education by studying affirmative action policies' effectiveness and consequences. Additionally, by studying compositional effects on student outcomes, my research helps to understand (1) further consequences of the underrepresentation of women in scientific fields and (2) the consequences of the policies implemented to improve this uneven distribution. As the evidence on affirmative policies in gender is scant, this chapter is particularly helpful in informing policymakers. Finally, looking at the issues of gender and linguistic composition helps to better understand the effects of mixing and interacting with students from different backgrounds.

1.6 Outline of the thesis

Chapter 2 presents the paper titled "The role of admission criteria in reducing gender inequalities in higher education". This paper exploits the sudden timing of a reform in the Chilean university admission system. The reform introduced a class ranking criterion that improves application scores for students that graduate with a GPA higher than their high school average. I argue that because the reform increases the importance of low stakes measures of achievement, it can benefit women. Thus, it can help reduce gender disparities in fields of study. First, I test whether the reform is neutral to gender by comparing men's and women's application scores. I also identify which students benefited from the reform by comparing applicants' admission priorities with the reform and a counterfactual scenario. I find that a larger proportion of women than men saw an improvement in their admission priority as a result of the reform. Then, I examine the effect of the reform on students' applications to fields, and I find that the policy change affected their preferences. Men and women who benefited from the ranking bonus were less likely to apply to, and enrol in programmes in lower-paid fields and more likely to apply to degrees in medicine and engineering instead. The effects were persistent across different levels of achievement in admission tests and did not significantly differ by gender. These findings, taken together with the targeting results, prove that the new criterion has a role in reducing gender inequalities in university.

Chapter 3 presents the paper "Policy evaluation of gender affirmative action in engineering schools". The paper evaluates the impact of two separate but contemporaneous efforts to increase female participation in engineering schools by two leading universities in Chile. I use a difference in difference approach to estimate the effectiveness of the policies and a peer effects model to explore further consequences of the reform on students' academic performance and drop-out rates. The paper finds that (i) both policies were successful in increasing women's enrolment in, and attendance at their engineering courses; (ii) they did not change the average academic ability of accepted students (as measured by their average composite application scores); (iii) the UCH policy led to lower first-year drop-out rates among women and improved performance in collaborative projects by both men and women, and had no significant effect on other first-year educational outcomes.

In Chapter 4, I present a paper written in collaboration with Dr. Chiara Orsini, Dr. Berkay Özcan, and Dr. Johann Koehler. In this paper, titled "Effects of team diversity on performance, perceptions, and predictions: Experimental evidence of gender composition and language", we randomised postgraduate students taking a common course in a UK university into small teams to investigate the role of gender composition and the share of non-native English speakers in four sets of outcomes: (1) academic outcomes (exam marks and classification), (2) self-perception of voice, (3) leadership role in the team, and, (4) expectations of academic performance. The Master's programme where our study took place has a significant proportion of non-native English speakers (close to 50%), plus a large share of women (80%). We used administrative and survey data to measure students' perceptions of teamwork and academic performance expectations. As we had exogenous variation in terms of group composition, we were able to estimate causal effects of group diversity on the outcomes of interest. We found compositional peer effects for all outcomes. For example, an increased share of women positively impacted students' grades and their perception of whether they felt their voice was heard in their team. We also found that non-native speakers had a lower performance than their native counterparts. However, the gap reduced when they had more women and non-native peers. Finally, we found no detriment to men in the minority; on the contrary, men in groups with more women were more likely to report taking leadership roles.

Finally, Chapter 5 presents conclusions derived from the thesis's main findings and related literature. First, I provide a brief summary of findings and contributions, then discuss my research and its potential policy implications. I focus on two main issues: reducing gender segregation in higher education and the importance of understanding peers' role on individual outcomes. Lastly, I provide some suggestions for avenues for future research.

Chapter 2

The role of admission criteria in reducing gender imbalances in higher education

2.1 Introduction

Despite the significant advances in female participation in higher education, gender segregation between fields of study persists. On average, among students enrolled in degrees from the fields of science, technology, engineering, and mathematics (STEM), only 35% are women. In contrast, women constitute 60-70% of students in arts, humanities, and social sciences (Chavatzia, 2017). This disparity deprives women from reaping the benefits of the higher economic returns that STEM subjects yield (Gerber and Cheung (2008), Arcidiacono (2004), Blau and Kahn (2017)). Moreover, if capable and talented women tend not to choose careers in the STEM fields, the talent pool from which companies in STEM fields may select employees is restricted, which might end up affecting economic efficiency, and growth (Xie et al., 2015).[1]

The literature that has tried to explain the reasons behind the gender imbalance in higher education has mostly focused on students' behaviors and gendered preferences. My paper adds to this literature by focusing on a structural factor, university admission systems.

In general, admission systems rely on two main elements: one-shot, high stakes admission exams, and measures of continuous assessment over a longer time horizon, such as high school grades. Prior studies have shown that, on average, women tend to perform worse than men in standardized tests, especially in mathematics (see Schleicher (2019) for evidence on PISA scores). However, the performance gap disappears when it comes to non-competitive or lower-stakes tasks (Cai et al. (2019), Schlosser et al. (2019), Iriberri and Rey-Biel (2019), Montolio and Taberner (2021)). Therefore, the decision regarding how much weight we give to each of these two measures of academic performance can have gendered consequences.

My empirical strategy exploits a reform implemented in Chile to the university admission system to analyze whether admission criteria, even when based wholly on objective factors, have a role in reproducing sex segregation in higher education. In Chile, university admissions are centralized and vacancies are limited. As in many countries, students not only specify the university but also the specific degree program they wish to attend in their applications, and admittance depends on their application scores. Until 2012, application scores were based solely on a national-level standardized exam and students' high school grades. However, the reform implemented in 2013 mandated universities to consider students' high school rankings as an additional admission criterion. High school rankings are a measure of the relative position of the students compared to their peers; hence, its mandatory inclusion increases the weight of measures of continuous assessment on students' application scores. Therefore, although the Chilean Council of Rectors designed the reform intending to democratize education and make it available to

[1] Xie et al. (2015) argue that scientific fields have an uncontested role in developing technological innovation and promoting economic growth. For a recent review of the literature on the relationship between gender inequality and macroeconomic outcomes see Kolovich et al. (2020).

a broader set of socioeconomic groups (CRUCH, 2019), I hypothesize that its inclusion might have had effects on (1) gender differences in applications scores and (2) the proportion of women attending more in demand and better-paid degree-programs, such as the ones in the STEM fields.

There are two crucial features of the reform that provide identification and allow me to test my hypothesis. First, the ranking score is an exogenous shock to the likelihood of gaining admission to more prestigious degree programs. Second, the computation of the ranking score generates variation among applicants because the GPA at which a student gets the ranking bonus varies from one school to another. In the analysis, I restrict the data to the first year of implementation of the policy to avoid endogeneity problems that strategic behavior from students might cause.

To understand the effect of the change in admission criteria, I answer two questions. First, is the reform neutral to gender? And second, how does benefiting from the reform affect students' study choices?

I use two strategies to answer the first question. First, I estimate a linear equation to identify differential trends in the application scores between men and women before and after the reform. Second, I simulate students' application scores without the ranking criterion and compare them with their actual final scores. While the differential trends analysis allows me to examine the gender gaps, the counterfactual analysis enables me to identify individuals who have benefited from the reform. I find that the reform has increased women's application scores relative to men. While on the first year after the implementation of the reform 78% of women had better application scores, the figure for men was only 67%. These results demonstrate that the reform has benefited women more widely than men and thus that it has had unintended gendered consequences.

Having established the effects of the reform on students' application priority, I focus on changes in students' preferences. I classify degrees into fields and estimate a multinomial logit model to calculate the probability of choosing and enrolling in a particular field on the condition of being positively affected by the reform. I control for test scores and high school performance to separate the effects of academic ability from the impact of the policy change. I pay special attention to the role that the reform has played in applications to degrees in the STEM fields. Majors in this field are usually more in demand than others, and thus their cut-off scores are higher. Thus, benefiting from the policy change increases the probability of being admitted to a degree program in this field.

I find glaring gender differences in the expected probabilities of applying to each field. Men, for instance, are considerably more likely to apply to engineering than to any other field. In contrast, women are substantially more likely to apply to a degree in the healthcare field unless they are at the very top of the mathematics achievement distribution. More importantly, I find that the policy change has affected students' choices. In general, men and women who benefit from the ranking bonus are less likely to apply and enroll in programs in the lower-paid fields. Instead, this group of applicants have an increased probability of choosing degrees in medicine and engineering. The effects are persistent across different levels of achievement in the admission tests; however, I find some heterogeneity in students' responses to the reform across different social backgrounds. For instance, for girls from private schools, the ranking bonus leads to an increase in the probability of applying to degree programs in engineering, whereas for girls from public schools it increases the likelihood of applying to degree programs in the medical field. It is outside the scope of this paper to explain the reasons behind the heterogeneity. However, I present statistics supporting the hypothesis that differences in salary expectations across social classes might drive the results.

The literature on the determinants of sex segregation in higher education has focused on various issues. Some of them are students' preferences (see for example, Zafar (2013), Wiswall and Zafar (2015), and Wiswall and Zafar (2018)); differences in mathematic achievement and academic preparedness (Turner and Bowen (1999), Card and Payne (2021)); gender differences in competitiveness (Buser et al. (2014), Buser et al. (2017)), and student's beliefs about their relative academic ability (Correll (2001), Rapoport and Thibout (2018), and Justman and Méndez (2018)).

In this paper, I link the evidence on the gender gap in performance in different types of tasks with its effect on students' fields of choice. Furthermore, my research adds a structural lens to previous related work by evaluating the impact of an admission system reform. Whilst some research has been carried out on the role of structural factors in field choice (see for example Papay et al. (2016) for the impact of test scores "labels" and Patterson et al. (2021) for the role of timing of course taking),[2] it has not extended to the gendered nature of admission systems and their relationship with subsequent disparities in fields of study.

Universities and governments across the world use aptitude and achievement tests to select applicants into degree programs (Edwards et al., 2012). In some countries, the tests constitute the sole criteria for selection, but more generally, as in Chile, the institutions employ the tests in conjunction with other measures of achievement. The results of this study are particularly significant for these institutions, as they provide empirical evidence on how to correct gender biases in admission systems that rely heavily on standardized test scores.

The rest of this paper is organized as follows. In Section 2.2, I provide some background on the Chilean admission system and describe the policy change. Section 2.3 discusses the empirical strategies. Section 2.4 describes the dataset. Section 2.5 reports the results on the targeting of the reform and the effects on student choices. Finally, Section 2.6 summarizes and discusses the results of the study.

2.2 Background

In many countries, university admission tests are conducted nationally, and students undertake them almost systematically. Brazil, China, Greece, Japan, Portugal, South Korea, South Africa, Spain, and Turkey are examples of countries that

[2] Papay et al. (2016) analyses the effects of receving a more positive "label" describing students scores on students decisions to attend tertiary education. Patterson et al. (2021) evaluates how the order in which students are assigned courses affects the choice of college major, they find that students are more likely to choose a major related to the courses they are taking at the time of the major selection process

employ general admission tests either as a sole or a supplementary determinant of entry (Edwards et al., 2012). Chile uses a combination of a national admission test and high school level achievement measures to determine entry to university. A public agency, the Department of Educational Evaluation, Measurement and Registration (known in Spanish for its acronym DEMRE) oversees the administration of the national admission test, the PSU (*Prueba de Selección Universitaria*), and the process of applications and admissions for most universities in Chile. 39 out of the 61 existing universities in Chile are part of this centralized admission system. It is important to note that the higher education institutions that participate in the centralized system are also the most selective and prestigious; every Chilean university listed in the QS-World-Ranking (2019), handles its admissions through DEMRE. Therefore, taking the tests and participating in the centralized admission process is highly desirable for high school graduates that wish to pursue a university degree.

The set of admission criteria used by the centralized admission system includes PSU scores, the high school grade point average, and since the 2013 reform, the high-school ranking. The PSU consists of four exams: language, mathematics, history and science. Every student must take the language and the mathematics test, but students can choose to take either the history or science test, or both. Students decide on which test to take depending on the degree programs they wish to apply for. The other two criteria are measures of achievement during high school; the first one is the NEM score (*Notas de Enseñanza Media*), which is calculated using the student's cumulative grade point average, denominated, and the second one is the ranking score (Ranking de Notas), which is a function of the student's relative position in their high school cohort (see section 2.3 for a detailed explanation of the raking score computation).

The admission process works as follows:

1. Students take the PSU test in a simultaneous process that happens once a year.
2. DEMRE processes the tests and sends students their results.

3. Students list up to ten degree programmes from their most preferred to least preferred ones and submit their list to DEMRE.

4. DEMRE computes the application score for each student for each degree programme in their list

5. DEMRE allocates applicants to degree programmes considering (1) students' application scores, (2) students' preferences over degree programmes, (3) degree programmes requirements and total vacancies.

6. Students are admitted to their most preferred degree for which they achieved a sufficiently high score.

Students declare their degree program preferences by submitting a list with their choices ranked in strict order of preference to the DEMRE online platform. As in many other countries' postsecondary education systems (McGrath and Frearson, 2016), a choice indicates both an institution and a major. For instance, a choice could be Engineering at the University of Chile. Because courses have limited vacancies, it is harder for applicants to gain admission to popular fields. Universities do not express preferences over applicants. However, they choose the weight of each admission criterion for each of their degree programs. Universities also choose which optional test (history or science) they require for each undergraduate degree program they offer. The weights of the admission criteria for each degree program at each university are public information. DEMRE calculates students' application scores using these weights (DEMRE, 2018).

2.2.1 Women in STEM in Chile

OECD countries, including Chile, have successfully managed to close the gender gaps in access to education. Moreover, nowadays younger women are more likely than younger men to achieve tertiary education (OECD, 2020a). However, the strong gender segregation in fields of study in higher education prevails. The figures for female participation in STEM degrees place Chile in the lower rankings among the OECD countries. For example, Chile is in the last position of the OECD rankings for sex distribution of STEM graduates: only 19% of STEM graduates in the country are women, compared to countries such as the UK that have a female participation in stem degrees of 35% (STEM Women, 2021).

Jobs in the STEM fields are also among the best remunerated (Deming and Noray, 2018). Therefore, the lack of women in these careers deepens the wage gap.[3] The OECD (2018), education at a glance report reveals that Chile has the highest wage gap between men and women with higher education. The study indicates that in 2015 women with higher education earned 65% of what men with the same educational level obtained.

Gender gaps in performance on standardized tests might be at the root of the disparities in fields of study as they begin during secondary education and persist into university admissions stage (Bobbitt-Zeher, 2007). Statistics for Chile show that in the last decade, while girls and boys have had similar levels of achievement in mathematics at the age of nine; at thirteen years old, girls achievement level has dropped below that of men. The gap increases in the last school measurement at the age of fifteen and persists in the mathematics national standardized test for admission to higher education (Arias et al., 2016).

In the Chilean case, scores in the mathematics admission test are particularly influential in determining the chances of being accepted into a STEM degree program. Therefore, the gender gap in test scores potentially becomes a determinant of the low participation of women in STEM.

More importantly, the evidence on the effect of competitive tests on gender gaps indicates that the gaps observed may not necessarily reflect a real difference in the math skills of men and women (see, for instance Niederle and Vesterlund (2010)). Therefore, relying heavily on test scores to make admission decisions could leave out women who are well prepared for the academic degrees they wish to attend.

[3]See Joensen and Nielsen (2016) for evidence that opting for advanced mathematics classes has sizable earning returns for girls. See also Blau and Kahn (2017) review on explanations for the gender wage gap.

2.2.2 Description of the reform

In 2013, Chilean university representatives implemented an educational reform that added a new university admission criterion, the ranking score, to the existing ones: national admission exam scores and the GPA score. The ranking score expresses the student's relative position among his/her peers during secondary education, using the performance of the students in the last three cohorts who graduated from that school as a reference. Thus, the class ranking criterion is a measure of continuous assessment. Therefore, by introducing it as a mandatory admission criterion, the reform lowered the weight of high-stake examinations (PSUs) and raised the importance of school performance (grades and ranking) earned over more extended periods in lower-stake settings.

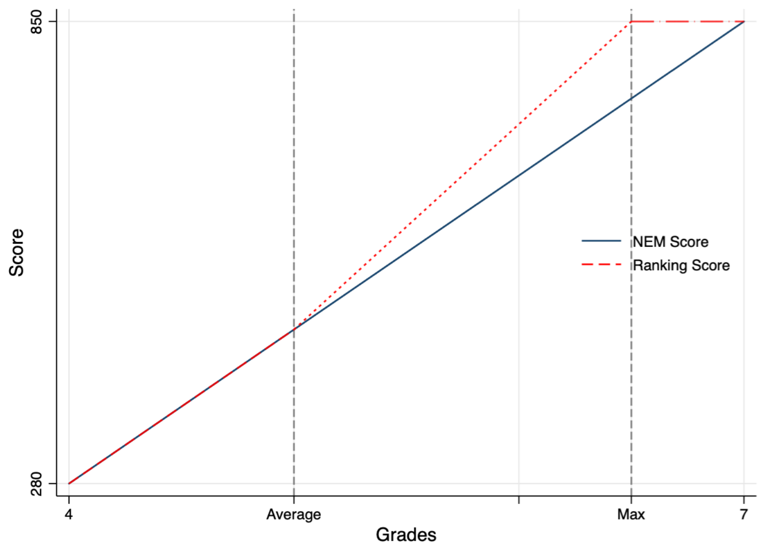
The new criterion recognizes the effort of students who perform well in their educational context, and it was introduced with the purpose of improving equity in access to university (DEMRE, 2019). The educational authorities decided to include the ranking score in the middle of the school year of 2012. Thus, the first generation of high school graduates affected by the reform only learnt about it a few months before graduation. As the ranking score considers students' performance in the four years of secondary education, the 2012 cohort of high-school graduates had little to no room to change their behavior in response to the reform.

DEMRE calculates each student's High School Ranking score using information particular to the student and two statistics from the educational institution that the applicant attended. The individual information corresponds to the cumulative grade average that the student obtained during their last four years of secondary education (GPA). The school information compromises two statistics, which DEMRE computes for each educational institution: the historical average, that is, the mean of the GAP of all of the students who graduated in the last three cohorts, and the historical maximum, which corresponds to the average of the maximum grade of each one of the three previous cohorts.

Students with a GPA above the historical average of their school get a higher ranking score than their NEM score. On the other hand, for the students who get GPAs equal to or lower than the historical average of the school, the ranking score equals the NEM score. In figure 1, this group of students is between the minimum GPA (4) and the average line. The formula to calculate the score is linear. The maximum score is 850 points, and the minimum is equal to the NEM score at the school average; thus, the minimum ranking score varies from one school to another. Students get to know their ranking score with a relatively high level of accuracy before taking the PSU tests.

Figure 1 illustrates the formula for a hypothetical school. The dotted line shows the ranking score. The ranking score for students with a GPA between the average line and the max line is a linear combination of the NEM score corresponding to the historical average of their school and 850. Students with a final grade above the maximum historical average of the school obtain 850 points.

Figure 2.1: Example of Ranking Score vs NEM Score



Notes: Example for an hypothetical school with an average class GPA between 4 and 7 (Average) and a maximum score (Max) below 7.

For students that graduate with a GPA above the historical average of their school, the ranking score can be interpreted as a bonus over their NEM score. I define the ranking bonus variable as:

- 1 if the individual's grade point average is in the region where the dotted line is above the solid line (Ranking score > NEM Score)
- 0 if not.

Two features of the policy change are particularly important for my analysis. (i) *Staggered timing*: DEMRE computes the ranking score using 4 years of individual information on grades so, when the reform was announced, students had already made their decisions on studying efforts. Therefore, the changes in admission priority are unexpected. (ii) *Variance across schools*: The minimum final GPA at which a student gets the ranking bonus, i.e. the historical average, varies across schools.

2.3 Empirical strategy

2.3.1 Applications scores gap

The ranking criterion is based on performance in continuous evaluations over a prolonged period of time. The stakes on each one of these evaluations are considerably lower than the stakes at the PSU test. Therefore, by adding the ranking criterion to the existent criteria the reform shifts some weight from high-stake measures of achievement to lower-stakes measure of achievement. Prior studies (Cai et al. (2019), Schlosser et al. (2019), Iriberri and Rey-Biel (2019), Montolio and Taberner (2021)) demonstrate that men tend to perform better in tasks where the stakes are higher; these results suggest that the policy change has the potential to improve women's application scores relative to men. In this section, I examine whether the above hypothesis is correct by analyzing differential trends in women's and men's application scores before and after the reform.

I first estimate the following linear regression to identify gender differences in the application scores before and after the reform:

$$Y_{j,t} = \beta_0 + \beta_1 R_t + \beta_2 G_j + \beta_3 R_t * G_j + \delta_t + \epsilon_{j,t} \tag{2.1}$$

Where $Y_{j,t}$ is the application score of student j (in time t), R_t is a dummy equal to 1 if the year $t \geq 2013$, G_j is a dummy equal to 1 if student j is a woman and 0 if not, and δ_t is a year fixed effect.

The interaction between the dummy variables for gender and the reform shows whether the application scores of women compared to men changed before and after the reform.

After identifying general gender trends, I move towards identifying the ex-post beneficiaries of the policy by simulating applicants' counterfactual scores. To analyze if a student is an ex-post beneficiary, I take her final application score for the degree she applied for as a first choice and compare it with the final score she would have received if she had applied the year before the reform.

To calculate the counterfactual case, I simulate the composite scores of applicants if the policy change had not been applied. I compute the counterfactual application score using students' real NEM score and PSU tests scores, but I modify the weights defined for each degree program to match the weights of the criteria from the year before the reform.

The example presented in Table 2.1 illustrates the calculus of the counterfactual case for a given student. The student in the example took the science test and got a ranking score higher than their NEM score. The degree program the student applied for decreased the weights of the mathematics PSU score and the NEM score in order to include the mandatory new criterion. Both the real and simulated scores are calculated as the weighted sum of the PSU scores, NEM score and Ranking score. While the student's real score (including the new criterion) is 710, the simulated one (if the weights had not changed to include the new criterion) is 702.5. In this example, the student has benefited from the policy because their real score is higher than their simulated score for the specific degree program presented in the example.

Table 2.1: Example calculation of real versus simulated composite score for a scientific degree

Criterion	Student score	2013 Weights	2012 Weights
Lenguage PSU Score	700	15%	15%
Mathematics PSU Score	750	40%	45%
Sciences PSU Score	600	20%	20%
History PSU Score	-		
NEM Score	700	15%	20%
Ranking Score	800	10%	0%
Composite Score		Real: 710	Simulated: 702.5

Notes: Student scores for hypothetical student participating in the 2013 admission process. 2013 weights are the current admission weights of a particular program, and 2012 weights are the pre-reform weights of said program.

2.3.2 Field Choice

Prior literature suggests that an improvement in a student's scores can affect their preferences. Avery et al. (2018), for instance, use data from the US and show that signals of high aptitude, specifically high scores, affect students' college major choices. These findings suggest that if the Chilean reform affects students' application scores, it could also affect students' preferences and choice of field of study.

In this section, I exploit the sudden inclusion of the ranking bonus to identify the effects of the reform on students' applications and enrolment in fields of study. I treat the introduction of the ranking score as an exogenous shock to the likelihood of gaining admission to more selective degree programs to evaluate whether an increase in application scores conduces female students to apply to degrees in male dominated fields.

When students apply to a university degree-program they are confronted with a vast number of options. To estimate the effect of the ranking criterion on students' choices, I use a discrete choice model where the independent variable is the field of the degree to which the student applied as a first choice, and the variable of interest is a dummy for the ranking bonus.

I assume that students choose a program by comparing the utility levels among the options that are available to them. Hence, I follow Fuller et al. (1982) and use a Multinomial Logit (MNL) approach to model the student's application decision. In particular, student s choosing major i enjoys the following utility:

$$u_{si} = W_s'\gamma_i + \varepsilon_{si} \tag{2.2}$$

- W_s' are the alternative invariant regressors including: Student s's school of origin, family income decile, PSU scores, NEM score and the ranking bonus dummy.
- ε_{si} is a random component that contains all the unobserved determinants of utility.

Individuals choose the alternative i that maximizes their utility; however, not all the options are available for each applicant. C_s denotes student s's choice set, which varies among applicants depending on the test they decide to take. For instance, if a student wants to pursue a scientific degree, there is no need to take the history test. A student deciding not to take the history test cannot apply for any art or humanities-related degrees. Figure 2.2 illustrates the decision process that applicants confront.

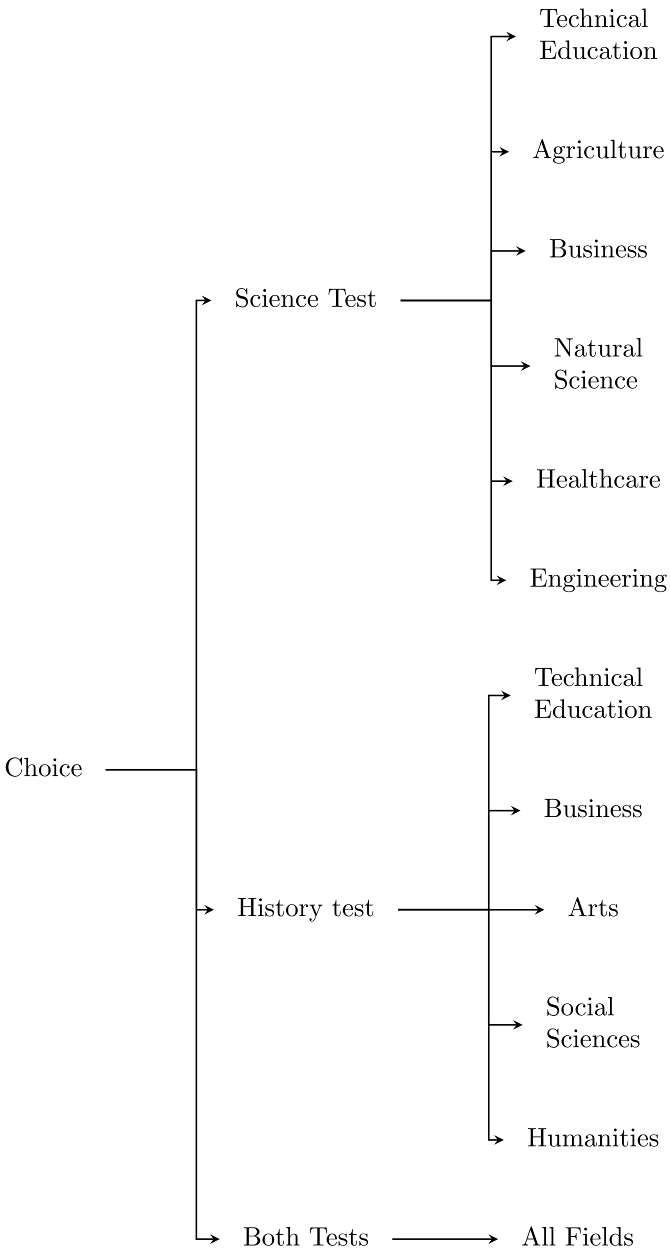


Figure 2.2: Applicants' decision process

Following Ben-Akiva and Lerman (1974) and Adler and Ben-Akiva (1976), I use a two-stage model as a deterministic rule to decide the availability (or unavailability) of a particular alternative. At the first stage, applicants choose which test to take, and at the second stage, they choose a field conditional on the test they took. Accordingly, to estimate the multinomial logit for the second stage, I separate the students into three subsamples: students who took the science test, students who took the history test and students who took both tests. However, as I am interested in the effect of the ranking criterion on STEM degrees, I only report the results for the two subsamples of students that can apply to degree programs in that field, i.e., students who took the science test and students who took both tests.

The set of options that an applicant faces C_s varies across the subsamples but is the same within each subsample. Following from equation 2.2, a student s that took the test t has a choice set C_t and will choose the alternative i if:

$$W'_{sti}\gamma_{ti} + \varepsilon_{sti} \geq W'_{stj}\gamma_{tj} + \varepsilon_{stj}, \forall j \in C_t \tag{2.3}$$

Because it is difficult to interpret the estimated coefficients of the multinomial logit models directly from their estimated coefficients, in the results section I report the fitted probabilities for the various levels of the response of the test choice and the field choice models. To calculate the fitted probabilities, I keep the ranking variable as a focal predictor, while the other predictors in the model are held to the subsample average values unless specified. For categorical variables like income band, I weight the levels of the non-focal factor by sample size. I also report the average marginal effect of benefiting from the ranking bonus on the predicted probabilities of choosing each field. I repeat this process for different levels of mathematical ability (measured by the PSU test) and for different schools of origin.

2.3.3 Identification

I restrict the analyses described in sections 2.3.1 and 2.3.2 to the first year of the implementation of the policy change to avoid endogeneity problems that strategic behaviour among students might cause. I assume that during the first year, the policy change did not provoke changes in students' choices through different mechanisms other than the change in their application scores. This assumption is based on the fact that DEMRE implemented the ranking in the same year that they announced it, which did not allow time for applicants to modify their behaviour.

Still, this estimation approach does not permit to fully disentangle the effect of being a student at the top of their class from the pure impact of the reform. In the presence of these confounding factors, an experimental setting could allow the complete separation of these two effects. In theory, the Chilean reform could be modelled as a quasi-experiment, where top-ranked students are treated by receiving a "ranking bonus". In such a case, the causal effects of the reform can be estimated using a difference-in-difference (DD) strategy. This approach has the advantage of removing biases in post-intervention period comparisons between the treatment and control group that could result from permanent differences between those groups, such as intrinsic differences between students at the top and bottom of the GPA distribution of their schools (Meyer, 1995). However, the existing data does not identify which students graduated at the top of their high school classes in pre-reform years.

Nonetheless, to alleviate concerns about potential confounding factors, I include a "quasi-experimental" analysis where I generate a proxy-treatment and a proxy-control group and use a DD strategy to estimate the causal effect of the reform on applications and attendance to engineering degree programmes.

Under an experimental framework, the treated group corresponds to the students whose GPA was above the historical average of their school. Thus, after the implementation of the reform, the treated and control group is easily identified by their ranking bonus status. Students who benefited from the ranking bonus are part of the treatment group, and the others are part of the control group. For the pre-reform cohorts, identifying top-ranked students in each school requires individual students' GPA, which is available, and, in addition, data on the average GPA of the student's high school cohort (or sufficient data to compute this average). The latter is not available; therefore, it is impossible to precisely identify the treatment group for the pre-reform period. Instead, I propose a proxy measure that relies on the existing school-level data and individual GPAs to determine the pre-reform treatment and control groups.

The available school-level data includes the type of school (private, semi-private and public) and location of the school (council level). This data allows me to construct a GPA average by type of school, council and cohort. Thus, for each cohort and council, I calculate GPA averages for three groups of students: students graduating from public schools, semi-private schools, and private schools. Then, I allocate students with GPAs of at least 0.5 standard deviations above the mean of their group to the treatment group and students with GPAs that are at least 0.5 standard deviations below to the control group. It is likely that students who are just above (or just below) the GPA average of their council/school-type group might not have been above (or below) the GPA average of their school. Dropping the observations close to the mean helps me to account for the variation in actual historical GPA of schools within a council/school-type group.

Given that the real treated group is observable after the reform, it is possible to use the post-treatment data to compare the proxy treated group with the real treated group and assess how similar they are. This assessment provides an indication of the fitness of the proxy measure. I use the 2013 data to test the proposed proxy measure and present the results in Table 2.2. Each column presents the percentage of students who were accurately predicted as receiving the bonus and not receiving it. For the 2013 cohort, the proxy measure correctly predicts 85% of students in the control group and 91% of the students in the treatment group. Although the proxy measure performs well in identifying the beneficiaries of the reform, it still has the disadvantage that it eliminates around 35% of the total number of observations (all observations that are half a standard deviation around the mean of each council/school-type group). Therefore, a limitation of this strategy is that it only allows the estimation of the effects of the reform for a selected sample of the treated students.

Table 2.2: Proxy-treatment fitness test

Proxy for ranking bonus (%)	Ranking Bonus (%)		
	0 (Control)	1 (Treatment)	Total
0 (Proxy-control)	85.37	14.63	100.00
1 (Proxy-treatment)	8.57	91.43	100.00
Total	39.77	60.23	100.00

Note: Estimation for the 2013 cohort (first post reform cohort).

Difference in differences and Triple difference in differences

To estimate the effect of the reform, I compare applications and enrollment to degree-programmes in engineering fields before and after the reform for beneficiaries and non-beneficiaries of the ranking bonus. For the pre-reform years, I use the proxy measure just described to identify the treated group. For post-reform period, I look at the outcomes of the students that were benefited by the ranking bonus against the ones who were not. Accordingly, I estimate the following difference-in-differences equation:

$$Y_{j,t} = \beta_0 + \beta_1 R_t + \beta_2 T_j + \beta_3 R_t * T_j + \delta_t + \epsilon_{j,t} \tag{2.4}$$

Where $Y_{j,t}$ is a dummy equal to 1 if student j applied (or enroled) to a degree programme in the engineering field (in time t), R_t is a dummy equal to 1 if the year $t \geq 2013$, T_j is a dummy equal to 1 if student j received the ranking bonus (after reform), or would have received according to the proxy measure (before reform), and δ_t is a year fixed effect.

I first estimate equation 2.4 separately for men and women, and then I also add the comparison between men and women to learn about the potential gendered effects of the reform. To add the further gender comparison, I use what is known in the literature as a triple difference strategy. This approach was first introduced by Gruber (1994) and has been used to study the effects of a variety of policies.[4] Equation 2.5 shows the specification of the triple differences estimation. Where G_j is a dummy equal to 1 if the student j is female.

$$Y_{j,t} = \beta_0 + \beta_1 R_t + \beta_2 T_j + \beta_3 G_j + \beta_4 R_t * T_j * G_j + \delta_t + \epsilon_{j,t} \tag{2.5}$$

2.4 Data

My data are drawn from the Chilean Centralized Admission System. The dataset contains individual-level data on all of the students who took part in the annual round of admissions. The data cover the period from 2008 to 2018. These are data about demographic and socioeconomic characteristics, academic performance, including high school records and admission test scores, and information on each student's applications and enrolment. Students from low socioeconomic backgrounds are slightly less likely to take the test than individuals from privileged backgrounds. Moreover, they are also less likely to apply to study a university degree. Hence, high achievers and individuals from medium and high socioeconomic backgrounds are overrepresented in the sample.

In 2013, the total number of applicants was 118212, from which 52.2% were women. 76976 enrolled in one of the degree-programmes offered by the centralised admis-

[4]Yelowitz (1995) for instance, uses this approach to estimate the effect of public health insurance on mother's labour market decisions, separating the effects by age of their children. Hamermesh and Trejo (2000) follow this approach, in the context of hours worked by gender, and Kellogg and Wolff (2008) for energy consumption by time of the day.

sion system, and 48.8% of the enrolled students were women.

Regarding student's background, 23.9% of the students who enrolled in university attended a private school, 51.9% attended a semi-private school and 24.3% came from a public institution. Type of school of origin is highly correlated to income. Table 2.14 (in the appendix) shows the family income bands, the percentage of applicants in each band, and the type of school of origin of students in each band. Only 4% of the students in the lower family income band came from private schools, while 78% of the students in the highest income band attended private education.

Table 2.3 presents statistics on test choice, average scores, applications, and enrolment for the 2013 cohort. The PSU consists of four tests; the minimum score for each test is 150 points and the maximum is 850. The top panel in Table 2.3 shows the proportion of men and women who took the Science, History, and both tests. The bottom panel shows the mean and the standard deviation of the PSU scores by gender. On average men performed better than women.

Table 2.3: Summary statistics: Admission process

		Men	Women
Test taken (Sample)			
Science	N	22,364	24,058
	%	48.2	51.8
	Total (N)	46,422	
History	N	14,894	18,393
	%	44.8	55.2
	Total (N)	33,287	
Both	N	12,121	12,490
	%	49.3	50.7
	Total (N)	24,611	
Admission scores			
Language	Mean	503.9	496.9
	s.d.	(111.3)	(106.2)
Math	Mean	514.7	487.3
	s.d.	(112.6)	(104.7)
History	Mean	513.3	488.7
	s.d.	(110.8)	(105.0)
Science	Mean	517.0	486.9
	s.d.	(109.3)	(105.8)
NEM score	Mean	570.3	592.3
	s.d.	(100.6)	(95.6)
Ranking score	Mean	590.7	617.0
	s.d.	(120.9)	(117.3)
Application Score	Mean	590.3	582.32
	s.d.	(129.75)	(136.25)
Applications and enrolment			
Applications	N	56,449	61,763
	%	47.75	52.25
Enrolment	N	38,918	37,058
	%	51.22	48.78

Notes: Mean and standard deviation (s.d.) of students who participated in the 2013 centralised admission process.

There were 1375 degree-programs offered in the 2013 centralized admission process. I classify them into ten fields of study: agriculture, arts, business, engineering, humanities, healthcare, natural sciences, social sciences, technical and other. Table

2.4 shows the average wage, the percentage of women and the percentage of students from each type of school by field. The best paid fields are engineering and medicine. Table 2.4 also shows the percentage of students who applied to a degree in each field, broken down by gender. The fields with more women are medicine, social sciences, arts, and humanities, while the fields with less women enrolling in 2013 are technical and engineering.

Table 2.4: Summary statistics: Field of study

Field	Average wage	Men	Women	Total	School of origin		
	Thousands of CLP ($)	(%)	(%)	(%)	Public	Semi-private	Private
Agriculture	731.44	3	3	3	0.23	0.50	0.27
Arts	709.70	5	7	6	0.19	0.45	0.35
Business	1066.68	12	9	10	0.20	0.41	0.38
Engineering	1223.06	36	12	24	0.24	0.55	0.21
Humanities	727.86	16	22	19	0.27	0.52	0.20
Healthcare	1084.58	14	30	22	0.24	0.54	0.22
Natural Sciences	1112.38	4	3	3	0.24	0.52	0.23
Social Sciences	707.95	8	12	10	0.23	0.50	0.26
Technical	858.29	1	1	1	0.32	0.63	0.05
Other	-	1	1	1			
Total	985.34	100%	100%	100%	0.24	0.52	0.24
N Observations	-	38,918	37,058	75,976	-	-	-

Notes: Data on average wage from MIFUTURO, a public database on labour market outcomes. The Data is generated by the Chilean Ministry of Education.

2.5 Results

2.5.1 Application scores

The first set of results demonstrate that the reform contributed to reducing the existing gender gap in application scores. Table 2.5 shows the OLS regression estimates of equation 2.1. The estimate for the gender dummy shows that women have statistically significant lower application scores than men. In contrast, the estimate for the interaction of the gender dummy and the reform dummy shows that the reform increased women's application scores relative to men by around 5 points.

Table 2.5: Differential trends in application scores between men and women before and after the reform

	(1) Application Score
After reform	0.904*
	(2.06)
Women	-3.643***
	(-12.00)
After reform and women	4.585***
	(11.58)
Year	0.852***
	(8.56)
Constant	-1102.7***
	(-5.51)
N	522341

Notes: N includes observations from 2009 to 2017, t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6 displays the percentage of students whose actual score was greater than the simulated score for the degree that they applied for as their first option. A higher percentage of women are part of the subset of students with higher composite scores because of the reform (78% of women versus 67% of men). When we look at women who applied to engineering degrees as their first option, we can see that 85% of them had a higher real composite score than their counterfactual score; meanwhile 73% of men were in the same situation. Even when for most of the students their composite scores were higher when using the ranking score as a factor, women benefited more broadly. Therefore, more women than men moved towards the top of the applicant distribution.

Table 2.6: Percentage of applicants with a higher real composite score than the simulated score

	All Degrees	Engineering Degrees
All	73%	76%
Men	67%	73%
Women	78%	85%

Note: Percentages calculated for the 2013 cohort

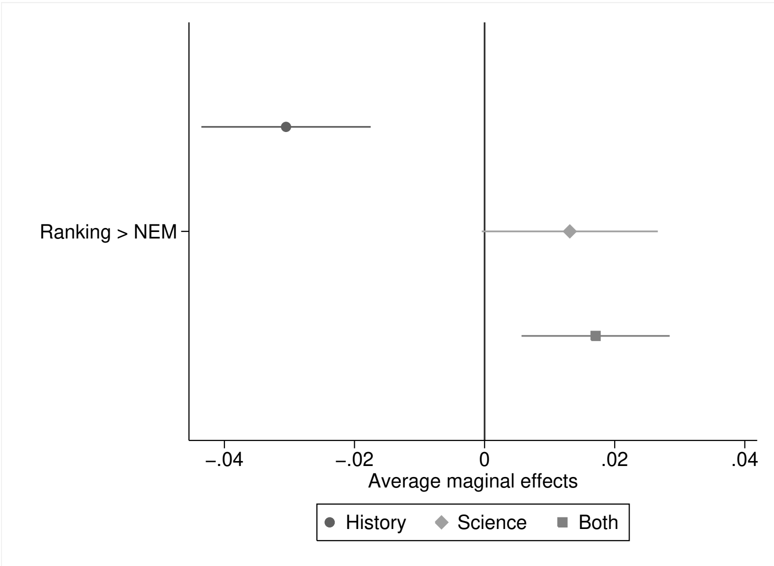
2.5.2 Test choice by gender

Before studying the effect of the reform on field choice, I estimate an MNL for test choice and report the predicted probabilities to take each test. The alternatives that a student faces are science or history or taking both tests.

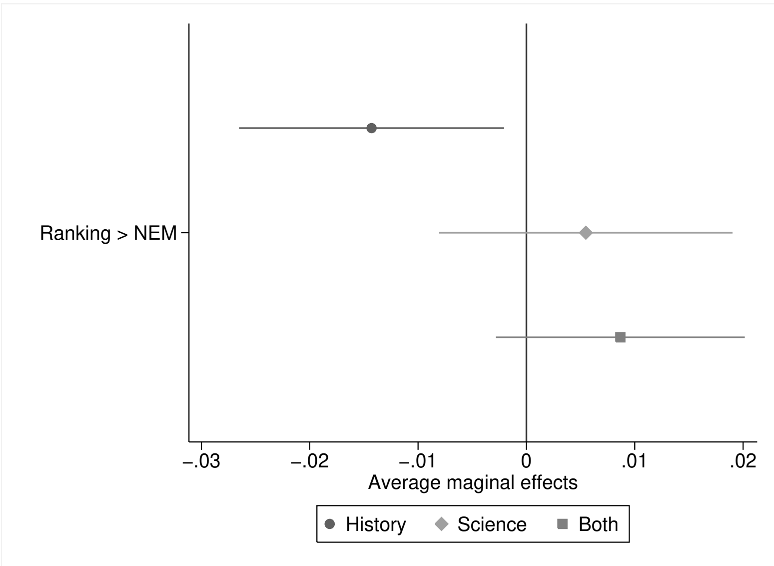
Table 2.15 displays the exponentiated coefficients of the regression with history as the reference outcome choice. Table 2.7 shows the expected probabilities of choosing to take the history, science or both tests for women and men, and the confidence bands for the fitted logits and probabilities. Overall, these results indicate that the reform has had bigger effects for women. The"average boy" is more likely to take the science test than the "average girl" (0.435 vs 0.459 when the ranking dummy is equal to 0 and 0.447 vs 0.459 when the ranking dummy is equal to 1). Men are also more likely to take both tests than women and less likely to take the history test, independently of the value of the ranking variable. Figure 2.3 presents the average marginal effect of the ranking bonus on test choice, separately for women and men. The figure also displays the 95% confidence interval for each of the three options (History, Science or Both Tests). The top sub-figure shows the results for women. For women in the beneficiaries' group, the average marginal effect of the ranking bonus on taking the science test is 2.3% and on taking both tests is 2.2%. The effects are significant at the 5% level. The bottom sub-figure shows the results for men. Men in the beneficiaries' group are also more likely to take both tests. The average marginal effect for that option is 1.7%, and it is statistically significant at the 5% level. The effect of the ranking bonus on taking the science test is not significant for men.

Figure 2.3: Average marginal effect of the ranking bonus on test choice

Women



Men



Notes: Ranking>NEM indicates benefitting from the reform. 95% confidence intervals.

Table 2.7: Predicted probabilities (PP) of test choice by gender

Test		(1) Women	(2) Men
History	$Ranking < NEM$	0.333***	0.257***
		(0.005)	(0.004)
	$Ranking > NEM$	0.300***	0.242***
		(0.004)	(0.004)
Science	$Ranking < NEM$	0.393***	0.437***
		(0.005)	(0.005)
	$Ranking > NEM$	0.407***	0.444***
		(0.004)	(0.004)
Both	$Ranking < NEM$	0.219***	0.240***
		(0.004)	(0.004)
	$Ranking > NEM$	0.236***	0.248***
		(0.003)	(0.004)
Observations		35,261	36,627

Notes: Column (1) and (2) present predicted probabilities for women and men. All predictors at their mean value, standard errors in parentheses.

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

2.5.3 Field choice by gender

Turning now to the second stage of students' decision-making process, I analyze how the ranking bonus affects field choice. I use the multinomial logit approach to estimate the probability of applying to each field as a first choice. Given that students can list up to ten preferences in their application, they can afford to list very selective degrees as their first choice even if there is a small probability of being admitted. However, there is growing evidence in the literature that students make their decision on their course of study based on their beliefs about their relative academic abilities. For instance, Correll (2001) shows that gender differences in perceptions of mathematical competence partly explain the gender gap in participation on STEM careers as perceptions of ability influence high school and college students' educational decisions. Rapoport and Thibout (2018) find that girls and boys performing equally in the same subjects choose majors differently because girls under-estimate their science skills when choosing their educational and career paths. In a similar line, Justman and Méndez (2018) find that female students require stronger prior signals of mathematical ability to pursue male-dominated subjects, such as STEM. The literature has demonstrated that students' scores in standardized tests not only determine the probability of being admitted to certain degrees but also act as ability signals for students, affecting their choices. I argue that the ranking score also provides students with information on their academic ability. In this section, I present the empirical results regarding the effects of the new signal.

Table 2.8 displays the predicted probabilities of choosing a degree in each of the following categories: engineering, natural sciences, health and other. The values presented in the table are based on the results of the multinomial logit for field choice. Columns (1) and (2) display the results for women and men in the subsample of students who took the science test only and columns (3) and (4) display the results for the subsample of students who took both tests (science and history).

There are marked gender differences in the expected probabilities of applying to each field in both subsamples. While men are considerably more likely to apply to engineering than to any other field, women prefer healthcare degrees. The differences are more striking in the science subsample. For the average girl, the expected probability of choosing an engineering degree is 20% when she is not benefited by the ranking bonus and 22% when she is. Meanwhile, the expected probability of the average boy choosing a degree in the engineering field is 63% or 64% depending on the ranking bonus.

Table 2.8: Predicted probabilities (PP) of field choice by gender

	Science test		Both tests	
	(1)	(2)	(3)	(4)
Choice	PP Women	PP Men	PP Women	PP Men
Other				
Ranking $< NEM$	0.137***	0.130***	0.544***	0.463***
	(0.007)	(0.006)	(0.013)	(0.011)
Ranking $> NEM$	0.118***	0.109***	0.504***	0.384***
	(0.004)	(0.004)	(0.008)	(0.009)
Engineering				
Ranking $< NEM$	0.192***	0.618***	0.141***	0.417***
	(0.008)	(0.008)	(0.009)	(0.011)
Ranking $> NEM$	0.206***	0.647***	0.166***	0.488***
	(0.005)	(0.006)	(0.006)	(0.009)
Medicine				
Ranking $< NEM$	0.616***	0.182***	0.293***	0.089***
	(0.010)	(0.006)	(0.012)	(0.006)
Ranking $> NEM$	0.626***	0.185***	0.300***	0.097***
	(0.006)	(0.005)	(0.007)	(0.005)
Natural Sciences				
Ranking $< NEM$	0.055***	0.071***	0.022***	0.032***
	(0.005)	(0.004)	(0.003)	(0.004)
Ranking $> NEM$	0.051***	0.059***	0.030***	0.031***
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	12,887	14,952	8,008	8,857

Notes: All predictors at their mean value. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The effect of benefiting from the new ranking score: AVME by mathematics PSU score

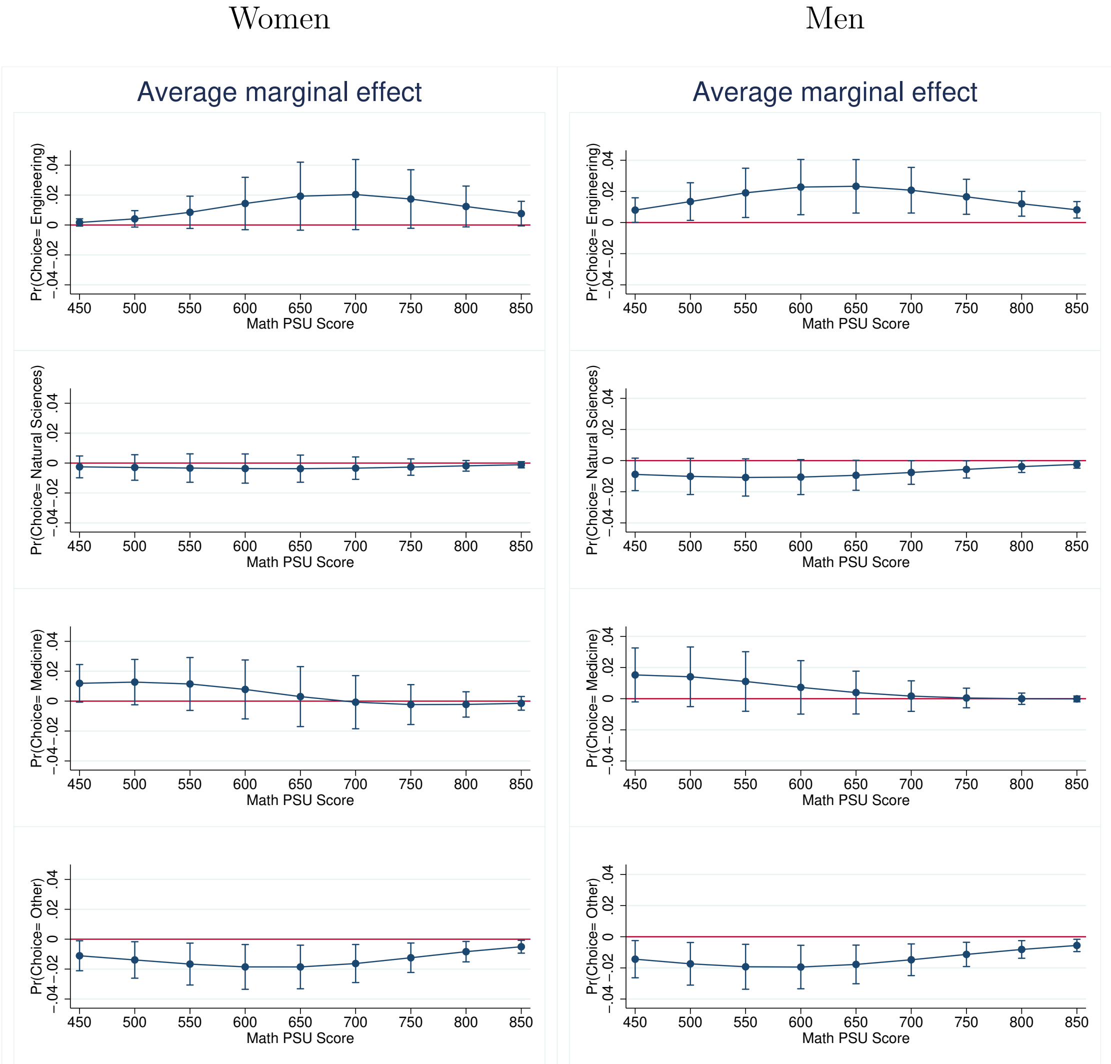
Math scores are strong predictors of choice.[5] Thereby, I investigate potential heterogeneities in the effects of the reform by estimating the marginal effects of the ranking dummy on choice probabilities by various levels of achievement in mathematics.

[5]See odd-ratios of math scores on field choice in tables 2.16 and 2.17 in the appendix.

Figures 2.4 and 2.5 show the average marginal effect at the mean (MEM) of being benefitted by the ranking bonus on field choice, the results are disaggregated by different levels of attainment in the mathematics PSU test. Figure 2.4 shows the results for students who took the science test only and figure 2.5 shows the results for students who took both tests. The left-hand column presents the average marginal effects for women. Each of the sub-figures contains the result for a different option (engineering, natural sciences, medicine or other) The MEM is on the y-axis and the mathematics test score is on the x-axis. The right-hand column displays the same results for men.

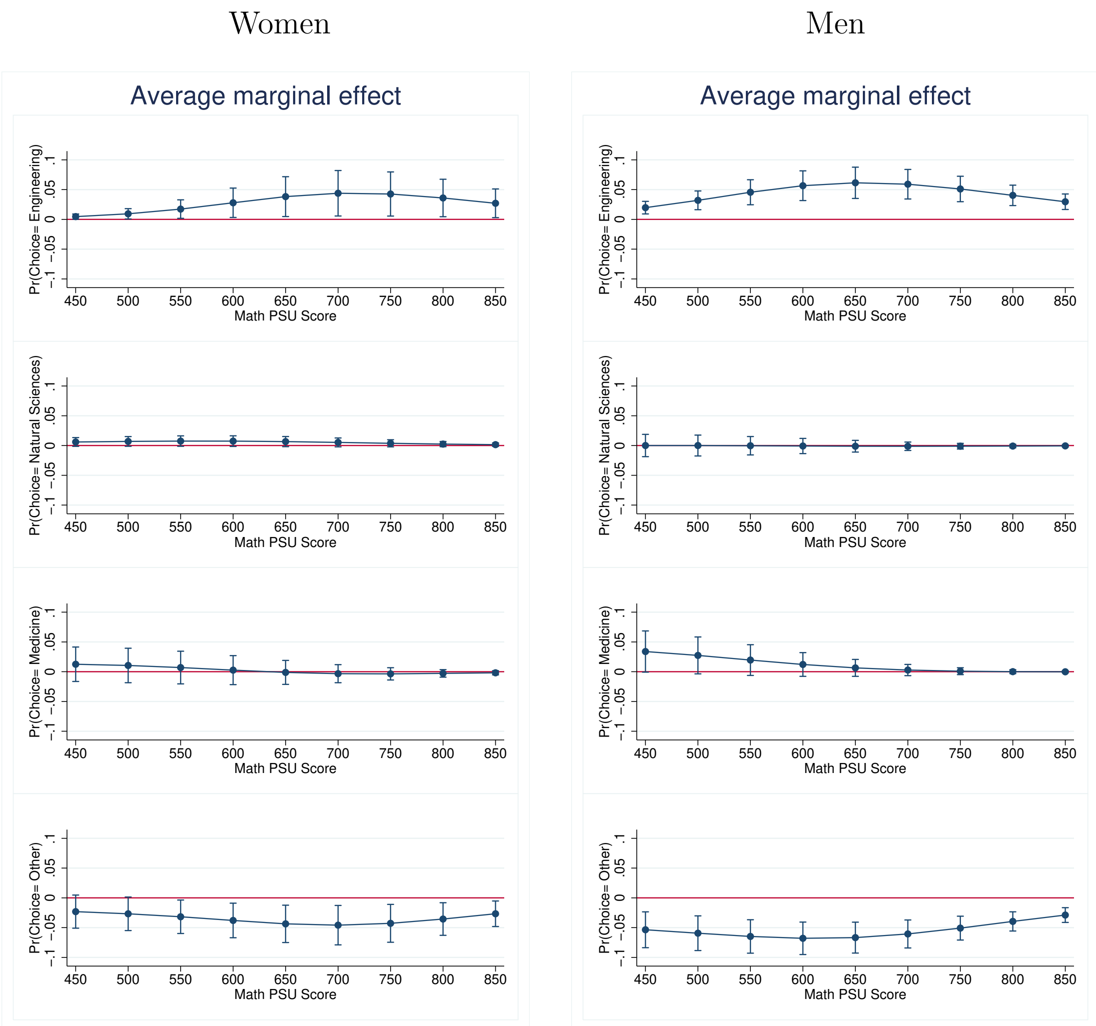
Women and men in the science subsample are less likely to choose a degree in the "Other" category when they benefit from the ranking bonus. This effect is consistent and statistically significant across levels of performance in the mathematics test. The drop in the likelihood is compensated for by an increase in the probability of choosing degrees in medicine at lower mathematics performance levels (scores lower than 600 points). Women and men in the science subsample are more likely to choose a degree in the engineering field when they benefit from the ranking bonus. The effect is consistent across all levels of performance in the mathematics test. However, it is only statistically significant for boys. In contrast, there is a negative average marginal effect of the ranking bonus on choosing a degree in the "Other" category that is statistically significant for boys and girls.

The results for the students that took both tests follow a similar trend. However, the effects are slightly higher for students in this subsample. The MEM on choosing engineering degrees is positive and statistically significant for women who scored more than 600 points in the mathematics test.



Notes: y-axis presents probability of chosing each field. The average marginal effect of the ranking bonus is presented at different levels of mathematic achievement, while all other variables are held at the sample average (Family income, school, and other test scores). Estimations with 95% confidence intervals

Figure 2.4: Average marginal effect of the ranking bonus on field choice by gender and math scores - Science subsample

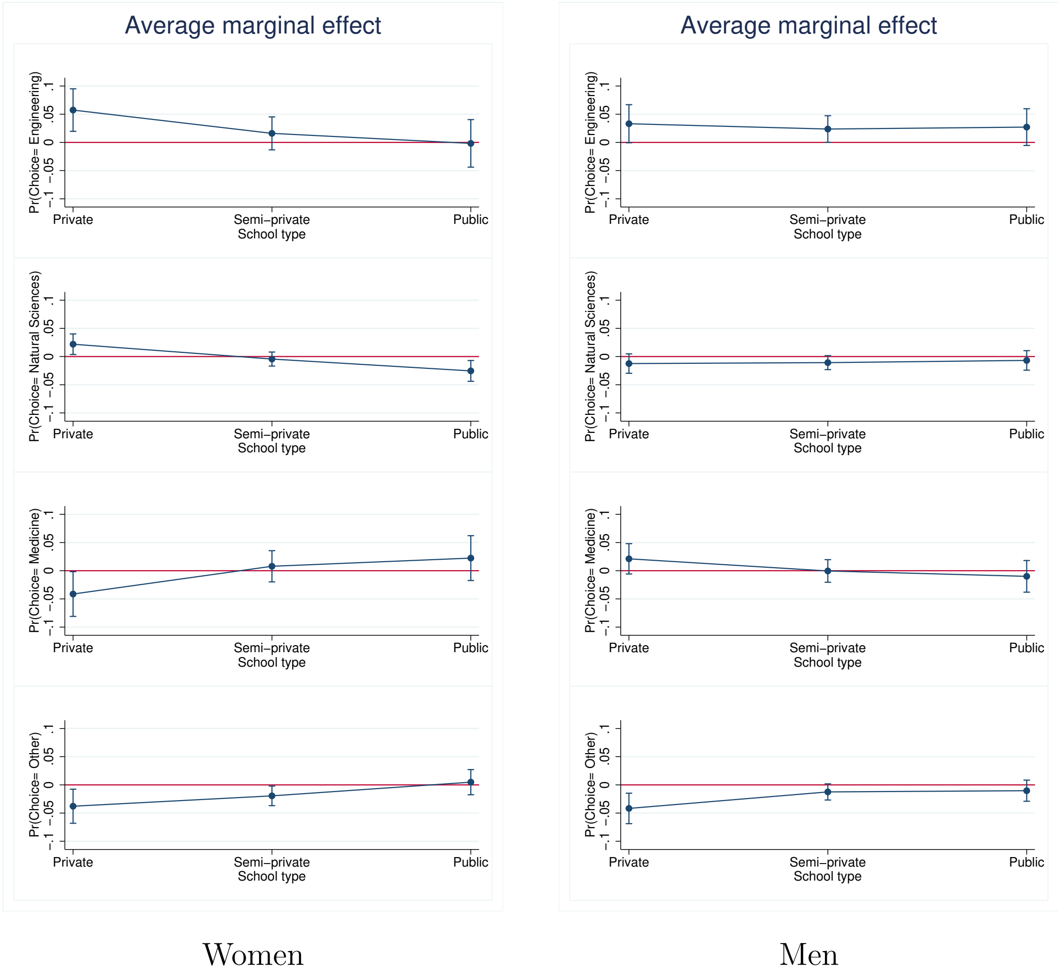


Notes: y-axis presents probability of chosing each field. The average marginal effect of the ranking bonus is presented at different levels of mathematic achievement, while all other variables are held at the sample average (Family income, school, and other test scores). Estimations with 95% confidence intervals

Figure 2.5: Average marginal effect of the ranking bonus on field choice by gender and math scores - Both tests subsample

Average marginal effect by school of origin

In Chile, the type of school of origin is very closely linked to class (Valenzuela et al., 2014). From Table 2.14 in the data section, we can see that families in the lower-income bands primarily send their children to public or semi-private schools, families in the middle choose semi-private schools, and families in the higher income bands choose private schools. Therefore, disaggregating the results by school of origin allows us to analyze class heterogeneity in the reform's effects on student choices. Figure 2.6 and 2.7 present the average marginal effect of the ranking bonus on field choice disaggregated by school type. The panels on the left-hand side show the results for women, and the right-hand panels show the results for men in terms of choosing a degree in the fields of engineering, natural sciences, health, or other.

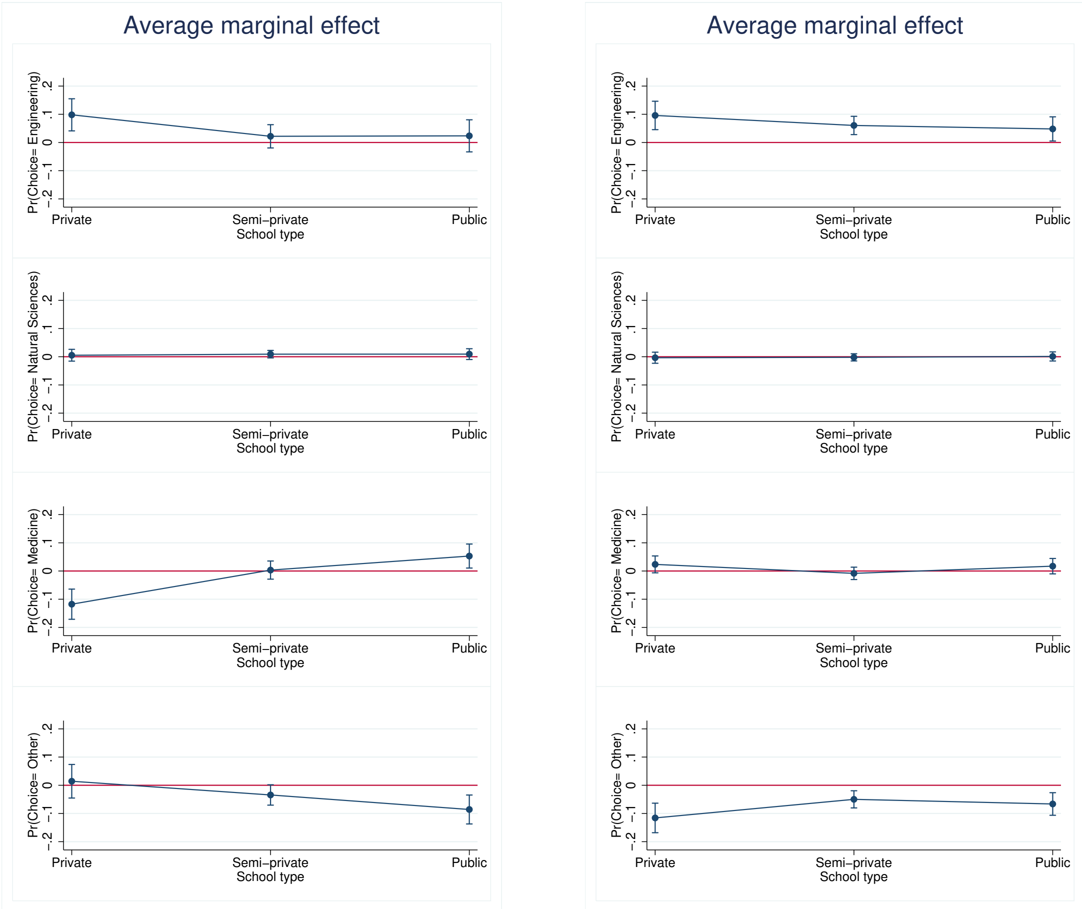


Notes: y-axis presents probability of chosing each field. The average marginal effect of the ranking bonus is presented for students from each type of school. Math scores are held at 650 points, while all other variables are held at the sample average (Family income, and other test scores). Estimations with 95% confidence intervals

Figure 2.6: Average marginal effect of the ranking bonus on field choice by gender and school type - Science subsample

The graph at the top-left in Figure 2.6 shows the average marginal effect on the probability of choosing an engineering degree for women who took the science test only. Girls who attended private schools are, on average, around 6% more likely to choose a degree program in the engineering field when they benefit from the reform. Meanwhile, there is no significant effect on girls from public or semiprivate schools. Girls from private schools are also more likely to choose a degree in natural sciences, while the opposite is true for girls from public schools. For degrees in the health field, the trend is reverted. The average marginal effect on choosing a degree in health is negative for girls who graduated from a private school. The graph at the top right shows the results for boys in the science subsample. There are no significant effects on choosing a degree in engineering for boys from any type of school. The graphs at the bottom show the effects on choosing degrees in other fields; the effects are negative and statistically significant for girls and boy from private schools but are not for students from public schools.

Figure 2.7 shows the results for students who took both tests. Girls from private schools are around 10% more likely to apply for a degree in engineering when they benefit from the reform, and around 10% less likely to choose a degree in health. In general, boys are more likely to choose a degree program in the engineering field when they benefit from the ranking bonus.



Women Men

Notes: y-axis presents probability of chosing each field. The average marginal effect of the ranking bonus is presented for students from each type of school. Math scores are held at 650 points, while all other variables are held at the sample average (Family income, and other test scores). Estimations with 95% confidence intervals

Figure 2.7: Average marginal effect by gender and school type - Both tests subsample

The effect of the reform on field choice is heterogeneous across types of schools. In general, boys and girls who get the ranking bonus are less likely to apply to degree programs in fields other than medicine, engineering, and natural sciences. For girls from private schools, the ranking bonus leads to an increase in the probability of applying to degree programs in engineering. Meanwhile, for girls from public schools, the bonus increases the likelihood of applying to degree programs in the medical field. The effect is particularly prominent for girls who took both tests. Meanwhile, boys from public schools are less sensitive to the ranking score.

Although I do not test what mechanism are behind this result, the academic literature provides some insights on possible explanation for the heterogeneity of the effect. A few students point out to different expectations for students from different social backgrounds. For instance, investigating beliefs about salaries by major Conlon (2019) shows that there is large heterogeneity across individuals from different socioeconomic backgrounds. Furthermore, Zimmerman (2019) and Friedman (2020) show that students from lower socioeconomic backgrounds tend to have lower incomes than their peers, which suggests that the heterogeneity in beliefs might be well funded. Tables 2.12 and 2.13 (in the appendix) present the statistics on income distribution across fields for women and men, respectively. If, in accordance with the findings of Zimmerman (2019) and Friedman (2020), women from public schools expect to get incomes in the bottom half of the income distribution, then they will be better off choosing degrees in the healthcare field. On the contrary, if women from private schools expect salaries at the top end of the distribution, they are better off attending degrees in the engineering field.

In addition to earnings expectations, the cost of education might differ for applicants from different types of schools. Degrees in engineering have a longer real duration than degrees in medicine. The length of a degree is particularly important for students from lower socioeconomic backgrounds, for whom being outside the labor market is more costly than for students from higher socioeconomic backgrounds. Engineering degrees are then more costly and potentially less profitable for girls from public schools, which might explain the differences in behavior across classes.

2.5.4 Enrolment by gender

The reform influenced students' courses of study by changing their final applications scores. A change in the final application scores impacts both students' applications and their likelihood to be accepted on their degree of choice. Final enrolments depend on students' choices, their application scores and degree programs vacancies. Therefore, along with evaluating students' first choice, I analyze the impact of the reform on final enrolment. I estimate the multinomial model from equations (2.2) and (2.3) but use enrolment instead of applications.

Table 2.9 shows the predicted probabilities for enrolment, which follow a very similar pattern than applications (Table 2.8). Men are considerably more likely to enroll in a degree in engineering than women from both subsamples. When looking at the effect of the ranking variable on choosing an engineering degree, we can see that there is little to no difference for men in the science subsample but a significant one for men in the "both tests" subsample. Women who benefit from the ranking bonus are more likely to enroll in engineering degrees, independent of the test they took. bonus are more likely to enrol in engineering degrees independently of the test they took.

Figure 2.11 in the appendix shows the average marginal effect of the ranking bonus on enrolment for students who took both tests. Girls and boys at all levels of mathematics achievement are more likely to enroll on a degree in engineering when they benefit from the reform. By contrast, the average marginal effect of choosing a degree in a field other than engineering, health and natural sciences is negative for pupils across different mathematics scores.

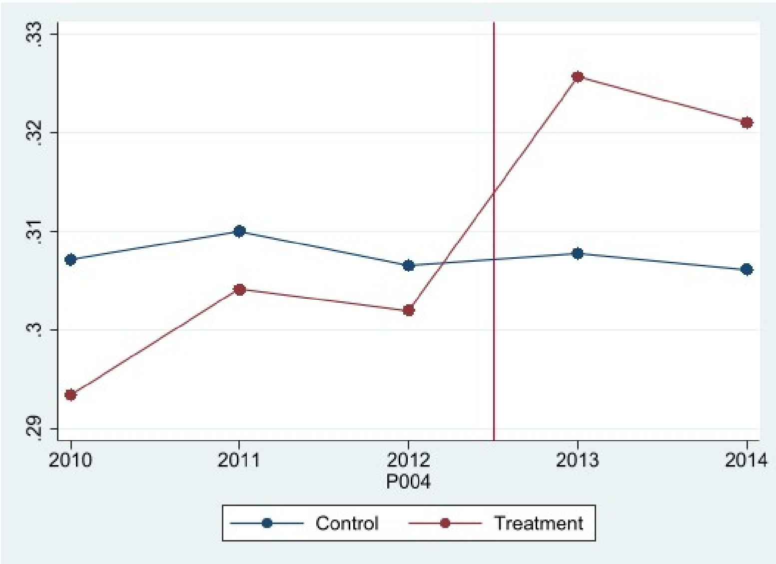
Table 2.9: Predicted probabilities (PP) of enrolment by gender

	Science test		Both tests	
	(1)	(2)	(3)	(4)
Choice	PP Women	PP Men	PP Women	PP Men
Other				
Ranking $< NEM$	0.186***	0.142***	0.378***	0.305***
	(0.009)	(0.006)	(0.010)	(0.008)
Ranking $> NEM$	0.164***	0.148***	0.335***	0.271***
	(0.005)	(0.005)	(0.006)	(0.006)
Engineering				
Ranking $< NEM$	0.246***	0.666***	0.182***	0.514***
	(0.010)	(0.009)	(0.008)	(0.009)
Ranking $> NEM$	0.261***	0.665***	0.201***	0.547***
	(0.006)	(0.007)	(0.005)	(0.007)
Medicine				
Ranking $< NEM$	0.504***	0.119***	0.394***	0.121***
	(0.012)	(0.006)	(0.010)	(0.006)
Ranking $> NEM$	0.515***	0.126***	0.411***	0.125***
	(0.007)	(0.005)	(0.006)	(0.005)
Natural Sciences				
Ranking $< NEM$	0.065***	0.072***	0.047***	0.060***
	(0.006)	(0.005)	(0.004)	(0.004)
Ranking $> NEM$	0.060***	0.060***	0.053***	0.057***
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	9,686	12,179	12,126	13,287

Notes: All predictors at their mean value. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

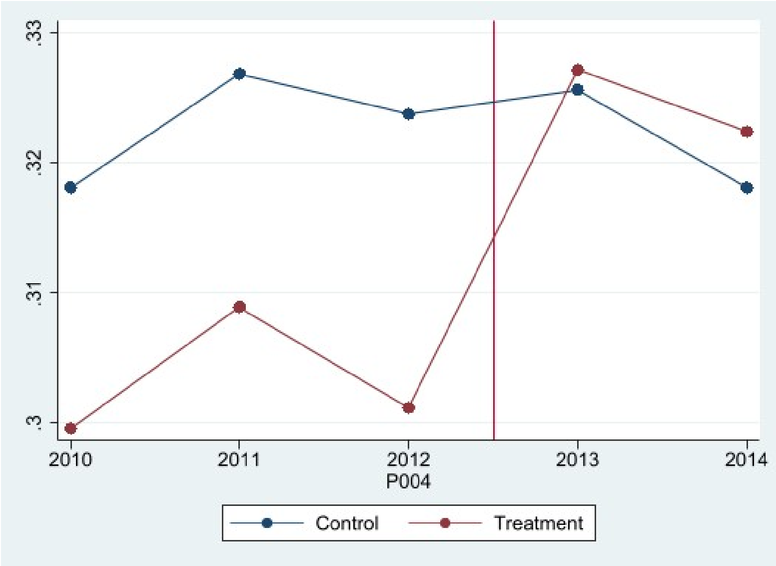
2.5.5 Quasi experimental Analysis

In this section, I present the results of a triple difference in differences estimation (DDD) on applications and enrolments to engineering degrees. Because of data constraints I define the pre-reform treatment and control groups using the proxy-measure described in section 2.3.3. Before presenting the results of the double and triple differences estimation for applications and enrolments, I show that the treatment and control groups have parallel trends in outcomes. Figure 2.8 shows the proportion of students who apply for an engineering degree as their first option each year, separated by treatment status. The red line shows the trend for the treatment group, and the blue line the trend for the control group. Note that the pre-reform cohorts (2010-2012) data are from the proxy-treatment and proxy-control groups. Figure 2.9 shows the proportion of students who enrolled in an engineering degree by treatment status in the same pre- and post-reform period. From Figures 2.8 and 2.9, we can see that both outcomes, applications and enrolments, follow similar trends across the treated and untreated groups before the treatment and that there is a clear change in trends after the reform.



Note: To the left of the red vertical line are the pre-reform years and to the right the post-reform years.

Figure 2.8: Proportion of students who apply to an engineering degree as their first option by treatment status



Note: To the left of the red vertical line are the pre-reform years and to the right the post-reform years.

Figure 2.9: Proportion of students who enrolled in an engineering degree by treatment status

Tables 2.10 and 2.11 show the estimated coefficients of the differences-in-differences (DD) and triple differences (DDD) estimation of applications and enrolments for engineering degrees. For Table 2.10, the dependent variable is a dummy that indicates if the student applied to a degree program in the engineering field as their first option or not. Column (1) shows the DD estimation results for women. The DD coefficient indicates that being benefited by the ranking bonus caused an increase in women applications of about 1.5 percentage points. Column (2) presents the results of the same equation but for the men's sample. For men, being a beneficiary of the ranking bonus increases their applications by approximately 3 percentage points. Results in column (3) show the results of the DDD equation. The DDD coefficients indicate that the difference in the effect of the ranking bonus between men and women is statistically significant at the 0.001% level. Thus, even though men and women are more likely to apply to engineering degrees when they receive the ranking bonus, the effect is larger for men. Table 2.11 shows the results for equations (3) and (4) for enrolment as the dependent variable. The results for the DD estimation (columns 1 and 2) and the results for the DDD estimation (column 3) show a similar pattern to the results for applications. Women are around 1.6% more likely to enrol in engineering degrees when benefited from the ranking bonus, while men in the treatment group are 2.8% more likely to enrol in this field. The DDD coefficient in column 3 shows that the difference in the reform effect between men and women is also statistically significant at the 0.1% level. These results confirm the results from the MNL analysis presented in section 2.5.3 The reform caused an increase in application and enrollments for engineering degrees for both men and women, but, at the individual level, the effect is bigger for men.

Table 2.10: DD and DDD estimation for applications to engineering degree-programmes

	Women	Men	Whole Sample
Time	0.009	0.016*	0.03***
	(1.46)	(2.43)	(8.61)
Treated	-0.004	-0.002	0.009***
	(-1.01)	(-0.56)	(3.48)
DD	0.015*	0.031***	
	(2.37)	(4.42)	
Sex			-0.201***
			(-79.58)
DDD			-0.016***
			(-3.81)
Constant	0.182***	0.288***	0.320***
	(4.68)	(7.17)	(11.29)
N	70035	76517	146552

Notes: All models include year fixed effects, region fixed effects and controls for language and mathematics PSU scores, t statistics in parentheses.
* p <0.05, ** p <0.01, *** p <0.001

Table 2.11: DD and DDD estimation for enrolment to engineering degree-programmes

	Women	Men	Whole Sample
Time	0.004	0.0169**	0.029***
	(0.70)	(2.61)	(7.45)
Treated	-0.022***	-0.012**	-0.004
	(-5.65)	(-2.64)	(-1.75)
DD	0.016**	0.028***	
	(2.68)	(4.14)	
Sex (Female=1)			-0.189***
			(-74.90)
DDD			-0.0176***
			(-4.26)
Constant	0.147***	0.335***	0.329***
	(4.19)	(9.24)	(12.68)
N	69139	78795	147934

Notes: All models include year fixed effects, region fixed effects and controls for language and mathematics PSU scores, t statistics in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.6 Conclusions

This paper examines the role of university admission systems as a structural factor reproducing sex imbalances in higher education. I used data from Chile and took advantage of a reform that comprised a new criterion for admission based on school performance. I hypothesized that the reform was not neutral to gender and tested this hypothesis through two main analyses. First, I analyzed the reform's targeting properties using linear regression to assess differential trends in women's and men's application scores and simulations that compared applicants' admission priorities with the reform and with a counterfactual scenario. By doing this, I determined whether there were gendered consequences of the reform on its beneficiaries. Then, I analyzed the effect of benefiting from the reform on students' application choices.

The first analysis revealed that women saw an improvement in their admission priority relative to men because of the reform. Then, the examination of students' applications reveals three main findings:

(i) For test choice, the new criterion has a bigger effect on women than men. Furthermore, only women see an increase in the marginal probability of choosing to take the science test when they benefit from the policy change. The importance of this result lies in the consequences it has for applications to degree programs; this is because taking the science test opens up the option of applying to degrees in engineering and healthcare. The choice of test and field of study are intrinsically linked but not simultaneous. Therefore, I cannot identify the full effect of the reform on field choices as part of the effect is absorbed in test choice.

(ii) There are substantial gender differences in the expected probabilities of applying to each field. When following a degree in the sciences, men are considerably more likely to apply to engineering than any other field. In contrast, women are substantially more likely to apply to a degree in the healthcare field. These results are consistent with the ones of Bordón et al. (2020) and Rapoport & Thibout (2018), who find similar gender differences in the field of choice in France and Chile, respectively.

(iii) I find that the policy change did affect students' choices. In general, men and women who benefited from the ranking bonus were less likely to apply to and enroll in programs in the lower-paid fields. This group of applicants exhibited an increase in the probability of choosing degrees in medicine and engineering. The effects were persistent across different levels of achievement in the PSU tests. However, I found some heterogeneity in the response of students from different social backgrounds. For girls from private schools, for instance, there was an increase in the probability of applying to degree programs in the field of engineering. Meanwhile, for girls from public schools, the ranking bonus increased the likelihood of applying to degree programs in the medical field.

These findings show that putting less emphasis on competitive tests while making room for alternative measures of achievement can help to reduce the gender imbalances in fields of study. The new admission criterion allowed more women to apply and enroll in the more in demand and competitive degree programs by improving their admission priorities and altering their behavior.

I conclude that admission systems are not neutral to gender. Understanding that they are part of a structure that might lead to inequality is crucial to inform better policy. My study provides empirical evidence on how using alternative achievement measures can help correct gender biases in admission systems that rely heavily on standardized test scores. This information can be used by university authorities and policymakers to develop targeted interventions to reduce the gender imbalances in fields of study.

2.7 Appendix

2.7.1 Distribution of monthly income by field

Table 2.12: Average monthly income in clp for women

Field	Percentile 10 ($)	Percentile 25 ($)	Percentile 50 ($)	Percentile 75 ($)	Percentile 90 ($)
Agriculture	309877.1	392987.6	544097	779134.3	1141376
Arts	323842.4	423716.7	578620.2	798886	1109849
Business	383792.5	551476.7	835070.4	1230930	1765173
Engineering	410892.2	620236.9	916401	1338152	1980126
Humanities	340934.9	449518.1	618396.2	861954.3	1144469
Medicine	431885	672386.1	970992	1292744	1622273
Natural Sciences	354267.5	470130.8	759308	1086226	1677584
Social Sciences	333085.9	444691.1	603838.6	803209.6	1078483
Technical	353396.9	476840.9	665324.2	973220.8	1465607
Total	379199.7	545983.5	782281.7	1081872	1460345

Source: MIFUTURO, a public database on labour market outcomes for Chilean tertiary education graduates.

Table 2.13: Average monthly income in clp for men

Field	Percentile 10 ($)	Percentile 25 ($)	Percentile 50 ($)	Percentile 75 ($)	Percentile 90 ($)
Agriculture	317758	412568.4	576095.5	835167.5	1288816
Arts	332875.6	442182.5	608332.2	843603.1	1172325
Business	385587.2	556947.4	854250.6	1259097	1801892
Engineering	412489.7	626383.8	930528.2	1362678	2023264
Humanities	340908.5	458453	647090.4	928571.9	1254299
Medicine	440881.4	760813.7	1123389	1494092	1884198
Natural Sciences	358726.5	475113.5	784362.3	1137079	1733295
Social Sciences	334784.5	451712.1	619557.5	828302.3	1128148
Technical	353396.9	476840.9	665324.2	973220.8	1465607
Total	385519.6	571410.6	840332.7	1198219	1696039

Source: MIFUTURO, a public database on labour market outcomes for Chilean tertiary education graduates.

2.7.2 Family income and school of origin

Table 2.14: Family income and school of origin

Family Income Band	Lower bound	Upper bound	Percentage of students		School of origin	
	Thousands of CLP ($)	Thousands of CLP ($)	(%)	Public (%)	Semiprivate (%)	Private (%)
1	0	144	7.53	44	52	4
2	144	288	19.60	39	57	3
3	288	432	15.39	32	62	5
4	432	576	10.26	27	64	8
5	576	720	7.80	23	64	12
6	720	864	5.73	20	62	17
7	864	1008	5.65	16	59	25
8	1008	1152	3.87	13	52	35
9	1152	1296	3.17	10	49	40
10	1296	1440	5.59	4	25	71
11	1440	1584	1.43	7	42	50
12	1584	-	12.99	3	19	78
Total			100	24	52	23

Source: DEMRE data.

Table 2.15: Odd ratios for test choice

	Women subsample		Men Subsample	
	Science	Both	Science	Both
	(1)	(2)	(3)	(4)
Ranking Dummy =1	1.093**	1.102**	1.032***	1.068***
	(0.024)	(0.027)	(0.026)	(0.029)
NEM Score	1.004***	1.004***	1.005***	1.005***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.110**	0.058***	0.105**	0.058
	(0.052)	(0.038)	(0.049)	(0.039)
Akaike Inf. Crit.	114,057.4	114,057.4	101,842	101,842

Notes: All models include Includes controls for family Income and School of origin Dummies. Ranking Dummy is equal to 1 for students who benefitted from the reform. ***Significant at the 1 percent level, **Significant at the 5 percent level, *Significant at the 10 percent level.

2.7.3 Odd ratios for field choice

Table 2.16: Odd ratios for field choice - Students who took the science test only

	Women subsample			Men Subsample		
	ENGINEERING	MEDICINE	NATURALSCIENCES	ENGINEERING	MEDICINE	NATURAL SCIENCES
	(1)	(2)	(3)	(4)	(5)	(6)
Raking Dummy	0.218**	0.165**	0.071	0.222***	0.192**	-0.001
	(0.157)	(0.137)	(0.236)	(0.131)	(0.157)	(0.214)
Semi-private School	0.564***	0.219**	0.204	0.465***	0.661***	0.551***
	(0.102)	(0.088)	(0.150)	(0.081)	(0.097)	(0.129)
Public School	0.691***	0.271**	0.254	0.476***	0.704***	0.611***
	(0.122)	(0.106)	(0.181)	(0.097)	(0.115)	(0.151)
MathScore	0.014***	-0.007***	-0.002	0.009***	-0.012***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
LengScore	-0.003***	0.002***	0.000	-0.003***	0.002***	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
SciencScore	-0.003***	0.009***	0.007***	-0.001	0.017***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
NEMscore	0.001**	0.004***	0.001	0.002***	0.005***	0.002***
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Constant	-5.643***	-3.107***	-4.767***	-3.740***	-7.257***	-6.338***
	(0.408)	(0.351)	(0.572)	(0.318)	(0.378)	(0.512)
Observations	12,887	12,887	12,887	14,952	14,952	14,952

Notes: The reference level "Other" includes the following fields: agriculture, Business and Technical education. All models include Includes controls for family Income and School of origin Dummies. ***Significant at the 1 percent level, **Significant at the 5 percent level, *Significant at the 10 percent level.

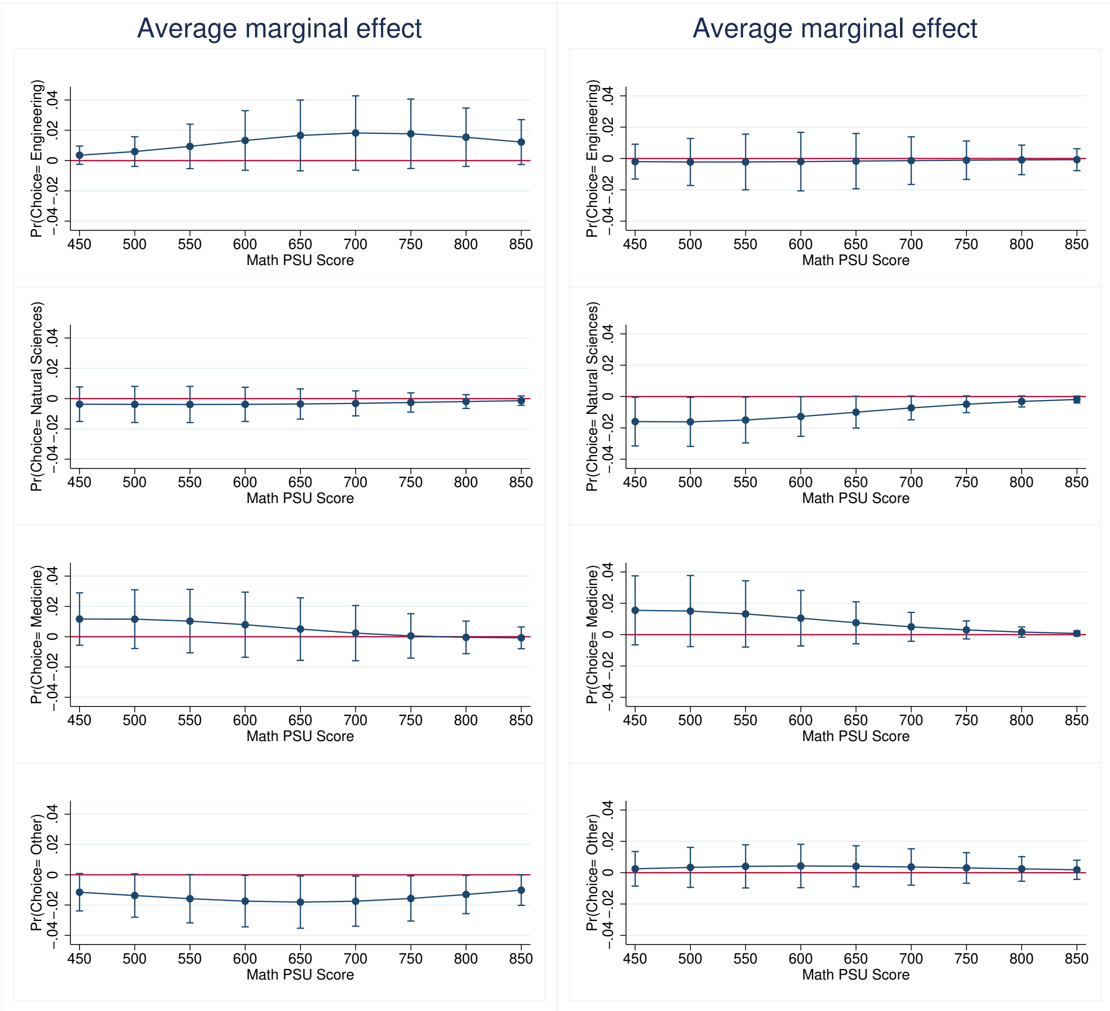
Table 2.17: Odd ratios for field choice - Students who took both tests

	Women subsample			Men Subsample		
	ENGINEERING	MEDICINE	NATURALSCIENCES	ENGINEERING	MEDICINE	NATURAL SCIENCES
	(1)	(2)	(3)	(4)	(5)	(6)
Ranking Dummy	0.240**	0.099	0.395**	0.344***	0.274***	0.175
	(0.094)	(0.077)	(0.199)	(0.068)	(0.102)	(0.179)
Semi-private School	0.639***	0.223**	0.307	0.590***	0.614***	0.366*
	(0.129)	(0.097)	(0.237)	(0.087)	(0.133)	(0.198)
Public School	0.493***	0.179	0.274	0.523***	0.566***	0.395*
	(0.146)	(0.110)	(0.280)	(0.099)	(0.150)	(0.223)
MathScore	0.012***	-0.007***	-0.000	0.010***	-0.009***	-0.005***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
LengScore	-0.010***	-0.006***	-0.008***	-0.009***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
SciencScore	0.005***	0.016***	0.014***	0.004***	0.020***	0.016***
	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.002)
NEMscore	0.002***	0.003***	0.000	0.001	0.004***	0.001
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Constant	-6.290***	-3.541***	-6.118***	-4.029***	-6.671***	-6.626***
	(0.442)	(0.343)	(0.813)	(0.296)	(0.436)	(0.742)
Observations	8,008	8,008	8,008	8,857	8,857	8,857

Notes: The reference level "Other" includes the following fields: agriculture, Business and Technical education. All models include Includes controls for family Income and School of origin Dummies. ***Significant at the 1 percent level, **Significant at the 5 percent level, *Significant at the 10 percent level.

2.7.4 Results for enrolment

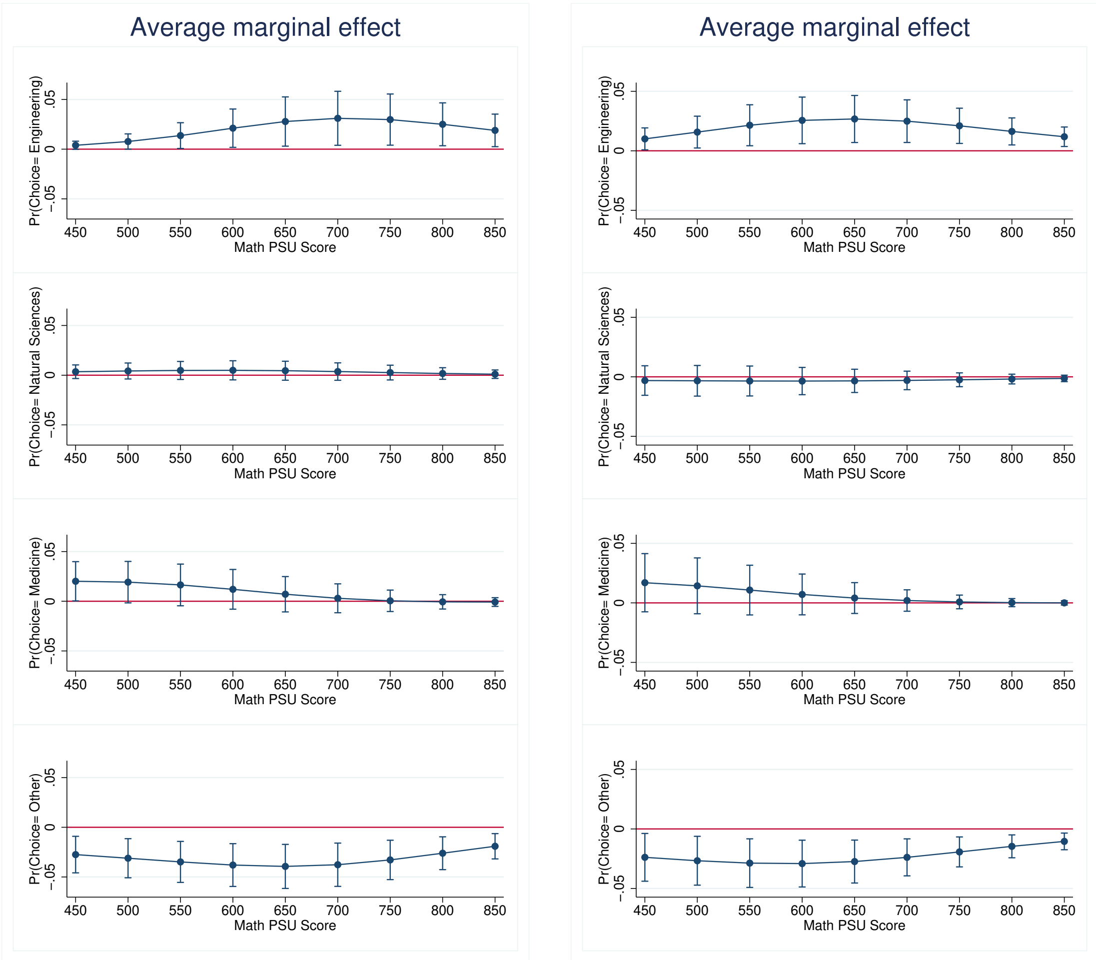
Women | Men



Note: Estimations with 95% confidence intervals

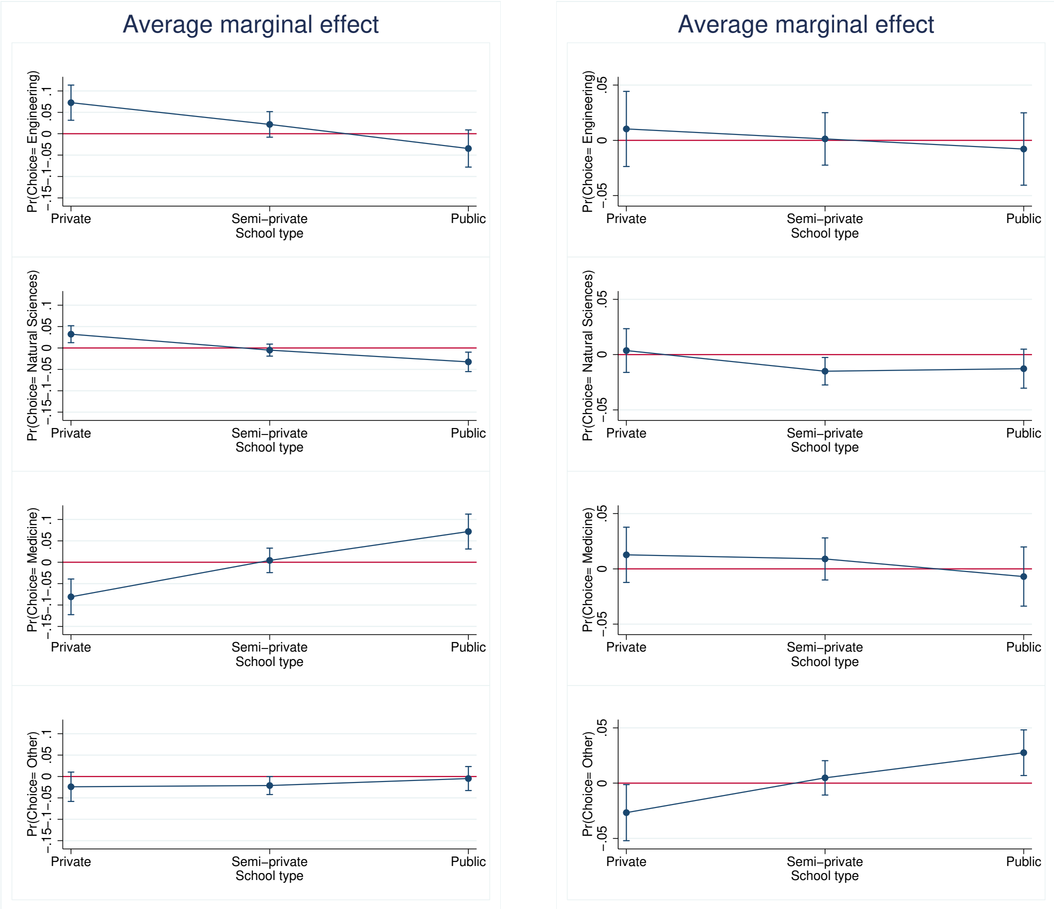
Figure 2.10: Average marginal effect on enrolment by gender and math scores -Science subsample

Women Men



Note: Estimations with 95% confidence intervals

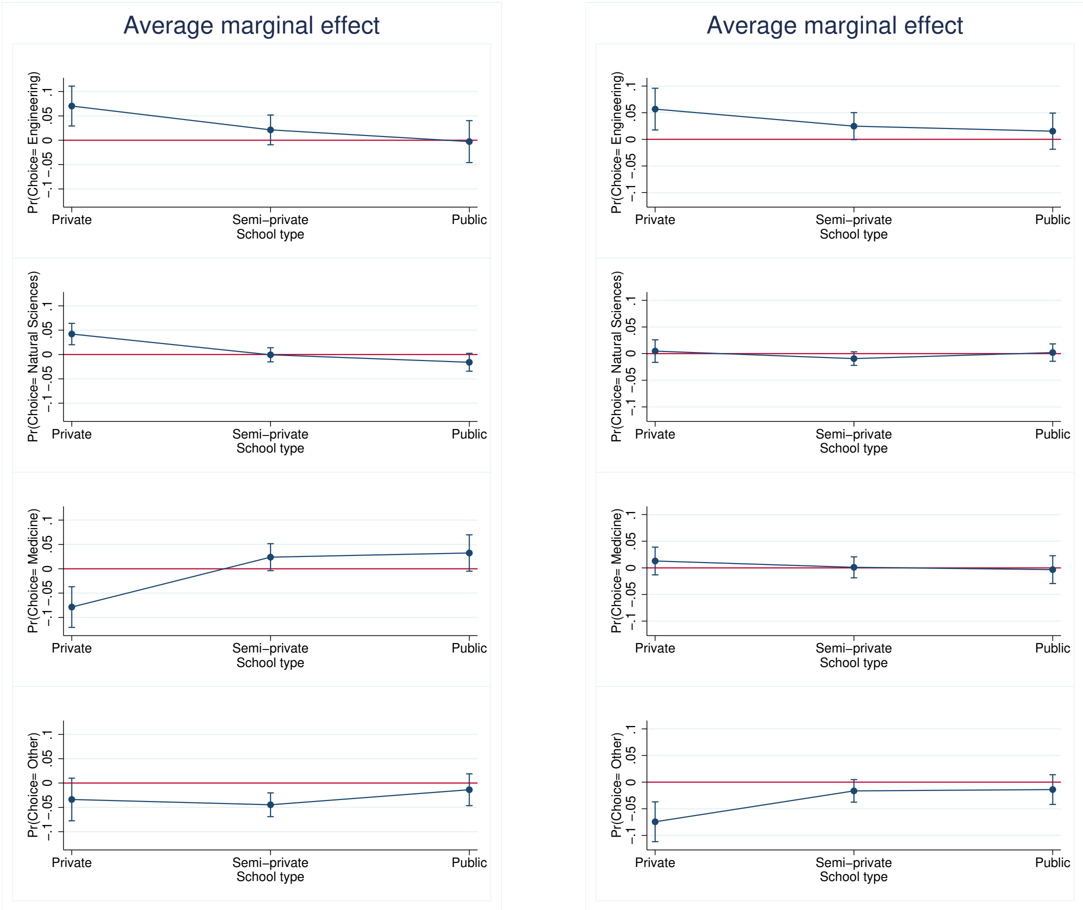
Figure 2.11: Average marginal effect on enrolment by gender and math scores - Both tests subsample



Women Men

Notes: Estimations with 95% confidence intervals. Marginals effects calculated at Math score=650

Figure 2.12: Average marginal effect on enrolment by gender and school type - Science subsample



Women Men

Notes: Estimations with 95% confidence intervals. Marginals effects calculated at Math score=650

Figure 2.13: Average marginal effect on enrolment by gender and school type - Both tests subsample

Chapter 3

Policy evaluation of gender affirmative action in engineering schools

Introduction

Although women constitute half of the population, they are still a minority in certain fields of study. On the global scale, only 35% of students enrolled in degrees from the fields of science, technology, engineering, and mathematics (STEM) are women (Chavatzia, 2017). This underrepresentation of women in science is disadvantageous for women (Blau and Kahn, 2017) and potentially harmful for society in general (Kolovich et al., 2020).[1] If it is in the interest of policy makers and university administrators to correct this imbalances, gender affirmative action in scientific fields can potentially be a powerful and effective tool.

[1] In a recent review of the extent, trends, and explanations of the gender wage gap, Blau and Kahn (2017) point to the lack of women in STEM fields as a relevant explanation of the gender wage gap. Whereas Kolovich et al. (2020), argue that gender inequalities in education slow down economic growth.

Higher education institutions have used affirmative action policies for two main reasons: (1) to reverse some of the effects of past and contemporary discrimination that have affected a particular group, and (2) to respond to their institutional interest in maintaining a sufficient share of minority students on their campuses, and to gain the positive academic and social benefits of a diverse student body. Although the intent is noble, the use of this type of policy has often been controversial. The critiques surround the issue of displacement, i.e., who the interventions leave out, and how affirmative action policies harm the intended beneficiaries by placing them in degree programmes for which they are ill prepared. [2]

Although the literature has contributed to informing this debate, it has done so by evaluating interventions that are focused on addressing disadvantages related to race, class, or caste, which are the ones that have most commonly been implemented in universities. Additionally, the evaluation of the role of affirmative action in bringing the benefits of a diverse body of students has remained mostly unexplored. In this paper, I take advantage of two gender affirmative action policies implemented in Chilean engineering schools, and evaluate them in terms of their effectiveness in reversing discrimination, and in regard to the effects of increasing the share of women on the academic outcomes of all members of the student body.

I use Chile as the case study, as it is a pioneer in the use of gender affirmative action in STEM and it provides a unique opportunity to learn about the effectiveness and consequences of this type of intervention. In 2014, the two most prestigious Chilean universities implemented policies that sought to increase women's participation in their engineering schools. The University of Chile (UCH) expanded its capacity by around 4% to add 40 extra female-only seats to each incoming cohort. In parallel, the Pontifical Catholic University of Chile (PUC) implemented changes, not in its admission process but in the curriculum and female representation among faculty members.

There are several characteristics of the Chilean case that make it particularly

[2] See Teigen (2000) for an overview of the main arguments in favour and against the use of affirmative action.

interesting to study. First, the higher education system is centralised, and the admission rules are clear and public. Second, while both universities sought to increase women's participation, they used very different strategies. Thus, this allows me to analyse two distinct policies at once. Finally, a crucial advantage of this case is that the rest of the universities in the country did not implement any related policies, and therefore, the setting provides a natural experimental context that allows for isolating the effect of the reforms on achieving their goals.

I use administrative data from the Chilean centralised higher education admissions system and data on students' academic performance from UCH to answer three main questions: (1) Do the policies work? (2) Do they hinder the academic ability distribution of incoming students? And finally, (3) are there any effects of bringing more women into the engineering school on the academic performance of their peers? To answer these questions, I carry out three sets of analyses.

First, I study the effectiveness of these two distinct policies in attracting more women to the engineering schools. I use the centralised admissions system data and a difference-in-differences approach to analyse the effect of both policies on women's applications and attendance. I compare the UCH and PUC outcomes with those of the other 28 Chilean engineering schools. I find that the PUC and UCH initiatives caused an increase in the share of women who apply and enrol in their engineering degree programmes. Moreover, in the case of UCH, the increase in women's enrolment was far beyond the mechanical 4% increase that the 40 extra seats were expected to bring.

Then, I look for changes in the average application scores of men and women admitted to UCH and PUC before and after the reform. Students' academic ability is relevant for university admission because it can affect the quality of the educational institution or potentially harm the intended beneficiaries. I compare the trends in the average application scores of men and women before and after the interventions, and I find that there was no significant change in the average academic ability of the students. Through these two first analyses I show that the gender affirmative policies implemented by both universities changed the gender composition of the incoming cohorts of the engineering schools without having strong effects on students' average academic ability.

In a third set of analysis, I assess how exposure to more female classmates, and to higher achieving peers affects educational outcomes. I use a peer-effects model and data from UCH to determine the effects of classroom gender composition and average academic ability on students' academic outcomes. I look at students' grades in four first-year subjects, the scores on a collaborative engineering project and drop-out rates. I find that increases in female participation positively impact students' performance in the collaborative project and lower the probability of drop-out for women.

This paper contributes to the literature on affirmative action in higher education in various ways. Because affirmative action is most commonly implemented in universities to correct race or ethnic imbalances, most studies on affirmative action in higher education have focused either on race-related policies (see for example, Arcidiacono (2005), Long (2007), Epple et al. (2008), Howell (2010), Francis and Tannuri-Pianto (2012a), Francis and Tannuri-Pianto (2012b)) or caste affirmative action policies (Bertrand et al. (2010), Frisancho and Krishna (2016), De Zwart (2000), Bagde et al. (2016)). Regarding gender imbalances, Lihamba et al. (2006) and Onsongo (2009) study affirmative action policies in Tanzania, Kenya, and Uganda, where the gender imbalance is prevalent across disciplines. To the best of my understanding, this is the first paper that provides causal estimates of the effectiveness of gender affirmative action policies that aim to increase female participation in STEM degrees. As the roots and structure of gender discrimination are different from those of race and ethnic discrimination (Sidanius and Veniegas, 2013), providing more evidence on the effects of affirmative action policies based on gender can help to better inform future policy.

In contrast to the UCH and PUC interventions, previously studied affirmative action policies tend to raise the share of students from the minority group using a deterministic approach, that is, through strict quotas that define the percentage of the minority group before the application and admission stages. Therefore, by examining the UCH and PUC policies, I also contribute to the literature on affirmative action in higher education by revealing the effects of alternative forms of affirmative action. This can be particularly helpful in settings where strict quotas policies are difficult to implement. Finally, I contribute to this literature by analysing further or indirect effects of these policies. Whereas other authors (see for example Bagde et al. (2016), Frisancho and Krishna (2016)) have studied the effects of affirmative action on the academic performance of the beneficiaries, by measuring peer-effects, I can identify the impact of the affirmative action policy on the academic outputs of beneficiaries and non-beneficiaries.

I conclude that, first, both policies were effective in increasing female participation in engineering schools. Second, the policies did not significantly affect the distribution of academic ability in incoming cohorts. Third, increases in the share of females did not hurt the academic achievement of peers, but helped reduce the probability of drop-out for girls. These results indicate that the reforms increased women's participation in engineering degrees by enhancing the admission rates and improving women's adherence to the degree programme. Taken together, these results show that we can correct past discrimination by using affirmative action policies without hurting academic quality or students' academic performance. Moreover, I provide evidence that by bringing more women into these male-dominated fields, we can also reduce the drop-out rates of women and, therefore, their persistence in STEM.

The rest of this chapter is structured as follows: Section 1 describes the two policy changes. In Section 2, I present the data and summary statistics. Then I present the empirical strategy and results for each research question in three separate sections. I start with the analysis of applications and attendance in Section 3, and then changes in the academic ability distribution in section 4, and finish with the effects on academic performance of peers in Section 5. Section 6 presents a discussion and conclusions.

3.1 Background and policy experiment

There are many possible methods that universities can use to raise the share of students that belong to a minority or disadvantaged group, for example, changing the admissions criteria for certain minority applicants (increasing the percentage of applicants in these groups that qualify for admission), undertaking targeted recruiting (which might increase the number of applicants from minority groups) or directly admitting a higher share of minority students who apply, among others. The latter option is the most traditional form of affirmative action and, although it is the most straightforward method, it can be controversial or even illegal. In the USA, for instance, public university systems were forced to eradicate the policies that explicitly favoured applicants because of their race.[3]

The top X per cent programmes implemented in Texas, California, and Florida are well-studied examples of policies that aim to increase the number of minority students who qualify for admissions.[4] These three US states grant automatic access to public universities for students whose grade point averages place them in the top x percent of their high school class. The programmes emerged as a response to the ban on traditional forms of affirmative action. The idea behind this was that students in the top 10 percent of their high school class might be fairly representative of the state as a whole. Therefore, guaranteeing admissions would indirectly increase the share of students from minority ethnic backgrounds.

While the US has banned explicit quota systems and point systems for minority applicants, university administrators still implement them in other countries. For instance, universities in India and Brazil have successfully implemented caste and race policies (see Bertrand et al. (2010) and Francis and Tannuri-Pianto (2012a)).

In Chile, where participation of women in STEM is low,[5] two pioneering univer-

[3] See Long (2007) for an overview of the history of affirmative action policies in the US.

[4] See for instance, Cortes and Lincove (2016), Cullen et al. (2013), Long and Tienda (2008), Niu et al. (2006).

[5] According to the Institute of UNESCO Statistics (UIS), Only 19% of STEM graduates in Chile are women.

sities have implemented distinct policies aimed at increasing the share of women on their engineering schools. The two universities, University of Chile (UCH) and Pontifical Catholic University (PUC), are affiliated with the Chilean centralised admission system described in Chapter 2, Section 2.2.

The University of Chile is a public university with five campuses all located in Santiago. It has more than 40,000 undergraduate and graduate students and offers more than 70 different bachelor and professional degrees, 39 doctoral programmes and 115 master's programmes (UCH, 2019). Its engineering school has 13 academic departments, 236 full-time professors, and close to 5500 students (UCH Engineering, 2019).

The Pontifical Catholic University of Chile is a traditional private university. In 2020, PUC offered 104 undergraduate degree programmes, 97 master-level programmes and 35 PhD programmes, and had around 30,000 students. The Engineering School has 4441 regular undergraduate students. In addition, its faculty is made up of 130 full-time and 215 part-time professors.

In 2014, UCH implemented an affirmative action policy named *Programa de Equidad de Género* (PEG) to increase female participation in the student body of the engineering school. The policy added 40 female-exclusive vacancies for each admission process. The new vacancies represent close to 5% of the total engineering school offers. The university assigns these slots to the 40 female applicants with the highest scores that have not got a place through the regular admissions process.

In the same year, PUC implemented an initiative that was different but had the same purposes. The programme, called "Women in engineering", aimed to attract more women to their classrooms. PUC implemented three main changes: (1) they changed the engineering school curriculum and created new majors with specialities that combine engineering with disciplines of more significant interest to women, such as biomedicine, architecture, and design; (2) they increased the number of female faculty members in the engineering school from 1 to 23; and (3) they organised talks so that female engineers could go to schools to talk about their experiences in their jobs to expose high school girls to female role models in engineering.

While the UCH policy operates at the admissions stage, the PUC policy tries to increase the application rates of women who qualify for admission.

3.2 Data and summary statistics

I use administrative data from the centralised admissions system for the analyses regarding attendance at engineering schools, and the academic ability of enrolled students. This dataset includes gender, age, academic background, high-school ranking, tests scores, preferences and final enrolment for every student who participated in a round of the national application process from 2007 to 2018. For the analysis of the academic performance of engineering students I use a second dataset provided by the University of Chile. The dataset contains individual-level data on first-year students covering the period from 2009 to 2017. These data include demographic characteristics, past measures of academic performance, including high school grades and admission tests scores, and grades for five first-year subjects, Algebra, Calculus, Physics, Computer Science and Introduction to engineering.

3.2.1 Centralised admission system data

The number of high school graduates that participate in the centralised admission process has been increasing in the last decade, and ranges from 90 thousand to 150 thousand each year. During the same period, female participation has increased to the point that there is no gender gap in participation in higher education. Table 3.1 presents relevant statistics on applications and enrolments by gender for UCH, PUC, all engineering degrees, and all degrees. The data for engineering degrees comes from the 30 universities that are part of the centralised admissions system that have at least one engineering programme. For those universities with more than one engineering programme, I have aggregated the data to have a single measure of female participation for each university.

From Table 3.1, column (4) we can see that during the 2009-2017 period the proportions of female applications (top panel) and enrolments (middle panel) in any university degree programme are close to 50%. In contrast, from column (3) we can see that when looking at engineering degrees only women still constitute a minority. Both the share of females among applicants, and the proportion of women enrolling in engineering degrees range from are around 26%.

Table 3.1, columns (1) and (2) shows the same statistics for UCH and PUC. Each year PUC has around 650 vacancies on its engineering programme. On the other hand, UCH provides approximately 700 under regular admissions (the average over the period is 692) and an additional 40 that are female-only seats. The data presented in this table only includes information for regular admissions, that is, it excludes the 40 extra seats. However, in both universities, the proportion of women applying to, and enrolling in the engineering schools has increased over the years.[6]

Table 3.1 also shows average application scores by gender. While the PSU test scores[7] range from 150 to 850, the average application scores of students who enrolled in a degree programme are in the top half of the score distribution. The statistics below are the average and standard deviation for students who enrolled in: (1) engineering at UCH, and (2) engineering at PUC, (3) any engineering degree, (4) any degree programme. On average, men have higher application scores, independently of their degree. The table also shows that the average application scores for students enrolled in engineering degrees at either UCH or PUC are higher than those enrolled in engineering degrees at other universities.

[6] See Table 3.14 in Appendix 3.7.1 for data on applications and enrolment by year.
[7] The PSU (*Prueba de Selección Universitaria*) is the national university admission test.

Table 3.1: Summary statistics: Centralised admission system

		(1) UCH	(2) PUC	(3) Engineering Degrees	(4) All Degrees
Applications					
Men	Mean (N)	1247	1134	18361	53699
	%	72.8	77.5	74.5	47.2
Women	Mean (N)	465	423	6295	60062
	%	27.2	22.6	25.5	52.8
All	Mean (N)	1713	1557	24657	113761
	%	100.0	100.0	100.0	100.0
Enrolment					
Men	Mean (N)	547	498	13601	33512
	%	79.1	78.3	73.3	51.0
Women	Mean(N)	144	132	4966	32170
	%	20.9	21.7	26.7	49.0
All	Mean (N)	692	629	18568	65682
	%	100.0	100.0	100.0	100.0
Application scores					
Men	Mean	748.6	768.3	604.3	598.5
	s.d.	29.1	28.0	116.6	123.9
Women	Mean	744.4	764.6	597.7	592.4
	s.d.	23.7	37.6	123.3	129.6
All	Mean	747.7	767.5	602.5	595.5
	s.d.	28.1	30.4	118.5	126.8

Notes: Mean (N) is average number over the period 2009-2017. Scale of application scores is 150-850. Columns "Engineering Degrees", and "All degrees" shows application statistics of students who submitted an application to at least one engineering degree programm, or to at least one degree programme (in any field), respectively. To see the data disaggregated per year see Appendix 3.7.1.

3.2.2 Data on academic performance

The data on academic performance contains information from students attending the UCH engineering school. Students who gain admission to this school have application scores that are much higher than those in other degrees or even in other engineering schools (see table 3.15). The student body also has average PSU mathematics scores that are at the very top of the score distribution. Table 3.2 shows students' average mathematics, and language scores and high school grades by gender. Men have higher application scores and mathematics scores than women on average, while women have higher language scores and high school grades. The table also shows the average grades for first-year subjects by gender. The grades are on a scale of 1 to 7. On average men have higher grades for every course except the introduction to engineering.

Table 3.2: Average Applications scores and average grades by gender

	Men	**Women**	**Total**
Previous Measures of Academic Achievement			
High School Grades	706.17	710.92	707.18
	(75.39)	(144.82)	(94.67)
Mathematics Test	779.96	754.69	774.54
	(42.6)	(42.51)	(43.82)
Language Test	688.05	698.52	690.3
	(57.92)	(54.38)	(57.33)
First year Academic Performance			
Computer Science	5.35	5.12	5.31
	(1.05)	(1.03)	(1.05)
Introduction to Engineering	5.88	5.98	5.9
	(0.49)	(0.48)	(0.49)
Physics	4.99	4.69	4.92
	(0.95)	(0.89)	(0.95)
Calculus	4.55	4.28	4.49
	(0.95)	(0.88)	(0.94)
Algebra	4.68	4.45	4.63
	(0.96)	(0.89)	(0.95)
Overall average	5.35	5.12	5.31
	(1.05)	(1.03)	(1.05)

Notes: Scale of previous measures of academic achievement is 150-850. Scale of first year academic performance is 1-7. Average values include whole sample (cohorts from 2009-2017). s.d. in parenthesis

3.3 Applications and attendance

To study the effectiveness of the two interventions in increasing women's participation in engineering schools, I look at changes in the application and enrolments rate of women in the treated schools and compare it to those rates in all of the other engineering schools.

There are a few mechanisms through which the UCH policy could increase the number of women applying to their degree programmes. For instance, women whose application scores are just below the previous year's cut-off might perceive a higher likelihood of admission than women with similar scores in the pre-reform cohorts. Even though students can apply for up to ten degree-programmes, the increase in the chance of being admitted could encourage applications from risk-averse women with scores around the cut-off. Thus, this effect would increase applications at the bottom of the score distribution. Other potential mechanisms include changes in preferences and aspirations. Shifts in preferences might occur if women are inclined to study or interact with other women. There are several reasons why women might prefer to interact with other women. For example, the literature on gender composition finds that women tend to perform better (Booth and Yamamura, 2018)[8], be more confident in their performance, and be more influential among their peers (Stoddard et al. (2020), Born et al. (2018)) in environments with a higher proportion of women. If this is the case, female applicants will be more likely to consider pursuing an engineering programme if they anticipate an increase in the share of women students. Lastly, the policy might also prompt changes in aspirations, whereby women who had not previously contemplated a career in engineering might consider it after learning about the policy. Lloyd et al. (2008) found an equivalent effect of the Texas top 10% plan on changing aspirations and expectations of minority students. If either of the last two mechanisms is at play, women's applications will increase at all levels of the application scores distribution; thus, the proportion of women admitted through regular admission should also increase.

UCH and PUC successfully raised the percentage of women in their engineering schools. In UCH, women's enrolment in 2013 for first-year students was 20.1%, and since then this percentage has increased consistently. Consequently, in 2018 the share of women among first-year students was 32.8%. At the same time, PUC's initiative increased the percentage of females enrolled in engineering from 19.5% in 2013 to 28.7% in 2018. Still, even though the data shows an increase in female enrolment at both engineering schools, we cannot immediately conclude that this

[8] Studying speedboat racesBooth and Yamamura (2018) shows that women perform better in single-sex competitions. Similarly, Dustmann et al. (2018) provides causal evidence that girls perform better when they attend single-sex schools.

increase is a direct result of the policies, as there could be confounding factors at play. Cultural trends, such as shifts in gender attitudes towards sciences, could also explain the changing gender composition of new cohorts of engineering students.

I use a differences-in-differences (DD) approach to isolate the effect of the treatments from potential confounding factors. The approach has been widely used in economics to identify the causal effect of a treatment (often the adoption of a policy change) in the absence of truly experimental data (see e.g. Angrist and Krueger (1999), Athey and Imbens (2006), Bertrand et al. (2004), Blundell and Dias (2009), Heckman et al. (1999), Lechner et al. (2011), Meyer (1995)). In its canonical format, the model considers two time periods, before and after the treatment, and two groups, the treated group and the control group. If the parallel trend assumption is fulfilled, one can measure the average treatment effect on the treated by comparing the average difference in outcomes experienced by the treated group to the average difference in outcomes experienced by the control group.

While UCH and PUC implemented policy changes that aimed to increase female applications, the other engineering schools in the country did not. Thus, this setting provides a natural experiment that allows me to compare the pre- and post-reform outcomes of the treated and untreated universities to estimate the causal effects of the reforms. In this case, there are two different treatments, one for UCH and one for PUC. Thus, each treatment group is composed of one treated unit, in contrast, the control group is composed of all the remaining universities.

Identifying the treatment effects in the DD approach requires the parallel trends assumption to hold. The assumption requires that in the absence of interventions, the difference between the outcomes of the treated and untreated groups remain constant over time. To test whether this case fulfils the assumption, I follow a “placebo strategy” like the one used by De Chaisemartin and d’Haultfoeuille (2020). The test consists of setting a placebo treatments at k periods earlier than the actual treatment and then, estimating the average effects for the periods between the placebo treatment and real treatment using the DD approach (Equation

3.1). If the estimates of the effects are not statistically significant, we can conclude that the assumption is fulfilled.

Then, I use a DD model to estimate the effect of the UCH and PUC policies on female applications and enrolments in their engineering schools. Some notation is helpful in presenting the estimator. Let $Y_{i,t}$ be the percentage of females applicants (or enrolled) in the first year of an engineering programme at university i at time t. The population is observed in several periods pre-treatment, and several periods post-treatment. Let $D_{1,i,t} = 1(D_{2,i,t} = 1)$ denote whether university i was exposed to the treatment 1(2) at period t. $D_{1,i,t} = 0(D_{2,i,t} = 0)$ if not. For $t \geq 2014$, individuals with $D_{i,t} = 1$ are the adopters, and those with $D_{i,t} = 0$ are the non-adopters. Since interventions were implemented in 2014, for $t < 2014$, $D_{i,t} = 0$ for all i. Treatment 1 is the one implemented in UCH and treatment 2 is the one implemented at PUC.

Hence, consider the following regression model:

$$Y_{i,t} = \beta_0 + \gamma_i + \delta_t + \beta_1 D_{1,i,t} + \beta_2 D_{2,i,t} + \epsilon_{i,t} \tag{3.1}$$

Where γ_i are university fixed effects and δ_t are time fixed effects. $\epsilon_{i,t}$ is an error term, and β_1 and β_2 are the causal effects of interest.

As the university is the level of analysis, the sample is small (N=318). Thus, I estimate errors and confidence intervals using *case* bootstrap,[9] with N=1999 random samples from the joint distribution of the terms in the model and the response.

[9] For details on bootstrapping method see Davison et al. (1997).

3.3.1 Results

Before presenting the DD estimates, I show the results of the placebo tests for applications and enrolment in tables 3.3 and 3.4. The tables below show the DD estimates of Equation 3.1 for applications and enrolments for the hypothetic scenario in which PUC and UCH implemented their initiatives $k = 3$ years before the actual year of implementation. Thus, in this estimation, 2007-2009 is the pre-treatment period, and 2010-2013 is the post-treatment period. In Table 3.3, the first column presents the estimates for percentage of female applicants among students who applied to university i as their first choice, second column is for percentage of females among those who applied to university i as their second choice, and in the third column "First or second choice" the estimates is for the pool of students who applied to university i either as first or second choice. Table 3.4 presents estimates for percentage of women among students enrolling in university i. None of the coefficient estimates is statistically significant, suggesting that the parallel-trend assumption is fulfilled for both applications and enrolments.

Table 3.3: Placebo test for applications

	Dependent variable:		
	Percentage of females among applicants		
	First Option	Second Option	First or Second Option
β_1	0.002	0.109	0.003
(UCH)	(0.009)	(0.043)	(0.010)
β_2	0.018	-0.010	0.016
(PUC)	(0.010)	(0.027)	(0.010)
Observations (N)	167	167	167
R^2	0.716	0.451	0.716
Adjusted R^2	0.638	0.300	0.638

Notes: β_1 and β_2 are the DD estimated coefficients as if treatment had happened in 2010. The dependant variable for "First or Second Option" is the percentage of women in each cohort t who ranked a university i in the first or second place in their application list. Bootstrap errors in parenthesis, *p<0.1; **p<0.05; ***p<0.01

Table 3.4: Placebo test for enrolment

	Dependent variable:
	Percentage of females enrolled in first year
β_1	0.001
(UCH)	(0.015)
β_2	0.009
	(0.014)
(PUC)	
Observations (N)	168
R^2	0.699
Adjusted R^2	0.613

Notes: β_1 and β_2 are the DD estimated coefficients as if treatment had happened in 2010. Bootstrap errors in parenthesis, *p<0.1; **p<0.05; ***p<0.01

Applications

I use Equation 3.1 to estimate the effects of both interventions on the share of females among applicants for each engineering school. I look at applications among students who ranked an engineering degree as their first, second option, and either first or second options. Given that the UCH policy requires women to put their engineering school as their first preference to be eligible for a female exclusive slot, we can anticipate a bigger increase in applications as first preference over second.

The first column of Table 3.5 presents the results for share of women applying to university i in the first option. We can see that because of the programme, 10.4% more female students chose engineering at UCH as their first option, whereas PUC experienced a 5.3% increase in the proportion of women that chose their programme as their first option. These results are statistically significant for both universities.

The second and third columns of Table 3.5 present the results for applications in the second option, and in the first and second options. Again, both policies caused an increase in the share of women applying to engineering programmes.

Table 3.5: DD for applications

	Dependent variable:		
	Percentage of females among applicants		
	First Option	Second Option	First or Second Option
β_1	0.104***	0.101***	0.085***
(UCH)	(0.012)	(0.038)	(0.009)
β_2	0.053***	0.051*	0.059***
(PUC)	(0.012)	(0.028)	(0.012)
Observations	318	318	318
R^2	0.747	0.473	0.707
Adjusted R^2	0.708	0.393	0.663

Notes: β_1 and β_2 are the DD estimated coefficients. The dependant variable for "First or Second Option" is the percentage of women in each cohort t who ranked a university i in the first or second place in their application list. Bootstrap errors in parenthesis, *p<0.1; **p<0.05; ***p<0.01

Enrolment

Table 3.6 shows the difference-in-difference (DD) coefficients for UCH (β_1) and PUC (β_2) treatments on enrolments. The first column presents the DD estimates for the percentage of women enrolled in engineering programmes under the regular admission process only. The second column presents DD estimates for all enrollments; thus, in this specification, $Y_{UCH,t}$ considers women enrolled via regular admission and via female-exclusive seats. As the policy implemented in PUC is more subtle, the computation of $Y_{PUC,t}$ is identical in both specifications.

Table 3.6: DD for enrolment

	Dependent variable:	
	Percentage of females enroled in first year	
	Regular admission	All
β_1	0.044**	0.084***
(UCH)	(0.022)	(0.022)
β_2	0.072***	0.072***
(PUC)	(0.022)	(0.022)
Observations	318	318
R^2	0.68	0.700
Adjusted R^2	0.651	0.654

Notes: β_1 and β_2 are the DD estimated coefficients. The dependant variable "regular admission" includes only women enrolled via regualr admission, while the varible "All" includes women enrolled via PEG. Bootstrap errors in parenthesis, *p<0.1; **p<0.05; ***p<0.01

The results indicate that, after controlling for confounding factors, the percentage of females enrolled in UCH through the regular admission process increased by 4.4% as an effect of the gender affirmative action policy; and the total effect (including female-exclusive seats) was 8.4%. In the case of PUC, the increase was 7.7%, which is slightly below the total effect of UCH. Both results are statistically significant as well as economically relevant.

The results show that the UCH and PUC initiatives successfully increased the percentage of women in their engineering programmes. Both universities experienced considerable and statistically significant increases in women's applications and enrolments.

However, these results do not consider the potential spill-over effects of these policies on admissions into engineering programmes from other universities. Spill-overs could act in two possible directions. On the one hand, the policies could have increased enrolments at other universities because students who applied to an engineering programme at UCH or PUC as their first choice but were not accepted probably listed other engineering programmes below in their application ranking. On the other hand, UCH and PUC might be attracting female students who would otherwise have applied for an engineering degree but at a different university. If this were the case, spill-over effects would decrease the number of women enrolling in engineering degrees in the control group. The second case is worrisome because it would mean that the DD estimates are an upper bound of the effect of the policies. However, we can see from the data in Table 3.13 that the proportion of women enrolling in engineering degrees in other universities remained stable before and after the interventions took place.

3.4 Academic ability

Affirmative action policies also have the potential to affect the average academic ability of the incoming cohorts of students. This is particularly true for policies that increase the admission rate of students from minority backgrounds by lowering admission's standards.

The impact that lower average academic ability can have in universities and students is twofold. First, there are "quality effects". Arcidiacono and Lovenheim (2016) argues that more selective institutions have better students, who in turn generate positive peer effects. Thus, lowering admission standards hinders students' ability to benefit from those high-quality peers. Additionally, if more selective universities produce better outcomes for all students than less selective universities, relaxing admission criteria restricts the extent of potential benefits for minority students. Second, there are "match effects", which arise when low-ability students are placed in schools where their academic preparation is significantly below their peers. More selective universities may instruct at a faster pace and assume a certain level of knowledge. Therefore, it may be optimal for some students to attend a less selective university even if a more selective one is in their choice set.

The results from the previous section indicate that both UCH and PUC achieved their goal of increasing the share of women in their student bodies. However, it is not clear yet whether this outcome led to changes in the ability composition. This is particularly relevant for UCH, as their policy involves admitting students who would not have qualified for a place in the school without the programme.

In this section, I explore potential changes in the academic ability distribution of UCH and PUC students using data from the centralised admissions system. The aim is to evaluate whether the increase in the share of women was followed by a decrease in the average academic ability of those new students. To do so, I compare the application scores of men and women before and after the introduction of the policies at UCH and PUC.

The application scores are a linear function of students' national university admission exam scores and their high school records. In the case of UCH, the measure of women's average academic ability considers students who enrolled through the gender policy and those who enrolled through regular admission.

I first show the average application scores of women and men by year; these statistics provide information on the general trend for both institutions. Then, I examine whether the policies were accompanied by changes in the average academic ability of women relative to their male peers by estimating the following equation

$$Y_{j,t} = \beta_0 + \beta_1 R_t + \beta_2 G_j + \beta_3 R_t * G_j + \delta_t + \epsilon_{j,t} \tag{3.2}$$

Where $Y_{j,t}$ is the application score of student j (in time t), R_t is a dummy equal to 1 if the year $t \geq 2014$, G_j is a dummy equal to 1 if student j is a woman and 0 if not.

In this equation, the β_3 coefficient provides the relevant information on differential trends in the application scores of women compared to men. If the β_3 coefficient is positive, it means that after the reform, the average application scores of women increased relative to their male peers. On the contrary, if the coefficient is negative, women's application scores decreased relative to their male peers. If the latter is true, the increase in the share of women caused a decrease in the average ability of the minority group and a reduction in the average academic ability of the whole group.

3.4.1 Results

Figures 3.1 and 3.2 show the application scores quartiles by gender for UCH and PUC respectively. The application scores are on a scale of 0 to 8.5; across the three-year pre-reform plus three-year post-reform period, the application scores of students enrolled in UCH and PUC stayed in the range 7 – 8.5. The UCH average score in the same period was 7.48 and the s.d. was 0.26, while for PUC, the average score was slightly higher, 7.67 and the s.d was 0.23.

From the boxplots, we can see that there was no clear fall in the median application scores or in the minimum scores for women enrolling in UCH and PUC during the post-reform cohorts (2014, 2015, 2016). Thus, women's academic ability (measured by application scores) did not fall after the reform.

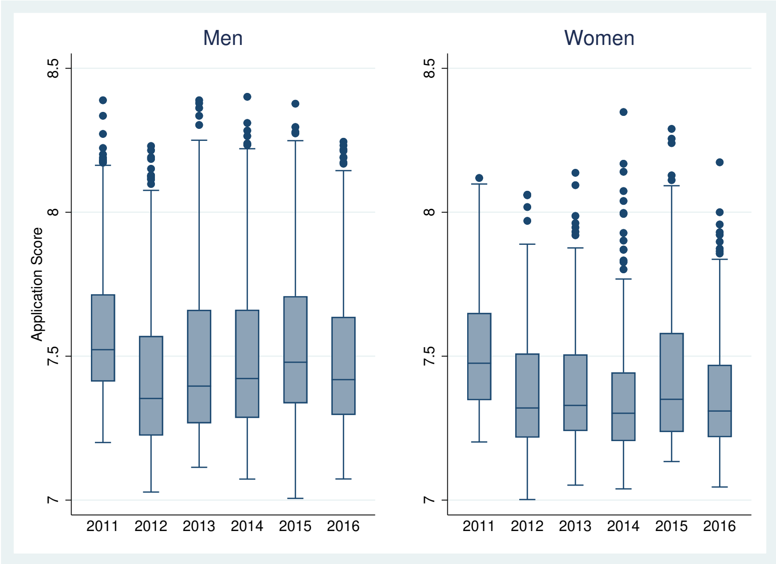


Figure 3.1: Application scores quantiles by gender - UCH

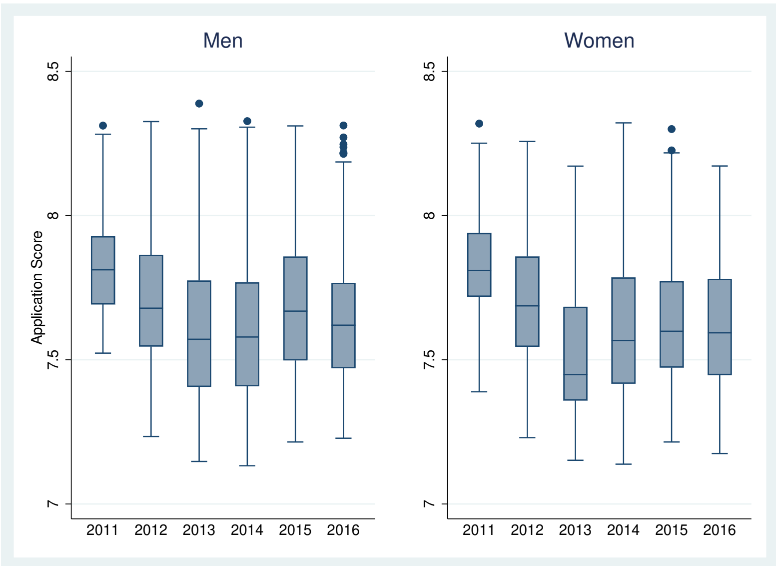


Figure 3.2: Application scores quantiles by gender - PUC

To compare these results to the trends in men's scores, I present the results for the analyses of differential trends in women's and men's average application scores before and after the reform in Table 3.7. The variable of interest is the interaction between the dummy "Women" and the dummy "Reform". The coefficients of the interaction are statistically significant for UCH, suggesting that the policy increased the gender gap in application scores. However, the value of the coefficient is particularly small, reaching approximately 0.06, which is around a quarter of a s.d. of the application score distribution of admitted students. This can probably be explained by the high density of applications around the cut-off score, which means that the first 40 women on the waiting list had scores that were not much lower than the last applicant admitted under regular admission. On the other hand, the coefficients of the interaction for PUC are not statistically significant; therefore, there is no evidence that the reform increased or reduced the gap between men's and women's average application scores.

Table 3.7: Differential trends for average aplication scores

	(1) UCH	(2) PUC
Reform	-0.08 ***	-0.18***
	(-6.65)	(-16.19)
Women	-0.05***	-0.04**
	(-4.52)	(-3.23)
Reform * Women	-0.05***	0.01
	(-3.31)	(0.98)
Constant	7.58***	7.83***
	(855.45)	(1032.99)
N	4145	3711

Notes: Dependent variable is average aplication score per cohort of incoming students at UCH and PUC respectively. Reform is a dummy equal to 1 for post-reform years. t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In sum, there were no significant changes in women's average application scores after the reform, and only small changes in the gender distribution of this ability measure. Changes in the gender composition of new cohorts were not followed by significant changes in the academic ability of incoming students. Thus, increases in the percentage of women stemmed mostly from a higher number of interested, well prepared female applicants rather than from a significant lowering of the admission scores.

3.5 Academic performance

In this section, I use a peer-effects model and data from the University of Chile to determine the effects of classroom gender composition and average academic ability on the academic performance of engineering students. I look at three outcomes: grades on first-year core subjects, grades on a collaborative engineering project, and drop-out rates.

Epple and Romano (2011) argue that "for given educational resources provided to student A, if having student B as a classmate or schoolmate affects the educational outcome of A, then we regard this as a peer effect" (p. 1054). Similarly, Sacerdote (2011) defines peer effects as "any externalities that spill over from peers' or peers' family background or current actions" (p. 250). Thus, I use a peer effect model to measure any externalities that might arise as a result of having more or less female peers.

There are a few challenges when estimating gender peer effects at university level and perhaps the most important of them is the selection problem. This problem occurs when a student who is predictably going to have a certain outcome seeks out, or is assigned to certain peers because of their predicted outcomes. This can severely bias the estimation, making it doubtful whether peer effects exist or not. The data from UCH is particularly helpful to estimate peer-effects because first-year engineering students are split into groups 'classrooms' for all of their teaching, and the allocation to the groups is done quasi-randomly.

The UCH allocation generates idiosyncratic variation in the gender composition (and average academic ability) among groups and makes the estimation of peer-effects plausible. The algorithm is based on application scores, and it attempts to keep a balance of average application scores in each classroom. Still, it generates exogenous variation in gender and any other demographic characteristics of new students. The classroom size is approximately 100 students. If there are a total of eight classrooms in a given year, the algorithm works as follows:

1. The student with the highest score goes to classroom 1
2. The student with the second highest score goes to classroom 2

   ...

8. The student number 8 on the list goes to classroom 8
9. The student number 9 on the list goes to classroom 1

   ...

16. The student number 16 on the list goes to classroom 8
17. The student number 17 on the list goes to classroom 1

    etc.

During the 2009-2017 period there were between seven and eight classrooms per cohort. Table 3.8 shows the average proportion of women per classroom, and the standard deviation of the proportion of women across classrooms for each cohort and for the whole sample.

Table 3.8: Proportion of women by classroom - UCH Engineering School

Year	Proportion of women Mean	s.d.	Classrooms (N)	Students (N)
2009	0.213	0.065	7	945
2010	0.171	0.033	8	999
2011	0.179	0.052	8	789
2012	0.228	0.037	8	767
2013	0.209	0.042	8	794
2014	0.297	0.080	8	820
2015	0.271	0.063	8	875
2016	0.290	0.049	8	957
2017	0.300	0.064	8	902
Total	0.240	0.074	71	7,856

Notes: Mean indicates the average proportion of women by classroom per year. s.d indicates the standard deviation.

All first term courses are compulsory and students in the same group share a physical classroom and attend lectures, seminars and teaching activities together during the first term. Thus, it is reasonable to conclude that the classroom is a relevant peer group for first-year students.

The first-term courses are Algebra, Calculus, Physics, Computer Science and Introduction to engineering. For the first four courses, which I refer to as "core courses", I use the average final grade as the measure of performance. For these courses, students take between 3 and 6 tests for each subject, and all students take these tests simultaneously, regardless of their classroom. Then, teaching assistants grade the tests based on a suggested step-by-step solution. On the other hand, the introduction to engineering course is assessed through a group work project. For this course, professors allocate the students within a classroom to 20 small group of 5 students each. Professors aim to have an even gender distribution among groups. Consequently, in classrooms with less than 20 women, groups are composed of either four men and one woman or five men. Evaluations in this course measure the results of the collaborative work, communication skills, and creativity. Although the proportion of women in each small team is not random, it is still exogenous as students cannot choose their teams.

I use a linear-in-means model, in the style of Manski (1993), to estimate gender peer effects on academic performance. In this model, first-year students' grades are explained by their individual characteristics, their group characteristics, a previous measure of academic ability (individual and group average), and the share of women in the classroom. I use average application scores as a measure of the group's academic ability instead of using the current average outcome of the group as an explanatory variable. The latter option can cause identification problems because, if peers influence a student, they also influence them.

An additional threat to identification in estimating gender peer effects is the inclusion of a gender dummy variable. I avoid this problem by estimating the equation through two separate regressions, one for males and one for females (See section 3.7.3 for a detailed discussion of this identification problem). Estimating the equation separately for males and females is also helpful because the effect of the variable of interest (δ) could differ by gender.

$$y_{ijc} = \alpha + \beta_1 X_{ijc} + \beta_2 \bar{X}_{-i,jc} + \gamma_1(ACA_{ijc}) + \gamma_2 A\bar{C}A_{-i,jc} + \delta P_{j,c} + \epsilon_{ijct} \tag{3.3}$$

- y_{ijct} is the final grade of a male/female student i in classroom j, and cohort c.
- X_i is a vector of other characteristics of student i, such as socioeconomic background and previous school (private, semi-private or public).
- $P_{j,c}$ is the proportion of women in classroom j in cohort c.
- ACA_{ijc} is a vector with admission test scores and high school grades.
- $A\bar{C}A_{-i,jc}$ is the average application score of peers.

3.5.1 Results

Grades

Tables 3.9 and 3.10 report the regression estimates of the effects of the percentage of women by classroom on classmates' grades for each subject. In table 3.9, I present estimates for Physics, Calculus, and Algebra. Columns (1), (3), and (5) show the estimates for women and columns (2), (4) and (6) show the estimates for men. The classroom gender composition has no significant effect on students' grades for any of these subjects.

In table 3.10, I present estimates for Computer Science and Introduction to engineering. Columns (1) and (3) report the estimates for women while columns (2) and (4) report the estimates for men. There are also no significant effects of the percentage of women in the class on computer science grades. In contrast, I find a positive and significant effect on grades in the introduction to engineering course. For women, a one-point increase in the percentage of women in the classroom increases the grades for the introduction to engineering by 0.0165 points. Then, an 8.4% increase, which is equivalent to the intervention effect, would cause an 0.14 increase in women's grades in this course. The increase is small, but statistically significant.

Individual application scores have a positive and significant effect on grades for men and women in every subject. For instance, a 1 point [10] increase in a female student's application score increases her physics grade by 1.839 points.

The classroom average application score is a measure of the group's academic ability. For women, being in a classroom with higher average application scores has a significant effect on reducing their grades in physics, calculus, algebra, and computer science. Meanwhile for men, higher average scores in the classroom either increase their grades or have no significant effect.

The effect of the classroom average application score on women' grades might seem counterintuitive. Students may benefit directly from higher ability classmates through knowledge spill-overs during class, or direct peer-instruction. Additionally, the average classroom ability can affect the overall standard, and students may be motivated to work harder to keep up with their high achieving peers. However, while high ability students can improve their peer's performance in many situations, contest theory suggests that large gaps in skills between individuals can have the perverse effect of reducing effort incentives. Brown (2011) provides empirical evidence for this theoretical prediction by showing that the presence of a superstar in a PGA golf tournament is associated with a lower performance among the other competitors. In a classroom setting, a high achieving classroom environment may negatively impact self-perception. The higher the ability of peers in a classroom, the harder it is to be ranked highly, and thus students might reduce their efforts.

[10] The application scores, mathematics scores and high-school grades are on a scale from 0 to 8.5. The original scale is up to 850, but I divided the scores by one hundred for ease of interpretation.

Moreover, classrooms where the average academic ability is higher might also be more competitive; Gneezy et al. (2003) and Niederle and Vesterlund (2007) show that the gender performance gap is exacerbated under competition.

Table 3.9: Estimated coefficients for grades on Physics, Calculus, and Algebra

	(1) Physics-W	(2) Physics-M	(3) Calculus-W	(4) Calculus-M	(5) Algebra-W	(6) Algebra-M
Percentage of women	0.000874	-0.00368	0.00430	0.00335	0.00250	0.00442
	(0.21)	(-1.48)	(1.11)	(1.45)	(0.59)	(1.80)
Average PSU score - Class	-0.457**	0.131	-0.355*	0.321***	-0.500***	0.130
	(-3.02)	(1.42)	(-2.54)	(4.04)	(-3.62)	(1.65)
Application Score	1.839***	1.327***	1.827***	1.203***	1.925***	1.286***
	(12.58)	(21.03)	(12.95)	(19.52)	(13.75)	(21.73)
Mathematics Score	-0.151	0.124**	-0.139	0.304***	-0.125	0.252***
	(-1.81)	(3.26)	(-1.76)	(8.23)	(-1.58)	(7.05)
High School Grades	0.000735	0.105***	0.00196	0.107***	0.0124	0.114***
	(0.04)	(6.12)	(0.12)	(6.46)	(0.77)	(7.21)
Constant	-4.561***	-7.529***	-5.927***	-10.12***	-5.379***	-8.742***
	(-3.69)	(-10.23)	(-4.95)	(-15.43)	(-4.66)	(-13.69)
Observations	1460	5312	1613	5658	1556	5599

Notes: Dependant variable is average mark in the course, columns names that end in W display estimated coefficients for women, and in M for men. All models include year fixed effects, and controls for socioeconomic background t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.10: Estimated coefficients for grades on Computer Science and Introduction to engineering

	(1) Computer Sc-W	(2) Computer SC-M	(3) Intro Eng-W	(4) Intro Eng-M
Percentage of women	0.00388	-0.00125	0.0165***	0.0134***
	(0.45)	(-0.32)	(5.68)	(8.25)
Average PSU score - Class	-0.539**	0.165	-0.120	0.175
	(-2.72)	(1.52)	(-0.47)	(1.37)
Application Score	2.462***	1.907***	0.421***	0.145***
	(7.64)	(15.19)	(4.90)	(4.00)
Mathematics Score	-0.644***	-0.345***	-0.148**	-0.0271
	(-3.50)	(-4.50)	(-2.96)	(-1.22)
High School Grades	0.0395	0.0145	-0.00450	0.0520***
	(0.34)	(0.35)	(-0.44)	(4.99)
Constant	-4.882**	-7.560***	4.419*	2.997**
	(-2.66)	(-8.17)	(2.29)	(3.11)
Observations	747	3093	1328	4940

Notes: Dependant variable is average mark in the course, columns names that end in W display estimated coefficients for women, and in M for men. All models include year fixed effects, and controls for socioeconomic background t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Collaborative engineering project

In the introduction to engineering course, students in each classroom are divided into 20 groups of around five students each. Professors aim to have at least one-woman per group; this is possible for classrooms with more than 20 women. The variable "low women" is equal to 1 where there are less than 20 women in the classroom and therefore not all groups have the participation of a woman. The variable "High women" is also a dummy variable that is equal to 1 when there are more than 30 women in the classroom, and therefore more than half of the groups have at least 2 women per group. Regression (3) shows that women's grades are higher when they are in a classroom with more than 30 women.

The explanation for the difference between these results and the ones for "core courses" might lie in the collaborative nature of the former. Although testing mechanisms are outside of the scope of this paper, evidence from the literature supports this hypothesis. Firstly, some evidence suggests that women perform better than men in cooperative work (Birch and Ladd (1997), Kolawole (2008)). Thus, an increase in the share of women might improve the results of the team. Secondly, the share of women itself can affect the dynamics of a team in a manner that is consequential to the results of the team. Stoddard et al. (2020) shows that in small groups with only one woman, the "token" woman participates less in group discussions, receives fewer returns on participation when they do, and struggles to convert their performance to influence. This can damage the performance of the team as they might miss valuable input from their female peers.

Table 3.11: Estimated coefficients for grade in introduction to engineering

	(1) Women	(2) Men	(3) Women	(4) Men
Low women	-0.073*	0		
	(-2.22)	(.)		
High women			0.132*	0
			(1.97)	(.)
Average PSU score - Class	-0.133	0.008	-0.259	0.008
	(-0.50)	(0.07)	(-1.01)	(0.07)
Application Score	0.415***	0.141***	0.424***	0.141***
	(4.77)	(3.88)	(4.88)	(3.88)
Mathematics Score	-0.139**	-0.026	-0.146**	-0.026
	(-2.75)	(-1.18)	(-2.88)	(-1.18)
High School Grades	-0.003	0.053***	-0.005	0.053***
	(-0.31)	(5.10)	(-0.49)	(5.10)
Constant	4.830*	4.482***	5.737**	4.482***
	(2.44)	(4.70)	(2.97)	(4.70)
Observations	1328	4940	1328	4940

Variable "low women" is a dummy equal to 1 for classrroms where there is maximum 1 women per team."High women" indicates that more than half of the team in the classroom have a least 2 women in the team. All models include year fixed effects, and controls for socioeconomic background. t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Drop-out

Table 3.12 presents the coefficients of a logit regression on drop-outs after the first term. While the percentage of women in the classroom has no significant effect on men's dropout rate, it does for women. Indeed, a one percent increase in the percentage of women produces a 0.112 decrease in the log-odds of dropping out after the end of the first term. I show predicted probabilities of dropping out by percentage of women in the group on Figure 3.3. The figure helps to understand the magnitude of the effect that the increase in the percentage of women caused by the reform has in reducing drop-out rates. For instance, while the predicted probability of drop-out for a woman in a group with 20% women is 3.8%, this number more than halves for a woman in a group with 28% women (to 1.4%). Thus, given that women also drop out less when they are in classrooms with higher shares of women, the reform can help to increase the persistence of women in the programme.

This finding supports those of Shan (2022), who conducted an experiment in an introductory economics course at a Swiss university and found that women assigned to female minority groups reported significantly less interaction with group peers and dropped out more.[11] Shan (2022) results suggest that isolation might explain the higher drop-out rates for women in highly masculine peer groups.

Table 3.12: Log-odds for drop-out logit

	(1) Women	(2) Men
Women percentage	-0.112**	0.0139
	(-2.87)	(0.70)
Average PSU score - Class	0.0199	0.00337
	(1.28)	(0.60)
Application Score	-0.0324*	-0.0195**
	(-2.05)	(-2.79)
Mathematics Score	0.00527	-0.00162
	(0.63)	(-0.45)
High School Grades	0.00698	0.00365
	(1.33)	(1.64)
Constant	-0.986	6.873
	(-0.07)	(1.38)
Observations	1567	5651

Dependant variable is a Drop-out dummy, equal to 1 if a student does not enrol in any course after the first term. Models include year fixed effects, and controls for socioeconomic background. t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[11] Booth et al. (2018) also provides complementary evidence for the UK, as they find that women in economic undergraduate courses drop-out less when in female-only classes. Additionally, Bostwick and Weinberg (2022) find that women are less likely to drop-out in the first year of a doctoral programme when in groups with more women.

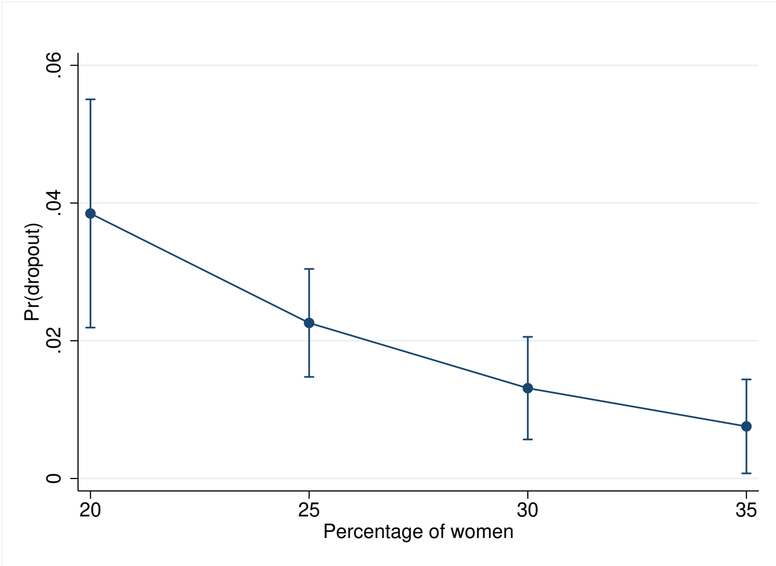


Figure 3.3: Predicted Probabilities of women drop-out

3.6 Conclusions

Although not frequently used, gender affirmative action can help to reduce the existing gender imbalances in STEM fields. This paper analyses the effects of two innovative and distinct policies that aimed to increase women's participation at the UCH and PUC engineering schools. As the universities that implemented the initiatives handle their admissions through the Chilean Centralised Admission system, I can compare their outcomes to the rest of the participating engineering schools. I take advantage of this "natural experiment" and use a difference in difference approach to investigate whether these policies were able to attract more women into their engineering departments and if there were any changes in the academic ability distribution of the new cohorts. In addition, I use data from UCH and a peer effects model to assess the implications for engineering students' academic achievement due to changes in the classrooms' gender composition and average academic ability.

I find that the UCH and PUC initiatives successfully increased the percentage of women in their engineering programmes. Both universities experienced considerable and statistically significant increases in women's applications and enrolments. In the case of UCH, the increase went significantly beyond the 40 seats that were allocated for females only, which proves that the initiative also increased the percentage of women attending their school under regular admission.

The changes in gender composition were not followed by changes in the average academic ability of incoming students. When looking at trends in the gender gap in application scores, I find a slight widening of the gap for UCH students, and no gap for PUC. The results suggest that increases in the percentage of women stemmed from a higher number of interested, well prepared female applicants rather than from a significant owering of the admissions standards.

Moreover, the increase in the share of women had no impact on students' academic achievement in core courses. I only find significant effects of the change in gender composition on performance in the collaborative engineering project, where a higher percentage of women in the relevant peer group increased men's and women's academic performance. The contrast in the results between collaborative and core courses might be explained from the nature of the gender dynamics that arise in team interactions. This theory is supported by (Stoddard et al., 2020) laboratory evidence.

I also find that women drop out less when they are in classrooms with higher shares of women. This last result is especially important for policy makers and university administrators, as it means that this kind of policy can increase female participation, not only by boosting admission rates, but also by increasing women's persistence in the degree programme.

3.7 Appendix

3.7.1 Summary statistics from Chilean centralised admission system data

Table 3.13: Applications and enrolment by gender - Whole sample

Year		Total Applications			Total Enrolment		
		Men	Women	All	Men	Women	All
All Degrees							
2010	N	47,651	49,499	97,150	27,606	24,553	52,159
	%	49.05	50.95	100	52.93	47.07	100
2011	N	46,908	48,628	95,536	27,387	24,050	51,437
	%	49.1	50.9	100	53.24	46.76	100
2012	N	55,184	61,152	116,336	36,963	35,337	72,300
	%	47.44	52.56	100	51.12	48.88	100
2013	N	56,449	61,763	118,212	38,918	37,058	75,976
	%	47.75	52.25	100	51.22	48.78	100
2014	N	56,210	62,951	119,161	39,246	38,043	77,289
	%	47.17	52.83	100	50.78	49.22	100
2015	N	58,859	66,729	125,588	40,384	39,200	79,584
	%	46.87	53.13	100	50.74	49.26	100
2016	N	64,655	77,251	141,906	41,082	41,077	82,159
	%	45.56	54.44	100	50	50	100
Engineering Degrees							
2010	N	17,051	5,487	22,538	12,023	4,254	16,277
	%	75.65	24.35	100	73.86	26.14	100
2011	N	17,002	5,831	22,833	12,089	4,377	16,466
	%	74.46	25.54	100	73.42	26.58	100
2012	N	18,404	6,228	24,632	14,377	5,003	19,380
	%	74.72	25.28	100	74.18	25.82	100
2013	N	20,299	6,993	27,292	15,620	5,592	21,212
	%	74.38	25.62	100	73.64	26.36	100
2014	N	19,833	7,172	27,005	15,670	5,657	21,327
	%	73.44	26.56	100	73.47	26.53	100
2015	N	20,422	7,128	27,550	16,001	5,828	21,829
	%	74.13	25.87	100	73.3	26.7	100
2016	N	21,141	7,419	28,560	15,952	5,916	21,868
	%	74.02	25.98	100	72.95	27.05	100

Notes: Total applications is the number (N) and proportion (%) of students who submitted an application to at least one degree programme, by gender. Total enrolment is the number (N) and proportion (%) of students who enrolled in a degree programme, by gender

Table 3.14: Applications and enrolment by gender - Engineering schools UCH and PUC

Year		**Total Applications**			**Total Enrolment**		
		Men	**Women**	**All**	**Men**	**Women**	**All**
				UCH			
2010	N	1,201	348	1,549	615	119	734
	%	77.53	22.47	100	83.79	16.21	100
2011	N	1,046	298	1,344	609	133	742
	%	77.83	22.17	100	82.08	17.92	100
2012	N	1,212	379	1,591	561	153	714
	%	76.18	23.82	100	78.57	21.43	100
2013	N	1,500	461	1,961	578	137	715
	%	76.49	23.51	100	80.84	19.16	100
2014	N	1,200	551	1,751	550	171	721
	%	68.53	31.47	100	76.28	23.72	100
2015	N	1,267	571	1,838	566	156	722
	%	68.93	31.07	100	78.39	21.61	100
2016	N	1,671	842	2,513	559	164	723
	%	66.49	33.51	100	77.32	22.68	100
				PUC			
2010	N	940	214	1,154	399	87	486
	%	81.46	18.54	100	82.1	17.9	100
2011	N	815	211	1,026	416	100	516
	%	79.43	20.57	100	80.62	19.38	100
2012	N	971	276	1,247	423	93	516
	%	77.87	22.13	100	81.98	18.02	100
2013	N	973	282	1,255	536	130	666
	%	77.53	22.47	100	80.48	19.52	100
2014	N	1,002	337	1,339	503	168	671
	%	74.83	25.17	100	74.96	25.04	100
2015	N	1,012	304	1,316	516	159	675
	%	76.9	23.1	100	76.44	23.56	100
2016	N	1,216	411	1,627	498	169	667
	%	74.74	25.26	100	74.66	25.34	100

Notes: Total applications is the number (N) and proportion (%) of students who submitted an application to UCH and PUC engineering schools, by gender. Total enrolment is the number (N) and proportion (%) of students who enrolled in the UCH and PUC engineering degree programmes, by gender

Table 3.15: Average Applications scores by gender

Year		All Degrees		Engineering All		Engineering UCH		Engineering PUC	
		Men	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
2007	Mean	610.61	605.67	610.43	604.28	744.22	742.58	769.45	770.13
	s.d.	93.07	89.95	90.72	90.21	25.29	23.39	21.26	19.6
2008	Mean	607.38	600.74	608.61	598.65	748.62	747.24	771.47	774.04
	s.d.	98.8	99.7	94.51	99.29	26.65	26.79	22.34	21.37
2009	Mean	612.43	603.29	614.28	600.47	743.46	743.04	766.25	763.77
	s.d.	95.64	99.92	92.21	100.07	23.44	24.37	22.53	22.43
2010	Mean	612.24	605.69	615.38	602.4	752.83	746.4	777.76	777.07
	s.d.	95.33	97	94.08	103.29	23.29	20.35	17.41	16.63
2011	Mean	612.5	605.65	615.43	601.77	757.57	751.52	782.75	781.73
	s.d.	98.43	99.72	98.49	106.74	22.46	21.69	16.21	17.28
2012	Mean	588.97	580.63	595	587.87	740.69	738.34	767.45	762.02
	s.d.	134.29	139	125.64	132.46	40.54	23.32	56.99	82.43
2013	Mean	590.3	582.32	595.74	587.09	747.5	739.71	761.05	751.41
	s.d.	129.75	136.25	124.64	131.04	28.12	21.75	25.04	22.73
2014	Mean	595.7	590.59	603.93	604.34	748.86	742.71	761.85	761.31
	s.d.	127.41	133.43	117.13	118.12	27.04	22.99	24.82	23.41
2015	Mean	599	593.28	607.06	605.23	753.65	750.05	769.06	764.87
	s.d.	126.99	136.05	120.19	123.96	25.94	25.7	23.32	23.08
2016	Mean	586.29	583.67	594.63	589.35	749.13	742.22	764.64	762.46
	s.d.	150.76	155.89	141.09	153.15	25.33	19.78	21.32	22.01
2017	Mean	589.4	586.19	597.42	596.89	745.77	745.37	764.95	759.03
	s.d.	145.13	153.47	136.93	148.63	40.97	26.08	26.29	60.88
Total	Mean	598.54	592.35	604.3	597.67	748.63	744.38	768.28	764.57
	s.d.	123.91	129.59	116.61	123.32	29.14	23.68	28.01	37.64

Notes: Mean and standard deviation (s.d.) of application scores for students who enrolled on: (1) Any degree programme, (2) Any engineering programme, (3) UCH engineering school, (4) PUC engineering school

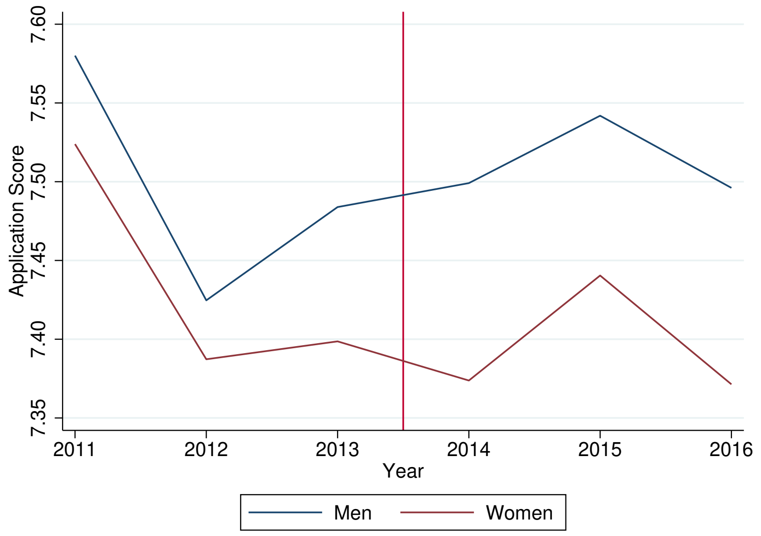


Figure 3.4: Average application score by gender - UCH

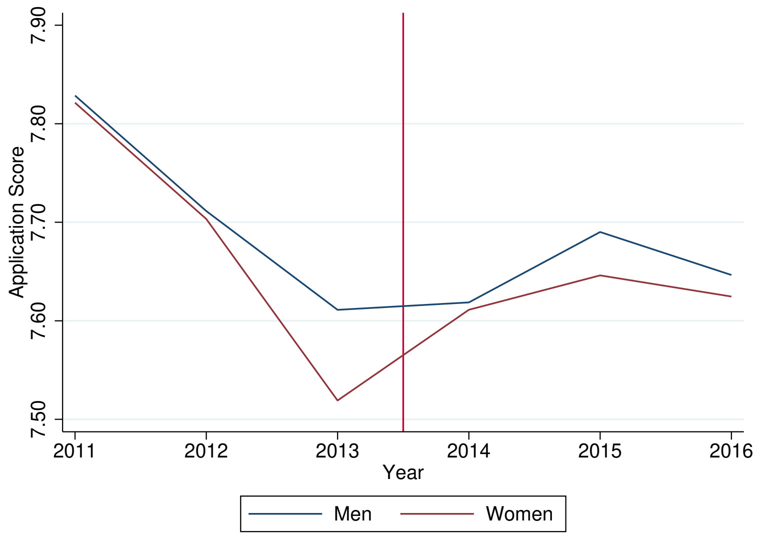


Figure 3.5: Average application score by gender - PUC

3.7.2 Summary statistics from UCH data

Table 3.16: Average Admission scores by subject and gender - UCH

Year		High School Grades			Mathematics Test			Language Test		
		Men	**Women**	**Total**	**Men**	**Women**	**Total**	**Men**	**Women**	**Total**
2009	Mean	702.53	742.61	710.8	775.06	765.68	773.12	697.25	708.26	699.52
	s.d.	57.62	40.92	56.92	38.26	39.76	38.73	54.12	50.61	53.56
2010	Mean	695.45	720.35	699.51	800.15	778.95	796.7	691.2	704.58	693.38
	s.d	68.24	46.18	65.77	35.75	37.67	36.88	55.98	51.49	55.45
2011	Mean	696.48	728.75	702.36	808.29	785.48	804.14	689.22	706.35	692.34
	s.d.	58.51	41.99	57.22	33.88	33.4	34.9	55.34	49.57	54.71
2012	Mean	701.18	734.5	708.45	775.42	757.23	771.45	681.73	703.01	686.37
	s.d.	56.39	42.81	55.43	39.96	37.17	40.05	54.04	51.91	54.26
2013	Mean	713.23	737.75	717.95	771.31	750.46	767.3	695.36	703.22	696.87
	s.d.	53.54	40.87	52.22	39.37	37.59	39.86	56.93	54.63	56.54
2014	Mean	717.43	734.63	721.44	774.16	751.13	768.79	687.44	689.25	687.86
	s.d.	44.77	38.82	44.03	44.5	41.5	44.86	58.83	53.27	57.55
2015	Mean	714.01	737.82	719.07	781.62	755.05	775.97	684.89	698.98	687.88
	s.d.	48.98	43.83	48.88	41.49	44.2	43.43	61.43	55.39	60.43
2016	Mean	724.94	749.85	730.51	768.67	733.74	760.85	689.94	699.68	692.13
	s.d.	45.25	37.32	44.8	38.42	35.52	40.48	57.06	53.66	56.43
2017	Mean	692.68	589.4	663	760.55	736.27	753.57	676.45	685.28	678.99
	s.d.	157.91	308.52	217.29	46.97	42.68	47.05	63.7	59.67	62.66
Total	Mean	**706.17**	**710.92**	**707.18**	**779.96**	**754.69**	**774.54**	**688.05**	**698.52**	**690.3**
	s.d.	**75.39**	**144.82**	**94.67**	**42.6**	**42.51**	**43.82**	**57.92**	**54.38**	**57.33**

Notes: The admission system uses a linear transformation to convert High schools grades into scores in the 150 to 850 range. Mathematics test and Language test are sections of the university admission tests (PSU) and the scale is 150-850.

Table 3.17: School of origin - UCH

Year		**Public**	**Semi-Private**	**Private**	**Total**
2009	N	153	157	286	596
	%	25.67	26.34	48	100
2010	N	154	153	362	669
	%	23.02	22.87	54.1	100
2011	N	185	159	348	692
	%	26.73	22.98	50.3	100
2012	N	143	167	314	624
	%	22.92	26.76	50.3	100
2013	N	144	191	289	624
	%	23.08	30.61	46.3	100
2014	N	157	166	294	617
	%	25.45	26.9	47.7	100
2015	N	189	211	287	687
	%	27.51	30.71	41.8	100
2016	N	165	231	283	679
	%	24.3	34.02	41.7	100
2017	N	180	273	316	769
	%	23.41	35.5	41.1	100
Total	N	1,470	1,708	2,779	5,957
	%	24.68	28.67	46.7	100

Notes: Number (N) and percentage (%) of first year UCH engineering students who graduated from a Public, Semi-private (partially financed by government), or Private school.

3.7.3 Identification of the gender composition effect

The first approach to evaluate gender peer effects would be a basic version of the Manski's linear-in-means modified to evaluate gender peer-effects, in this specificitation y_{ij} is an scalar outcome for a student i in classroom j (e.g. first-year student's Calculus grades). $\bar{y}_j = E(y|j)$, i.e. the mean outcome of students in classroom j. x_i is a gender dummy, and P_j is the proportion of females in classroom j. ϵ_{ij} are not observed attributes that affect y_{ij}. Following Manski (1993):

$$y_{ij} = \alpha + \gamma \bar{y}_j + \beta x_i + \delta P_j + \epsilon_{ij},\ E(\epsilon_{ij}|j, x_i) = j'\sigma \tag{3.4}$$

where $(\alpha, \gamma, \beta, \delta, \sigma)$ is a parameter vector. Here γ is the endogenous effect, β is the difference in males and females achievement, δ is the effect of having more females in the classroom, and σ captures the correlated effect.

The mean regression of y on (x_i, j) has the following form :

$$E(y|x_i, j) = \alpha + \gamma E(\bar{y}_j|x_i, j) + \beta E(x_i|x_i, j) + \delta E(P_j|x_i, j) + j'\sigma \tag{3.5}$$

but, $E(x_i|x_i, j) = P_j$ and $E(P_j|x_i, j) = P_j$, then:

$$\bar{y}_j = \alpha + \gamma \bar{y}_j + \beta P_j + \delta P_j + j'\sigma \tag{3.6}$$

Provided that $\gamma \neq 1$:

$$\bar{y}_j = \frac{\alpha}{1-\gamma} + \frac{\beta + \delta}{1-\gamma} P_j + j' \frac{\sigma}{1-\gamma} \tag{3.7}$$

Sustituting (5) in (2) gives the reduced form:

$$y_{ij} = \frac{1}{1-\gamma}\alpha + \beta x_i + ((\frac{1}{1-\gamma} - 1)\beta + \frac{1}{1-\gamma}\delta)P_j + \frac{1}{1-\gamma} j'\sigma$$

Since there is no self-selection, i.e., no correlated effects, one could impose $\sigma = 0$. However, the parameter of interest δ still cannot be identified. Instead of estimating δ, the coefficient that would be estimated through this equation is $((\frac{1}{1-\gamma} - 1)\beta + \frac{1}{1-\gamma}\delta)$. Therefore the Identification of the gender composition effect requires replacing $\bar{y}_j$ for a previous measure of academic performance and separete estimation of the regression for men and women.

3.7.4 Other estimations: Logit, binary outcome variable, pass or fail

Table 3.18: Log-odds for "pass" logit (1)

	(1) Physics-W	(2) Physics-M	(3) Calculus-W	(4) Calculus-M	(5) Algebra-W	(6) Algebra-M
pass						
Percentage of women	2.680	-0.514	0.148	1.366	1.156	2.370**
	(1.51)	(-0.50)	(0.12)	(1.84)	(0.77)	(2.81)
Average PSU score - Class	-1.571*	-0.199	-0.924	0.469*	-1.538**	-0.140
	(-2.45)	(-0.61)	(-1.78)	(2.02)	(-2.84)	(-0.54)
Application Score	5.352***	2.383***	3.461***	1.670***	3.975***	1.810***
	(6.63)	(6.71)	(6.52)	(7.12)	(6.53)	(7.43)
Mathematics Score	-0.382	0.312	-0.0683	0.817***	-0.00851	0.816***
	(-1.16)	(1.92)	(-0.28)	(6.85)	(-0.03)	(6.63)
High School Grades	-0.0382	0.221***	0.0145	0.240***	0.0533	0.246***
	(-0.50)	(3.52)	(0.32)	(4.83)	(1.06)	(4.96)
Public	0.0336	0.183	-0.182	0.273**	-0.147	0.463***
	(0.16)	(1.43)	(-1.17)	(2.80)	(-0.86)	(4.37)
Semiprivate	0.0452	-0.342**	-0.171	-0.154	-0.0661	-0.0688
	(0.22)	(-3.12)	(-1.16)	(-1.82)	(-0.40)	(-0.78)
Constant	-22.99***	-17.63***	-17.76***	-22.67***	-16.95***	-19.00***
	(-3.70)	(-5.83)	(-3.77)	(-10.39)	(-3.35)	(-8.12)
Observations	1473	5356	1627	5706	1567	5651

Notes: Dependant variable equals to 1 if student i passed the class. Columns names that end in W display estimated for women (and M for men). t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.19: Log-odds for "pass" logit (2)

	(1) Computer Sc-W	(2) Computer SC-M	(3) Intro Eng-W	(4) Intro Eng-M
Percentage of women	4.091	1.375	18.05	-5.971
	(1.46)	(0.98)	(1.63)	(-1.14)
Average PSU score - Class	-1.330	-0.0991	-6.179	-6.161
	(-1.68)	(-0.29)	(-0.39)	(-1.53)
Application Score	5.333***	3.261***	4.315	3.011
	(4.41)	(6.19)	(1.21)	(1.69)
Mathematics Score	-1.394*	-0.231	-0.225	-0.382
	(-2.31)	(-0.86)	(-0.14)	(-0.41)
High School Grades	-0.215	0.113	0.291	-0.721
	(-0.57)	(0.78)	(0.70)	(-1.18)
Public	-0.409	0.116	0.816	0.243
	(-1.51)	(0.74)	(0.66)	(0.41)
Semiprivate	-0.144	-0.119	0.267	-0.218
	(-0.52)	(-0.83)	(0.27)	(-0.48)
Constant	-16.67*	-20.60***	14.71	37.51
	(-2.12)	(-5.76)	(0.12)	(1.22)
Observations	753	3118	766	4972

Notes: Dependant variable equals to 1 if student i passed the class. Columns names that end in W display estimated for women (and M for men). t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.7.5 Other estimations: Heteregeneous effects

Table 3.20: Estimated coefficients of heterogenous model for grades - Men

	(1) Physics-M	(2) Calculus-M	(3) Algebra-M	(4) Computer SC-M	(5) Intro Eng-M
Percentage of women	0.00256	0.00626	0.00916*	0.00683	0.0150***
	(0.63)	(1.81)	(2.45)	(0.89)	(5.62)
Application Score - Q2	0.254*	0.195	0.190	0.346	0.0752
	(2.08)	(1.71)	(1.62)	(1.68)	(0.97)
Application Score - Q3	0.556***	0.364**	0.548***	0.717***	0.0848
	(4.59)	(3.22)	(4.61)	(3.43)	(1.10)
Application Score - Q4	1.092***	0.884***	1.120***	1.305***	0.147
	(8.67)	(7.28)	(8.89)	(5.92)	(1.84)
Application Score - Q2$\times$ Percentage of women	-0.00434	-0.00259	-0.000776	-0.00660	-0.00250
	(-0.85)	(-0.56)	(-0.16)	(-0.66)	(-0.75)
Application Score - Q3 $\times$ Percentage of women	-0.00987	-0.00393	-0.00808	-0.0100	-0.00202
	(-1.94)	(-0.86)	(-1.64)	(-0.97)	(-0.61)
Application Score - Q4$\times$ Percentage of women	-0.0135**	-0.00814	-0.0147**	-0.0160	-0.00222
	(-2.60)	(-1.67)	(-2.83)	(-1.53)	(-0.66)
Average PSU score - Class	0.00150	0.00339***	0.00149	0.00207	0.00172
	(1.61)	(4.20)	(1.88)	(1.86)	(1.34)
Mathematics Score	0.00224***	0.00401***	0.00324***	-0.000812	-0.000224
	(6.05)	(11.07)	(9.06)	(-1.15)	(-1.03)
High School Grades	0.00149***	0.00149***	0.00146***	0.00143***	0.000551***
	(8.91)	(9.15)	(9.29)	(3.72)	(5.40)
Public	0.141***	0.185***	0.238***	0.103*	0.0381*
	(4.91)	(6.51)	(8.33)	(2.34)	(2.26)
Semiprivate	-0.122***	-0.0247	-0.000771	-0.101*	0.0198
	(-4.44)	(-0.92)	(-0.03)	(-2.37)	(1.23)
Constant	0.720	-2.641***	-0.471	2.865*	3.974***
	(0.93)	(-3.80)	(-0.70)	(2.52)	(4.11)
Observations	5312	5658	5599	3093	4940

Notes: Includes year fixed effects, t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.21: Estimated coefficients of heterogenous model for grades - Women

	(1) Physics-W	(2) Calculus-W	(3) Algebra-W	(4) Computer Sc-W	(5) Intro Eng-W
Percentage of women	0.00256	0.00734	0.00264	0.0198	0.0215***
	(0.43)	(1.38)	(0.44)	(1.47)	(5.21)
Application Score - Q2	0.242	0.304	0.120	1.101**	0.272*
	(1.16)	(1.62)	(0.60)	(2.75)	(1.98)
Application Score - Q3	0.350	0.420*	0.251	0.734	0.292*
	(1.53)	(1.99)	(1.11)	(1.62)	(2.00)
Application Score - Q4	1.094***	0.956***	1.058***	1.370**	0.355*
	(4.52)	(4.13)	(4.40)	(2.68)	(2.39)
Application Score - Q2 × Percentage of women	-0.00560	-0.00720	0.00107	-0.0412*	-0.00891
	(-0.69)	(-1.05)	(0.14)	(-2.24)	(-1.61)
Application Score - Q3 × Percentage of women	0.00279	-0.00190	0.00386	-0.0110	-0.00845
	(0.31)	(-0.24)	(0.44)	(-0.52)	(-1.43)
Application Score - Q4 × Percentage of women	-0.00889	-0.00578	-0.00581	-0.0152	-0.00579
	(-0.96)	(-0.66)	(-0.63)	(-0.64)	(-1.00)
Average PSU score - Class	-0.00395*	-0.00270	-0.00406**	-0.00479*	-0.00113
	(-2.54)	(-1.88)	(-2.86)	(-2.37)	(-0.44)
Mathematics Score	0.000976	0.00144	0.00149*	-0.00169	-0.000888
	(1.24)	(1.91)	(1.98)	(-0.99)	(-1.89)
High School Grades	0.000515**	0.000538***	0.000647***	0.00298**	0.0000785
	(2.98)	(3.34)	(4.01)	(2.72)	(0.76)
Public	-0.0167	-0.0633	-0.0197	-0.0610	0.0297
	(-0.31)	(-1.24)	(-0.38)	(-0.70)	(0.94)
Semiprivate	-0.0174	-0.0456	-0.0351	-0.0759	0.0288
	(-0.34)	(-0.94)	(-0.72)	(-0.88)	(0.96)
Constant	6.049***	4.133***	5.458***	6.739**	6.761***
	(4.47)	(3.30)	(4.45)	(2.78)	(3.49)
Observations	1460	1613	1556	747	1328

Notes: Includes year fixed effects, t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figures 3.6 and 3.7 show the average marginal effect of the percentage of women for students in each quantile of the application scores distribution. (The regression coefficients are presented in tables 3.20 and 3.21)

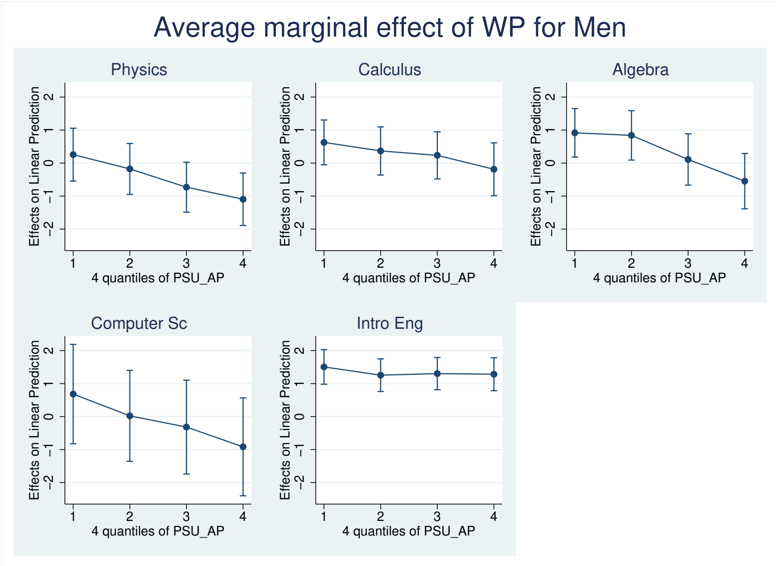


Figure 3.6: Marginal effect of women percentage on grades by application scores quantiles - Men

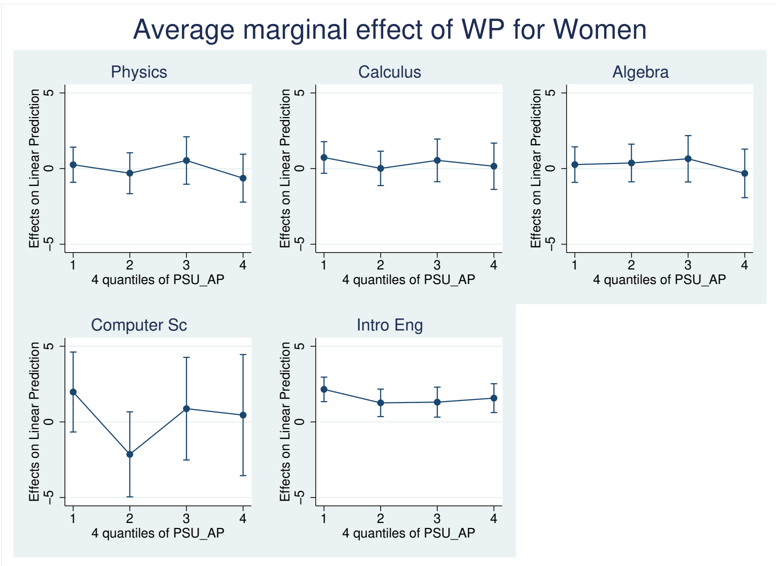


Figure 3.7: Marginal effect of women percentage on grades by application scores quantiles - Women

For women, the effect of the percentage of women on grades does not seem to vary across different levels of previous academic achievement. For men, there is a variation in physics, where students with lower application scores seem to have higher grades when they are in classrooms with a higher percentage of women.

Chapter 4

Effects of team diversity on performance, perceptions, and predictions: Experimental evidence of gender composition and language

Introduction

In recent decades, women's and immigrants' participation in higher education and labour markets has increased. In 2019, OECD countries received on average eight new migrants per thousand inhabitants, and on average, international students accounted for 13% of all enrolments in master's programmes and 22% of PhD enrolments OECD (2020b).According to UNESCO, women's participation rate in tertiary education almost doubled from 2000 and 2014. Similarly, in the labour force, the ratio of female to male participation in OECD countries has increased from 70% in 2000, to 77% in 2021 (World Bank, 2022).

Diversity's individual- and group-level effects could puzzlingly point in mixed directions. For firms, increases in gender, cultural, and linguistic diversity might bring the benefits of a broader range of views and skills, or they might generate communication difficulties and other forms of friction (Lazear, 1999). Likewise for university students, similar positive externalities might arise from exposure to a diverse student body, or it might disturb instruction pace and flow in ways that hinder student academic achievement (Diette and Uwaifo Oyelere, 2012).

This paper aims to shed light on that puzzle by estimating the causal effect of a team's gender- and linguistic diversity on students' performance, perceptions, and predictions. We randomize the gender- and linguistic composition of small teams of postgraduate students at a top UK University, and we estimate diversity's impact on individual performance with respect to academic outcomes; we estimate its impact on perceptions of dynamics of group deliberation like 'voice' and 'leadership'; and we estimate students' predictions of their own academic promise.

The study took place in two consecutive iterations of a compulsory one-term course with an annual cohort of around 180 Master's students (average age 25). We administered the study in two cohorts in 2020-2021 and 2021-2022). 80% of the students in each cohort were women and 49% were non-native English speakers. In this course, university officials assigned students into seminar groups (called seminars) in a way that was exogenous to their demographic characteristics. After this first random assignment, within each seminar group, we then randomised students a second time into smaller teams. We combine three data sources. First, we use administrative data with rich information on detailed student background from admission records, course selection, marks, course teachers's characteristics (gender and native English speaker status) and academic advisor. Second, for each cohort we collected survey data before the course commenced and again at its conclusion. Third, we gathered individual-level data on course performance from an end-of-term examination that accounted for 100% of the course grade.

Our answers to the research questions capitalize on three features of the study setting. First, there was no self-selection into the seminars and teams that we used to estimate compositional effects. Consequently, there is exogenous variation in the gender composition and in the distribution of native versus non-native English speakers (as well as any other demographic characteristics of students across groups). Therefore, our estimates of the effects of gender composition and the share of native speakers can be interpreted as causal.

Second, the course enrols a high percentage of international students. UK institutions are attractive to international students looking to complete or continue studies in an English-speaking country. According to OECD (2020b), The UK is the second receiving country of international students, only after the US. This course is not an exception; around 80% of the students are foreign-born, and around half are non-native English speakers. Thus, the setting maximized interactions between students with different language backgrounds.

Finally, the postgraduate programme and the course boast around 80% female participation. This allows us to study the effects of gender composition in teams in a female-majority setting. Previous literature has focused on the effects of gender composition on interactions and outcomes in male-dominated settings (i.e., STEM fields).[1] This is often done to understand the causes of low representation of women in these fields which pay on average higher wages than female-dominated fields. It is hypothesized that a high fraction of men in fields where women are a minority discourage other women from chosing these fields, and hence lead to a low representation of women. However, studying the effect of gender composition in settings where women are the majority is equally important. For example, these settings can help to clarify if findings of studies in male-dominated areas are explained by men's attitudes toward women or attitudes of a majority toward a minority regardless of the dominating sex.

We present estimates for three sets of outcomes. First, we report performance estimates for overall course mark. We find that, on average, native English speakers

[1] For instance, Oosterbeek and Van Ewijk (2014), and Shan (2022) study gender per effects for economics undergraduate students, where women constitue around 30% of the student body. Stoddard et al. (2020) study gender dynamics among students from an accounting undergraduate programme with a similar percentage of women participation.

perform around 4% better than non-native English speakers even after controlling for other factors. However, we do not find evidence that the share of non-native speakers in the group affects individuals' exams marks. Regarding gender, we detect a statistically significant effect for non-native speakers of an increase in the share of female peers on exam marks, equal to 2% higher for 1 more woman in the team. Second, we report perception estimates for measures of team dynamics. We find that non-native speakers feel less heard when in groups with more native speakers, as one more native speaker in the team causes a 2 points (out of 10) decrease in non-natives' self-assessment of the influence of their voice. We also observe that an increase in the proportion of women in the seminar group causes an increase in women's perception of being heard in their group. The effect is equal to 1 point (out of 10) for each additional woman in the seminar group. In terms of leadership, we find that men are more likely to consider themselves as leaders of their team when in groups with more women, which shows that men do not suffer negative effects from being in a gender minority. Finally, we present prediction estimates for changes in students' expectations of their own future performance. We find that students in groups with more native speakers lower their expectations. They become more pessimistic about their future academic achievement.

The findings are relevant for universities and employers who deal with internationalisation, diversity, and inclusion in the workplace in settings where women are the majority. Non-native speakers benefit from having more diverse peers, and native speakers are not harmed by increased shares of peers with different linguistic backgrounds. Men and women benefit from having more female peers even in female-majority settings. This evidence supports policies that promote internationalisation, and helps inform course designs for academics and teams' formation for employers.

This paper contributes in various ways to a growing literature on diversity in educational settings and in teams at the workplace. In particular, to the strands that investigate the effects of gender composition and presence of non-native speakers. In educational settings, previous research has focus mainly on primary and secondary education. At this level, the evidence on the effect of diversity in language and country of origin on pupils outcomes is mixed. For instance, in the USA, Diette and Uwaifo Oyelere (2012) find that the effect of non-native speakers on the native peers are heterogenous. They report small positive effect for the native students in the bottom and middle part of the achievement distribution, and small negative effects among those at the top. In contrast, Geay et al. (2013) finds no negative effect of presence of non-native English speakers in UK schools.[2] In terms of gender diversity, there is overall agreement in the literature that there are positive effects of higher shares of girls in the classroom on achievement of boys and girls.[3] However, much of these findings from primary and secondary education might not be relevant in settings with adults students.

Only a few studies investigate the role of a diverse environment in higher education settings in educational attainment. These studies have mostly focused on larger peer groups rather than in small teams. For instance, Braakmann and McDonald (2018) focus on the diversity among undergraduate students using the entire university cohort as the relevant peer group, and find heterogenoeus results. Treating the entire classroom as the peer group, Chevalier et al. (2020) study the effect of ethno-linguistic composition of a classroom on academic outcomes for undergraduate students in the UK. They find that Non-native speaker's benefit from greater linguistic diversity. Oosterbeek and Van Ewijk (2014) find no significant effect of the gender of peers at the classroom level on academic outcomes. Our study is the first to analyse the effects of peers' diversity in both small groups (e.g., 4-5

[2] Several studies investigate the effect of the share of immigrant children in students' academic outcomes. The results are also mixed. For instance, Gould et al. (2009) and Jensen and Rasmussen (2011) find evidence of negative effects of the share of immigrants in classrooms with data from Israel and Denmark respectively. Meanwhile, studying the impact of immigration in Austrian schools, Schneeweis (2015) finds negative effects on the academic outcomes of migrants, but no effect on native students.

[3] Lavy and Schlosser (2011) and Hoxby (2000) have studied the effect of peer's gender, and peer's gender and race (respectively) on pupil's academic outcomes. Both studies find that student's marks increase with the proportion of girls in the classroom. Some studies have explored longer term outcomes. For instance, Schneeweis and Zweimüller (2012), and Anelli and Peri (2019) have studied the effect of gender of school peers on student's choices of field of study. Similarly, Black et al. (2013) study the role of high school peers, particularly in terms of peer's gender and social class in determining student's future labour market, and other longer run outcomes. They find heterogeneous results, while teenage girls tend to benefit from higher proportion of girls, boys do not.

students) and middle-size groups (e.g., 8 to 16) as well as among adult students at the postgraduate level.

This is important because, by studying peer effects in small groups of postgraduate students, we can learn about team dynamics, and gain insights into the role of diversity along gender and native English versus non native English speakers in the workplace, where causal evidence is scarce. Apart from Battaglini et al. (2022), that studies the effect of working with female colleagues on federal judges' decisions of hiring female law clerks, the few causal evidence on team diversity comes from laboratory settings with undergraduate students.[4]

4.1 Background

4.1.1 Course setting

The experiment was nested in two consecutive iterations of an eleven-week postgraduate course module in the LSE's Department of Social Policy across the 2020 and 2021 academic years. The module, *Understanding Policy Research (SP401)*, is mandatory for all students who enroll on the one-year interdisciplinary MSc programme in International Social and Public Policy. The Master's programme attracts an internationally diverse student profile to study "how states and societies respond to global challenges of social, demographic and economic change, and of poverty, migration and globalisation." The SP401 curriculum equips students with the tools to evaluate applied policy research critically, through a grounding in concepts that draw on qualitative and quantitative research, the policy process, and applied social policy problems.

SP401 coursework comprised three components, each of which was repeated weekly

[4] For instance, Hoogendoorn and Van Praag (2012), and Hoogendoorn et al. (2013) study diversity in teams consisting of undergraduate students in business studies that start up a venture as part of their assignments. They study etchnic and gender diversity on teams performance on business sales and profits.

throughout the MSc programme's first term. First, students viewed a pre-recorded lecture that instructors uploaded online at the start of each week.[5] Second, students completed a team-based activity in their own time. Third, students undertook additional activities during a weekly in-person seminar meeting, whereupon their team convened with one to three other teams. Lecture topics ranged from introductions to quantitative and qualitative research on one hand to substantive overviews of key research issues in policy analysis and policy-making on the other. Seminar activities required students to apply concepts from the week's lecture to specific policy problems.

4.1.2 Data Collection

The study draws on three sets of data. First, we gathered administrative data from the university's Registrar on student characteristics. Those data encompassed self-report information on gender and language, as well as other demographic characteristics and admissions information.

Second, students completed two surveys. Students completed a baseline self-report questionnaire in the first week of the term. The baseline survey contained items relating to respondents' prior methodological training, their expectations for their own future performance in the course, and their self-assessed approaches to navigating professional group dynamics. Second, students completed a questionnaire in the term's final week. The endline survey duplicated items from the intake survey relating to expectations about performance in the course in particular and in the MSc in general. The endline survey also contained items that captured respondents' reflections about the group-work to which they had contributed, including self-assessments of their own sense of whether they had influenced the group-work.

Third, students completed a take-home online examination administered roughly one month after instruction concluded. The exam included many short questions, for which the expected answer ranged from a few words to a paragraph. The exam

[5] Pre-recorded lectures were implemented as a response to the COVID-19 pandemic.

posed questions from throughout the term's materials and mimicked the format of questions that students had completed in their study groups and seminar meetings. The exam mark accounted for 100% percent of final mark on the course. However, even though the course was mandatory, a fail did not automatically lead to non-completion of the MSc. Still, stakes are high, because if the exam mark fell into the bad fail category, the student may retake the exam the following year, which delays graduation in at least a year.

4.2 Data

4.2.1 Administrative data

Table 4.1 presents summary statistics for the administrative data for both cohorts of students. Panel 1 of Table 4.1 shows that students' average age is 24.6, 79.5% of students are women, and 51% of the students are non-native speakers. Panel (2) presents statistics about students' previous academic backgrounds. Most students have a Bachelor's degree as their highest qualification, and 18% of students hold a prior Master's degree before starting the MSc. Panel 2 also shows that 33% of students have either a completed or are pending qualification from a university in the United Kingdom. The bottom panel of Table 4.1 presents average mark for women, men, native speakers and non-native speakers. Men and women performed similarly (around 68/100); however, there were significant differences by native language. Native English speakers earned an average mark of 73 while non-native speakers earned an average mark of 65. Marks corresponded to the following classifications: marks in the 0-39 range equated to a "Bad Fail", 40-49 was a "Fail", 50-59 was a "Pass", 60-69 was a "Merit", and 70-100 a "Distinction". The bottom of panel (3) presents statistics of the proportion of student falling into each category by gender and native language.

Table 4.1: Summary statistics: Administrative data

		All	Men	Women	Non-native Speakers	Native Speakers*
(1) Demographic characteristics						
Proportion (%)		**-**	20.5	79.5	51	49
Total (N)		**376**	77	299	180	173
Age	Mean	**24.6**	25.4	24.4	25.3	24.1
	S.d	**3.9**	4.4	3.7	4.6	2.9
(2) Prior studies						
Highest qualification	Bachelor (%)	**81.9**	83.1	81.9	75.4	89.0
	Master (%)	**17.6**	16.9	18.1	24.6	11.0
	Yes (%)	**33.4**	37.6	32.1	22.2	42.2
Studied in United Kingdom	***No (%)***	**66.6**	62.3	67.9	77.8	57.8
	S.d	**3.9**	4.4	3.7	4.6	2.9
(3) Academic outcomes						
Mark	Mean		68.9	70.4	66.7	72.5
	s.d		13.2	12.6	12.8	12.1
	Bad fail (%)	**2.4**	3	2	2.8	1.2
	Fail (%)	**4.3**	4	4	3.3	2.9
Classification	Pass (%)	**14.6**	8	16	21.0	9.3
	Merit (%)	**24.7**	26	24	29.8	19.8
	Distinction (%)	**54.0**	60	53	43.1	66.9

Notes: *The LSE administrative dataset does not contain information on student's native language. Thus, summary statistics presented in the last two columns correspond to the sample of students who answer the baseline survey.

4.2.2 Survey data

4.2 presents summary statistics of the data we collected through the two surveys.

Baseline

The baseline survey asked questions about student's native language, usual role in group work, familiarity with relevant subjects, and their expectations of final mark on the course. In the top panel of 4.2, we present statistics for self-perception of leadership. Through the survey students reported what role described them best when working in groups. We provided four options, and included a brief description of what each role entailed. We were particularly interested in variation among students in their self-perception of leadership skills as these are linked to better labour market outcomes (see for example, Kuhn and Weinberger (2005)). Therefore, we derive the binary variable "Leadership role" which is equal to 1 if the student reported to be best described by the leadership role, and 0 if they chose any other alternative. About 30% of the students considered themselves leaders before starting the course. The baseline measure allowed us to estimate the course's effect on changes to students' perception of themselves.

Additionally, for non-native English speakers, we derived a variable that indicated the "distance" between the student's native language (as they reported in the baseline survey) and the English language. The variable allow us to capture heterogeneity between non native English speakers. We use Chiswick and Miller (2005)'s measure of linguistic distance which ranges from 1 to 3, in 0.25 increments, with three being the most similar to English. Among the sample, there was substantial diversity in students' native languages; the language scores spanned the whole range from 1 to to 3, with a mean of 1.9.[6]

Then, we present statistics of two variables that contain information about familiarity with relevant subjects and mean and standard deviation of expected final mark on the course. Most students reported some experience with research methods; on average, they were more familiar with qualitative than quantitative methods.

End of year survey

The end of year survey asked students to reflect on team dynamics, students' views of their future interactions in teams, and academic performance expectations. The bottom panel of Table 4.2 presents the mean and standard deviation by group for relevant end term survey data.

At the end of the course, we asked students three questions related to their perception of their "voice" in team interactions. Survey items probed student's level

[6] See Appendix 4.6.1 for details on student's country of origin, native language, and their correspondant measure of linguistic distance to English

of agreement with the statement *"My voice was heard during group discussions"*, and with two follow up items: *"Working in teams for SP401 made me more confident than before in voicing my view in future interactions"*, and *"Working in teams for SP401 made me more confident than before that my view will be heard in future interactions"*. Table 4.2 shows that most students in general agreed with the first statement, and both women and native speakers in particular were more likely than their counterparts to have agreed. For the two statements about confidence in future interactions, the average response was around 5 (neither agree nor disagree). Again, women and native speakers responded above 5 on average, which indicates that these students tended to agree more with the two follow up questions.

Table 4.2 also presents statistics for two variables related to leadership. The first is a binary variable that we constructed using data from the end of year survey item that collected information on each student's perception of their role in their team. As for the baseline leadership variable, the variable took the value 1 if the student answered Manager/Leader and zero otherwise. We also used this leadership variable in combination with the baseline leadership variable to measure changes in students' assessment of their role in teams. Table 4.2's bottom panel displays that around 30% of students reported that they had a leadership role in their teams, which is consistent with what they reported in the baseline survey. However, we observed within group-differences across baseline and end of term team roles. While 20% of men reported taking leadership roles at the start of the course, 30% reported the same at the end of the term. The opposite was true for non-native English speakers, who were more likely to report that they identified with the leader role in teams before the start of the course. The second leadership variable was constructed using peers' input. For the 2020-2021 cohort, we asked students to match each of the roles with one or more team members. Then, we constructed a binary variable equal to 1 if a student was mentioned as taking the Manager/Leader role by at least one other team member and zero otherwise. 30% of students were nominated at least once as the leader of the group by their peers.

Finally, Table 4.2 also presents the average expected mark at the end of the term.

At this time, average expected marks are lower than at the start of the course for all groups.

Table 4.2: Summary statistics: Surveys

Variable		**All**	**Men**	**Women**	**Non-native Speakers**	**Native Speakers**
(1) Baseline Survey						
Leadership role (0 if no, 1 if yes) *	Mean	0.3	0.2	0.4	0.3	0.4
	S.d.	0.5	0.4	0.5	0.4	0.5
	N	182	36	146	84	98
Language Score (1-3)	Mean	-	1.9	1.9	1.9	-
	S.d.	-	0.6	0.5	0.5	-
Familiarity with qualitative research methods (0-10)	Mean	5.9	5.6	5.9	5.8	6.0
	S.d.	3.4	2.1	2.5	2.4	2.5
Familiarity with quantitative research methods (0-10)	Mean	4.1	4.2	4.1	4.1	4.2
	S.d.	2.2	2.2	2.2	2.3	2.2
Expected Mark (0-100)	Mean	72.6	72.0	72.8	73.7	71.4
	S.d.	12.6	13.1	12.4	12.7	12.3
	N	355	71	284	182	173
Average response rate	%	93.9	92.2	94.3	-	-
(2) ***End of term survey***						
My voice was heard during group discussions (Agreement 0-10)	Mean	8.7	8.3	8.8	8.6	8.9
	S.d.	1.8	2.1	1.7	1.7	1.6
Leadership role in team (0 if no, 1 if yes)	Mean	0.3	0.3	0.3	0.2	0.4
	S.d.	0.5	0.5	0.5	0.4	0.5
Expected Mark End (0-100)	Mean	71.0	70.0	71.3	71.5	70.6
	S.d.	8.8	8.5	8.9	9.0	8.7
	N	187	39	148	98	75
More confident in voicing my view in future interactions (Agreement 0-10)	Mean	5.1	4.8	5.2	5.3	4.9
	S.d.	3.1	2.9	3.1	3.1	3.1
More confident that my view will be heard in future interactions (Agreement 0-10)	Mean	5.2	5.1	5.3	5.5	5.1
	S.d.	2.9	2.5	3.1	2.9	3.0
	N	85	22	63	33	47
Peers nominate as leader in team (0 if no, 1 if yes)**	Mean	0.3	0.3	0.3	0.2	0.3
	S.d.	0.4	0.5	0.4	0.4	0.5
	N	187	39	148	98	75
Average response rate	%	59.8	58.4	60.2	63.9	60.7

Notes: *2021-2022 cohort only, ** 2020-2021 cohort only. We obtain the data on native language from the baseline survey; thus, average response rate for native vs non-native speakers is in relation to baseline response.

4.3 Empirical strategy

We aim to understand the effect of native language and gender composition on academic achievement, team dynamics, and expectations of future performance. Estimating peers' effects on student outcomes can be challenging, mainly because of a self-selection problem. University students often get to choose which classes to take, which can lead to endogenous peer group composition. The endogeneity arises when an unobserved characteristic is correlated with the outcomes of interest and with the probability of joining a group. This endogeneity can severely bias the estimation of peer effects, as individuals in the same group will have correlated outcomes even if peer effects do not exist.

There is no selection problem in our setting because group formation is exogenous to peers' characteristics and expected results at every level of interaction: course, seminar group, and team. First, there is no self-selection at the course level because the course is compulsory for all Social Policy master's students. Second, course administrators allocated students to seminar groups independently from students' characteristics including gender and native language. Students could request changes in their allocated seminar group only if they confronted a timetable clash. Where such a clash necessitated a re-allocation, the course administrator relocated the student to any non-clashing seminar and thus preserved the allocation's exogeneity. However, to identify causal effects from the team composition, we further randomised students within each seminar into smaller groups of approximately four students in each. As seminar size varied between cohorts, there were either two or four teams per seminar, but the total number of teams remained constant across cohorts. Table 4.3 shows the average and standard deviation of the proportion of women and native speakers per team. As there is some variation in gender composition (s.d. 0.21), and on the proportion of native speakers across groups (s.d. 0.48), we can estimate compositional peer effects.

Table 4.3: Proportion of women and native speakers per team

Cohort	**Proportion of women**		**Proportion of native speakers**		**Groups**
	Mean	S.d.	Mean	S.d.	(N)
2020-2021	0.79	0.21	0.42	0.28	44
2021-2022	0.81	0.21	0.55	0.25	44
Total	**0.80**	**0.21**	**0.48**	**0.27**	**88**

4.3.1 Estimation

We use a linear model to estimate the causal impact of the proportion of women and proportion of native speakers on all the outcomes of interest. Manski (1993) introduced the original model to estimate peer effects, which attributes outcomes to individual characteristics and the characteristics of a group to which a student belongs. We extend and develop that model to capture the effects on each of the outcomes of interest of variation in gender composition and share of native english speakers as follows:

$$Y_{igs} = \alpha + \beta_1 W_g + \beta_2 NS_g + \beta_3 W_{gs} + \beta_4 NS_{gs} + \gamma X_{ig} + C_i + \epsilon_{igs} \tag{4.1}$$

where Y_{igs} is the outcome of interest for student i in group g and seminar s, W_g is the gender composition of group g, NS_g is the percentage of native speakers in group g. W_{gs} and NS_{gs} are the percentage of women and of native English speakers in seminar s, but excluding students in student i's own group g. X_{ig} is a vector of control variables including age, familiarity with subjects relevant to the course, dummies for English as first language and gender, student's highest level of education, previous UK studies, and teacher and academic advisers' characteristics. Additionally, when we estimate equation 4.1 for the subsample of non-native speakers, we include Chiswick and Miller (2005)'s language distance scores in X_{ig}. C_i is a dummy for the student's cohort.

Note that groups are a sub-unit of the seminar. Therefore, the composition of the seminar varies with the composition of the groups. Thus, excluding students in the same team for the measures W_{gs} and NS_{gs} helps to avoid multicollinearity problems. More importantly, as students can neither self-select into seminar groups nor into study groups, W_g , W_{gs}, PNS_g and PNS_{gs} are all exogenous to the outcomes Y_{igs}. Therefore, all the coefficients of interest, that is β_1, β_2, β_3, and β_4, provide causal information about compositional effects on the outcomes of interest.

We estimate equation 4.1 for three sets of outcomes: *performance* measured by academic achievement, *perceptions* of team dynamics, and *predictions* of academic expectations. Performance outcomes correspond to examination marks, for which we estimate equation 4.1 using ordinary least squares. We cluster standard errors by cohort and seminar level.

To estimate the effects on perceptions and predictions we use endline survey data. The key team dynamics outcomes comprised responses to items such as "Voice was heard", which was measured on a scale of 0-10; and "Leadership role", which was binary. The main variable on expectations is "Expected mark" which is measured from 0-100. For the continuous outcomes, we estimate the effects with a least squares regression, and a probit for binary outcomes. Additionally, because the endline survey has a 60% response rate, adjust for survey non-response before running the regressions on these variables. We adjust for nonresponse using the inverse probability weighting method.[7] First, we classify all individuals (responders and non-responders) into cells based on observable characteristics that predict whether an individual responds or not; we use student's cohort, seminar group, gender, native English status, and information on previous studies in the UK. Then, we calculate the response probability conditional on this set of characteristics and assign responders a weight corresponding to the inverse response probability of their cell.

For a detailed description of outcomes of interest see Appendix 4.6.3.

[7] For a detailed description of the Inverse probability weighting method see Hernán and Robins (2016)

4.4 Results

4.4.1 Performance on Academic outcomes

Table 4.4 presents the estimated coefficients of equation 4.1 for exam marks and for the "Distinction" outcome. Column (1) presents estimates for the regression on marks for the whole sample, and columns (2) to (5) for the regression on each group. We find no significant gender differences in exam marks, but native English speakers have significantly higher marks than non-native English speakers even when controlling for other factors. These results might be attributable to the difficulties that non-native English speaking students confront in learning in a second language (e.g., Bernhofer and Tonin (2022)).

In terms of effects of gender composition, we find that an increase in the percentage of women in the team and seminar caused an increase in exam marks for non-native English speakers. For instance, column (5) shows that for non-native speakers an increase in 1% of the percentage of women in the team caused an increase in 0.08 points in exam marks. Thus, in a group of 4, 1 more woman in the group (25% increase in percentage of women), caused an average increase of 2 points.

Table 4.4: Regression coefficients: Exam Marks

	(1) All	(2) Women	(3) Men	(4) Native Speakers	(5) Non-Native Speakers
Previous UK Studies	0.81	-0.26	7.52*	0.43	0.73
	(1.55)	(1.84)	(2.89)	(2.25)	(2.62)
Gender (Female=1)	-1.46			-1.81	-0.60
	(1.45)			(1.69)	(2.20)
Native English	4.20**	4.18*	2.22		
	(1.45)	(1.67)	(2.52)		
Percentage Native Speakers (Team)	-0.00	0.00	-0.04	0.01	-0.01
	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
Percentage Native Speakers (Seminar-O)	0.02	0.01	0.08	0.02	0.03 0.01
	(0.03)	(0.04)	(0.05)	(0.04)	(0.03)
Percentage of Women (Team)	0.01	0.01	-0.03	-0.05	0.08*
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Percentage of Women (Seminar-O)	0.06	0.06	0.03	0.06	0.09
	(0.03)	(0.03)	(0.10)	(0.05)	(0.05)
Language Score					4.39
					(2.33)
Constant	84.21***	81.66***	91.78***	92.65***	59.28***
	(5.52)	(6.32)	(16.82)	(16.03)	(11.87)
N	343	273	70	167	155

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. Exam Marks are in scale from 0-100. Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: Administrative records for cohorts 2020-2021, and 2021-2022. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, cluster standard errors in parentheses.

4.4.2 Perceptions of team dynamics

We now analyse the perception of team dynamics by testing the causal role of linguistic and gender diversity on voice and leadership.

Voice in teams

Table 4.5 presents the regression estimates for the item "My voice was heard during group discussions" (measured as level of agreement from 1-10). Column (1) shows that an increase in the percentage of women in the team and in the seminar group caused an increase in students' level of agreement. Estimates by subsample suggest that women and non-native speakers benefited from having more women in the group, as an increase in the percentage of women either in one's own team or in the seminar group caused an increase in the extent to which students agreed with the statement. For instance, for non-native speakers an increase in 1% of the percentage of women in the classroom caused an increase in the level of agreement of 0.02 points. Thus, in a group of 4, 1 more woman in the group (25% increase in percentage of women), caused an average increase of 0.5 points in level of agreement with the statement. On the other hand, an increase in the percentage of native speakers had the opposite effect on non-native English speakers, who reported that they felt less heard when working in teams with more native English speakers. The effect was similar in size to the effect of the share of women, but in the opposite direction.

Table 4.5: Regression coefficients: "My voice was heard during group discussions"

	(1) All	(2) Women	(3) Men	(4) Native Speakers	(5) Non-Native Speakers
Previous UK Studies	0.08	-0.19	-0.02	-0.56	0.38
	(0.26)	(0.24)	(0.65)	(0.28)	(0.39)
Gender (Female=1)	0.48			0.32	0.78
	(0.30)			(0.35)	(0.54)
Native English	0.54*	0.37	0.59		
	(0.21)	(0.22)	(1.22)		
Percentage Native Speakers (Team)	-0.01	-0.01	0.00	0.00	-0.02*
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Percentage Native Speakers (Seminar-O)	0.00	0.00	-0.01	-0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Percentage of Women (Team)	0.01*	0.01	0.01	0.01	0.02*
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Percentage of Women (Seminar -O)	0.01*	0.02*	-0.00	0.01	0.02
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Language Score					0.71
					(0.36)
Constant	4.02*	6.00***	9.61*	10.42***	1.42
	(1.73)	(1.16)	(4.31)	(2.35)	(2.19)
N	218	174	44	104	101

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable "My voice was heard during group discussions" measures the level of agreement with the statement in a scale of 0-10. Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: End year survey, cohorts 2020-2021, and 2021-2022. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, cluster standard errors in parentheses.

We also present estimates for two follow-up questions related to student perceptions of how interactions with their team would affect their future interactions. The first is "more confident in voicing their views" (measured as level of agreement from 1 to 10). In Table 4.6, column (4) shows that for non-native English speakers, an increase in the percentage of female peers caused an increase in their confidence in voicing their views.

The second follow up question pertained to students' perception of how others would receive their views, which was also measured on a scale from 1 to 10. Table

4.6, columns (5) to (8) present the regression coefficients of this item by subsample. Similar to the previous survey item, column (8) shows that an increase in the percentage of women in the team and in the rest of the seminar group caused an increase in reported confidence on how their views will be received in the future. Results from the voice items suggest that non-native speakers benefited from having peer groups with a higher percentage of women.

Table 4.6: Regression coefficients: "More confident in voicing my view" and "my voice will be heard"

	"More confident in voicing my view"				"My voice will be heard"			
	(1) Women	(2) Men	(3) Native Speakers	(4) Non-Native Speakers	(5) Women	(6) Men	(7) Native Speakers	(8) Non-Native Speakers
Previous	1.66*	2.38	1.87	1.43	0.69	1.10	0.62	0.27
UK Studies	(0.79)	(1.75)	(0.99)	(0.97)	(0.81)	(1.46)	(1.10)	(0.95)
Gender			-0.91	0.94			-1.17	1.92
(Female=1)			(0.98)	(0.97)			(0.96)	(0.97)
Native	-0.38	1.29			-0.75	1.40		
English	(0.97)	(1.63)			(0.91)	(1.36)		
Percentage Native	0.00	0.01	0.01	-0.01	0.00	0.00	0.00	-0.01
Speakers (Team)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Percentage Native	0.01	-0.03	-0.00	0.01	0.01	-0.04	0.01	0.00
Speakers (Seminar -O)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
Percentage of	-0.00	-0.02	-0.02	0.08**	0.01	-0.02	-0.02	0.09***
Women (Team)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Percentage of	0.02	0.04	-0.01	0.10**	0.02	-0.01	-0.02	0.09***
Women (Seminar -O)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.03)
Language				-0.78				0.16
Score				(0.83)				(0.75)
Constant	-3.20	10.55*	5.71	0.57	-1.71	12.66**	6.47	0.53
	(4.28)	(4.37)	(6.20)	(4.27)	(3.87)	(3.89)	(6.92)	(4.19)
N	59	21	47	30	57	20	45	30

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable "More confident in voicing my view" measure the level of agreement (from 0 to 10) with the statement: "Working in teams for SP401 made more confident than before in voicing my view in future interactions". Data source: End year survey 2021-2022. The dependent variable "My voice will be heard" indicates the level of agreement (from 0-10) with the statement "Working in teams for SP401 made more confident than before that my view will be heard in future interactions". Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: End of year survey 2021-2022. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, cluster standard errors in parentheses.

Leadership

In addition to voice, we present estimates of three measures of leadership. First, we present a binary measure of leadership based on self-reported data gathered in the end of year survey. The dependent variable is one if the student answered "Manager/Leader" to the question: "Which of these roles best describes your own role in your sub-group?" and zero if they chose any of the other alternatives (Sceptic/Thinker, Checker/Recorder, or Conciliator). Table 4.7 presents the estimated coefficients for this measure. Across the whole sample, native speakers were more likely to report that they adopted a leadership role, but there were no statistically significant compositional effects of language or gender. However, the estimated coefficients from the sub-sample regressions that appear in Table 4.7, column (3) show that men are more likely to consider themselves team leaders when they have more female peers.

Table 4.7: Probit regression coefficients: Identify as leader

	(1) All	(2) Women	(3) Men	(4) Native Speakers	(5) Non-Native Speakers
Previous UK Studies	0.20	0.13	-3.42	0.30	-0.09
	(0.22)	(0.25)	(2.73)	(0.30)	(0.42)
Gender (Female=1)	0.12			0.48	-0.85
	(0.25)			(0.30)	(0.47)
Native English	0.67**	0.80***	10.03**		
	(0.22)	(0.24)	(3.59)		
Percentage Native Speakers (Team)	0.00	0.00	0.06**	0.00	-0.01
	(0.00)	(0.00)	(0.02)	(0.00)	(0.01)
Percentage Native Speakers (Seminar -O)	0.01	0.01*	-0.16**	0.01	0.00
	(0.00)	(0.00)	(0.05)	(0.01)	(0.01)
Percentage of Women (Team)	-0.00	-0.00	0.02	-0.00	0.01
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Percentage of Women (Seminar -O)	0.00	-0.00	0.33**	0.01	-0.01
	(0.00)	(0.01)	(0.10)	(0.01)	(0.01)
Language Score					0.54
					(0.33)
Constant	-2.79**	-2.53*	-52.62**	-3.50	-2.95
	(1.07)	(1.15)	(16.14)	(1.96)	(1.72)
N	218	174	44	104	101

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable "Identify as leader" is a binary variable that indicates if a student indicates that the role that describes them best is Manager/Leader (value 1) or not (value 0). Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: end of year survey cohorts 2020-2021, and 2021-2022. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, standard errors in parentheses.

In Table 4.8 we present estimates for two outcomes, a measure of change in the leadership role (columns (1) to (4)), and the outcome "Others identify as leader" (columns (5) to (8)). The first measure is constructed by comparing students' responses in the baseline and end surveys. It is equal to 1 if the student reported not commonly taking the leadership position in their teams in the baseline survey but then reported to be the leader of their team in the end survey, 0 if they provided the same answer in baseline and end survey, and -1 if they reported taking leadership roles at baseline, but not a leadership role in endline. Thus, we interpret a positive coefficient as increasing willingness to take on leadership roles. The results are similar to the first measure of leadership. For men (Table 4.8, column 2), an increase in the percentage of female peers causes an increase in the probability of changing their perception of their role in a team, towards being more likely to identify as a leader. More specifically, for men an increase in 1% of the percentage of women in the team, causes an increase on the dependent variable of 0.02 points. Thus, in a group of 4, an additional woman in the group would cause an average increase of 0.5 (or 25%) in the change of willingness to lead.

Table 4.8 also shows the probit regression coefficients for peers' perceptions of leadership. This outcome is a binary measure equal to 1 for student i if someone in their team reports that student i had the lead role in the team. For non-native English speakers, being in groups with more native speakers makes them less likely to be nominated by others as the team leader, while being in teams with more women makes it more likely for them to be mentioned as the team leader.

Table 4.8: Regression coefficients: Change in Leadership self-perception and others identify as leader

	Change in Leadership self-perception				Others identify as leader (Probit)			
	(1) Women	(2) Men	(3) Native Speakers	(4) Non-Native Speakers	(5) Women	(6) Men	(7) Native Speakers	(8) Non-Native Speakers
Previous UK Studies	0.03	-0.25	0.10	0.01	0.32	1.28	0.06	0.35
	(0.12)	(0.20)	(0.14)	(0.13)	(0.31)	(1.01)	(0.39)	(0.44)
Gender (Female=1)			-0.02	0.06			0.48	-0.68
			(0.14)	(0.12)			(0.48)	(0.49)
Native English	-0.03	0.44			0.32	-0.43		
	(0.11)	(0.28)			(0.25)	(0.63)		
Percentage Native Speakers (Team)	-0.00	-0.00	-0.00	-0.00	-0.00	-0.02	0.00	-0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Percentage Native Speakers (Seminar -O)	0.00	0.00	0.00	0.00	0.00	-0.02	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)
Percentage of Women (Team)	-0.00	0.00	0.00	-0.00	0.01	-0.02	0.01	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)
Percentage of Women (Seminar -O)	0.00	0.02***	0.00	0.00	-0.00	0.01	0.01	-0.00
	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)	(0.01)
Language Score				0.18*				0.04
				(0.08)				(0.09)
Constant	0.00	-1.50*	-1.80*	-0.19	0.312	1.949	-1.228	-1.183
	(0.53)	(0.64)	(0.82)	(0.36)	(0.24)	(0.57)	(-0.51)	(-0.60)
N	101	28	70	53	130	34	71	80

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The variable "Change in Leadership self-perception" is the difference between the two binary "Identify as leader" variables (at end of term minus at baseline). Thus, the variable is equal to zero if there is no change, 1 if student answers leader at end but not at baseline, and -1 if student answers leader at baseline but not at the end survey. Data source: Baseline survey and end of year survey cohort 2021-2022. The "Others identify as leader" variable is is equal to 1 if a student was mentioned as taking the Manager/Leader by at least one member of their team, and zero otherwise. Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: end of year survey 2020-2021. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, standard errors in parentheses.

Predictions of academic achievement

Finally, we present the regression coefficients for three outcomes related to students' predictions of their academic performance on the course. The first of these outcomes is their expected mark on the course, according to what students reported at the end of the term before they sat the examination. Table 4.9 presents the estimated coefficients for this model. The results suggest that the share of native speakers had strong effects on expected mark, especially for non-native English speakers. An increase in the share of native English speakers in the team or seminar reduced the expected mark that non-native speakers report at the end of the term. We also find that an increase in the percentage of women in the team caused an increase in expected mark. This result is in concordance with the results for actual marks (presented in section 4.4.1), as an increase in the share of women in the group caused an increase in exam mark. Thus, students predicted the direction of the favourable effect.

Table 4.9: Regression coefficients: expected mark at end

	(1) All	(2) Women	(3) Men	(4) Native Speakers	(5) Non-Native Speakers
Previous UK Studies	-2.96*	-3.14	-1.91	-5.53***	1.91
	(1.41)	(1.60)	(2.95)	(1.58)	(2.42)
Gender (Female=1)	0.97			-0.23	4.21
	(1.52)			(1.89)	(2.54)
Native English	0.09	-0.93	3.30		
	(1.38)	(1.53)	(4.20)		
Percentage Native Speakers (Team)	-0.02	-0.02	-0.01	-0.01	-0.01
	(0.02)	(0.03)	(0.05)	(0.04)	(0.03)
Percentage Native Speakers (Seminar-O)	-0.00	-0.02	0.06	0.06	-0.07*
	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)
Percentage of Women (Team)	0.06*	0.06*	0.02	0.05	0.08*
	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)
Percentage of Women (Seminar -O)	0.04	0.03	0.05	0.08*	-0.02
	(0.03)	(0.03)	(0.07)	(0.03)	(0.04)
Language Score				-2.69	
					(1.56)
Constant	62.18***	60.02***	72.96***	56.37***	72.85***
	(5.94)	(6.69)	(15.56)	(10.72)	(7.83)
N	217	173	44	103	101

Notes: All models include controls for: age, Education level, Experience with quantitative methods, Experience with qualitative methods, Seminar leader gender, seminar leader native English, adviser gender, adviser native English. The dependent variable measures the student's expectation of their exam mark at the end of the course (on a scale from 0-100). Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: end year survey. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, cluster standard errors in parentheses.

We also analyse expected marks in comparison with student's original expectations of academic performance, which we collected in the baseline survey. Table 4.10 presents the estimated coefficients of the regression with the difference between expected mark at baseline versus end of term (see columns (1) to (4)). We find that students in groups with more native speakers experienced a larger decline in their predicted performance. Non-native English speakers have a higher initial expected mark than native speakers (average of 73.7 vs average of 71.4), and the decrease is on avarege leading to a better prediction of actual mark

Lastly, Table 4.10 columns (5) to (8) present the estimates for the difference between the expected mark at the end of the course with the actual mark. We do not find any statistically significant effect of the share of women nor of the share of native speakers on the gap between expectations and real mark.

Table 4.10: Regression coefficients: Change in mark's expectations and Difference between actual mark and expected mark

	Change in mark's expectations (Start to end)				Difference between actual mark and expected mark			
	(1) Women	(2) Men	(3) Native Speakers	(4) Non-Native Speakers	(5) Women	(6) Men	(7) Native Speakers	(8) Non-Native Speakers
Previous	2.39	-2.46	1.21	5.64**	3.93	9.27	8.67**	-3.74
UK Studies	(1.23)	(5.38)	(1.60)	(1.86)	(2.56)	(5.34)	(2.60)	(4.12)
Gender			-1.54	1.50			0.39	-0.33
(Female=1)			(3.12)	(2.86)			(2.62)	(4.85)
Native	1.18	12.69			5.54*	5.04		
English	(1.29)	(9.47)			(2.47)	(5.25)		
Percentage Native	-0.07**	-0.16	-0.06	-0.11**	0.03	-0.03	0.01	0.02
Speakers (Team)	(0.02)	(0.08)	(0.05)	(0.03)	(0.04)	(0.08)	(0.05)	(0.05)
Percentage Native	-0.06*	0.10	0.04	-0.09*	0.02	0.15	-0.03	0.11
Speakers (Seminar -O)	(0.03)	(0.11)	(0.04)	(0.04)	(0.05)	(0.10)	(0.04)	(0.06)
Percentage of	0.05	0.07	0.06	0.06	-0.03	-0.01	0.00	-0.04
Women (Team)	(0.03)	(0.08)	(0.03)	(0.05)	(0.05)	(0.09)	(0.05)	(0.06)
Percentage of	0.01	-0.13	0.03	-0.06	0.01	0.00	-0.01	0.05
Women (Seminar -O)	(0.03)	(0.09)	(0.03)	(0.05)	(0.04)	(0.12)	(0.05)	(0.07)
Language				-2.93				9.58**
Score				(1.88)				(2.88)
Constant	5.36	-27.44	-26.62	7.34	16.76	-8.13	-0.21	-12.58
	(6.52)	(35.93)	(20.91)	(11.29)	(9.50)	(20.99)	(14.24)	(13.97)
N	173	44	103	101	172	44	103	100

Notes: All models include controls for: age, education level, experience with quantitative methods, experience with qualitative methods, seminar leader gender, seminar leader native English, adviser gender, adviser native English. The dependent variable " Change in mark's expectations" corresponds to the difference between the student's expected mark at the end and start of the course (Expected Mark at end – Expected Mark at baseline). Data source: baseline survey and end year survey. The dependent variable "Difference between actual mark and expected mark" is the difference between the student's exam mark and their expected mark at the end of the course (Mark - Expected Mark at end). Seminar - O refers to other members in the seminar group (excluding individual's own team). Data source: administrative records and end year survey. $^{\star}p < 0.05$, $^{\star\star}p < 0.01$, $^{\star\star\star}p < 0.001$, cluster standard errors in parentheses.

4.5 Conclusions

We implemented a field experiment to measure the role of gender composition and the share of native English speakers in small groups of postgraduate students. We study a cohort of students from a MSc programme in a top UK University. The programme is female dominated, and very international, which makes it an ideal setting to study how internationalisation and the increase in women's participation in higher education affects group interactions as well as individual outcomes. To do so, we randomised students into small teams to generate exogenous variation in the demographic characteristics of students across groups. Then, we analysed the effects of group composition on three sets of outcomes: (1) academic *performance*, (2) *perceptions* of team dynamics, (3) and *predictions* of academic performance.

Regarding individuals' academic performance, we find that non-native English speakers do worse than their native peers but that this gap closes when they have more female peers. For non-native English speakers, an increase in the proportion of women in the seminar group causes an increase in exam marks and on the probability of graduating with distinction.

We also find that group composition affects outcomes related to team dynamics. Firstly, women and non-native speakers were more likely to report their voices were heard when in groups with more women. Non-native speakers agreed less with the same statement when in groups with more English native speakers. Secondly, when looking at leadership measures, we find that men were more likely to identify as their team leaders when they were in seminar groups with more women. Also, when we compare this result with students' answers in the baseline survey, we see that men were more likely to change their answers (from not feeling like a leader to being the leader of their group) when they collaborated with more women. We also find that non-native speakers were less likely to be signalled as the team leader by their peers when they collaborated with more Native English speakers.

Taken together, these results indicate that group composition, at least in the gender and linguistic dimension, plays a relevant role in both academic outcomes and students' self-perception of their group work interactions and related skills. More specifically, we show that non-native speakers benefit from having peers from diverse linguistic backgrounds. At the same time, we do not see any adverse effect of a higher share of non-native speakers on native speakers' outcomes.

Furthermore, in contrast to what previous studies have found for women in minority status, men do not seem to suffer from being in the minority in any outcome. Related literature has shown that women are less likely to take leadership roles when working in male-dominated environments. For instance, in a laboratory setting, Stoddard et al. (2020) finds that women are less likely to be considered influential or chosen as a spoke-person in predominantly male teams. However, it is unclear if the effect is caused by "tokenism" (the effect of being in the minority) or if it's caused by gender attitudes. Our analysis in a female-dominated setting suggests that the latter is true. Contrary to when women are in the minority, men's attitudes towards leadership are positively affected, i.e., they are more willing to take leadership roles when in groups with more women.

Our results are relevant for course design and team formations in educational settings, where students often need to work on group projects during their studies, and to the industry and the public sector. In terms of linguistic diversity, as non-native speakers' performance and perception of being heard improve when non-native speakers are in teams with a low presence of native speakers, higher education settings and workplaces should consider creating teams where there is a large enough proportion of non-native speakers, especially in settings where teams need to solve highly complex problems when everybody's opinion must be considered to lead to the best outcome. Additionally, this allow non-native speakers to practice and develop interpersonal skills that are highly valued in the labour market. Regarding gender composition of teams, our results suggest that women benefit from being in teams where there is a large fraction of women in them. Since the teams that we created had a low presence of males in them due to the female-dominated nature of the course we study, our findings suggest that for universities and workplaces it may make sense to create female-only teams, especially when it is important that everyone's opinion is heard, which is the case when complex problems are at hand.

4.6 Appendix

4.6.1 Student's country of birth and native language

Table 4.11: Number (N) of students by country and cohort

Country of birth	2020-2021 (N)	2021-2022 (N)	Total (N)	Country of birth	2020-2021 (N)	2021-2022 (N)	Total (N)
Albania	1	0	1	Mauritania	1	0	1
Argentina	2	2	4	Mexico	3	2	5
Armenia	0	1	1	Nepal	1	0	1
Australia	2	1	3	Netherlands	2	0	2
Bahrain	0	1	1	Nigeria	2	0	2
Bangladesh	1	2	3	Norway	0	3	3
Belgium	0	1	1	Pakistan	5	5	10
Brazil	2	1	3	Panama	1	0	1
Bulgaria	2	0	2	Paraguay	1	0	1
Burma (Myanmar)	0	1	1	Peru	0	1	1
Canada	5	7	12	Philippines	2	1	3
Chile	0	3	3	Poland	2	2	4
China	26	21	47	Qatar	0	1	1
Colombia	1	5	6	Romania	1	1	2
Dominican Rep.	1	0	1	Russia	0	1	1
England	38	38	76	Saudi Arabia	0	1	1
Eritrea	0	1	1	Scotland	2	0	2
FYR Macedonia	1	0	1	Singapore	0	3	3
Finland	1	0	1	South Africa	0	1	1
France	9	5	14	South Korea	4	5	9
Germany	1	5	6	Spain	2	1	3
Ghana	0	1	1	Sri Lanka	0	1	1
Greece	2	0	2	Sudan	1	0	1
Hong Kong	3	6	9	Sweden	0	1	1
Hungary	1	0	1	Taiwan	0	1	1
India	12	13	25	Thailand	0	2	2
Indonesia	2	0	2	Turkey	3	1	4
Ireland	0	2	2	USA	18	20	38
Italy	7	10	17	Ukraine	2	0	2
Japan	2	2	4	Uruguay	1	1	2
Jordan	1	0	1	Utd Arab Emts.	1	1	2
Kazakhstan	1	0	1	Venezuela	1	1	2
Kenya	1	1	2	Vietnam	1	0	1
Lebanon	1	0	1	Wales	1	1	2
Lithuania	1	1	2	Zimbabwe	0	1	1
Luxembourg	0	1	1				
Malaysia	1	0	1	**Total**	**186**	**191**	**377**

Source: University administrative data records for the course

Table 4.12: Non-native English speakers Language scores (Linguistic Distance)

Native Language	**N**	**Score**	**Native Language**	**N**	**Score**
Arabic	4	1.5	Mandarin	6	1.5
Bengali	1	1.75	Mandarinchinese	1	1.5
Bulgarian	1	2	Nepali	1	1.75
Burmese	1	1.75	Norwegian	1	3
Cantonese	5	1.25	Polish	5	2
Chinese	38	1.5	Portuguese	3	2.5
Dutch	2	2.75	Punjabi	1	1.75
French	14	2.5	Rumanian	2	3
German	5	2.25	Russian	1	2.25
Greek	1	1.75	Spanish	23	2.25
Gujarati	1	1.75	Swedish	3	3
Hindi	3	1.75	Tagalog	1	2
Hungarian	1	2	Tamil	3	1.75
Indonesian	1	2	Telugu	1	1.75
Italian	14	2.5	Thai	1	2
Japanese	4	1	Turkish	3	2
Korean	7	1	Vietnamese	1	1.5
Malayalam	2	1.75	**Total**	**182**	**1.92**

Notes: N represent the total number of students who reported each language as their Native language. Score is the Chiswick and Miller (2005) measure of linguistic distance from each native language to English. The measure ranges from 1 to 3, with three being the most similar to English.

4.6.2 Survey Attrition

Although most students answered the baseline survey, there is some attrition at the point of the end survey. Table 4.13 presents descriptive statistics for the group of end survey respondents (Response=1), and non-respondents (Response=0). Table 4.14 presents the odd ratios of the logistic regression used for the inverse probability weighting.

Table 4.13: Covariate's mean value for respondents and non-respondents

Covariates	Means	
	Response=0	**Response=1**
Age	24.7	24.5
Female=1	0.79	0.80
Native English=1	.50	0.47
UK Studies=1	0.35	0.32
Highest level of education (Master) = 1	0.20	0.19
Proportion Native Speakers (Team)	45.8%	51.4%
Proportion Native Speakers (Seminar -O)	42.1%	47.8%
Proportion of Women (Team)	78.2%	80.5%
Proportion of Women (Seminar -O)	78.2%	81.6%
Expected Mark	71.5	73.1
Experience with Quantitative methods	4.0	4.2
Experience with Qualitative methods	5.7	5.9
Adviser Gender (female=1)	0.54	0.46
Adviser Native Language (English=1)	0.64	0 .55
Total (N)	153	226

Notes: Seminar - O refers to other members in the seminar group (excluding individual's own team). Data Source: Baseline survey and administrative records

Table 4.14: Odd ratios for end survey response

	Response
Cohort	0.82***
	(0.24)
Year Born	0.01
	(0.03)
Female =1	-0.05
	(0.28)
Native English =1	-0.18
	(0.23)
UK studies =1	-0.24
	(0.25)
Seminar Group	0.03
	(0.02)
Constant	-1668.34***
	(466.69)
N	355

Notes: Odd ratios. $^{\star}p < 0.05$, $^{\star\star}p < 0.01$, $^{\star\star\star}p < 0.001$, standard errors in parentheses.

4.6.3 Description of dependent variables

Table 4.15: Description of dependent variables

Variable	Description	Range	Data Source	Questionnaire Item
Academic performance				
Exam Marks	Mark on the course's final exam. This is the only assessment in the course.	0-100	Administrative records	
Perceptions of team dynamics				
"My voice was heard during group discussions"	The variable indicates the level of agreement with the statement "My voice was heard during group discussions"	0-10	End of year survey	Level of agreement from 0 to 10 with the following statements: (1) "My voice was heard during group discussions"
"More confident in voicing my view"	The variable indicates the level of agreement with the statement "Working in teams for SP401 made me more confident than before in voicing my view in future interactions"			(2) "Working in teams for SP401 made more confident than before in voicing my view in future interactions" (2021-2022 only)
"My voice will be heard"	The variable indicates the level of agreement with the statement "Working in teams for SP401 made me more confident than before that my view will be heard in future interactions"			(3) "Working in teams for SP401 made more confident than before that my view will be heard in future interactions. (2021-2022 only)
Identify as leader	Binary variable that indicates if a student indicates that the role that describes them best is Manager/Leader (value 1) or not (value 0)	0,1	End of year survey	"Which one of these roles do you think describes you best when you do group work?" (Baseline) "Which of these roles best describes your own role in your sub-group?" (End of year)
Change in Leadership self-perception	We construct this variable by combining the end survey leadership variable plus a baseline survey leadership variable. The baseline leadership variable takes value 1 if the student answers Manager/Leader, and zero otherwise. Then we compute the change in leadership as the difference between the two binary leadership variables (at end minus at baseline). Thus, the variable is equal to zero if there is no change, 1 if the student answers leader at end but not at baseline, and -1 if the student answers leader at baseline but not at the end survey.	-1,0,1	Baseline and End of year survey	(a) Manager/Leader: provides leadership and direction for the group, (b) Sceptic/Thinker: ensure the group avoids premature agreement, push the group to explore all possibilities, (c) Checker/Recorder: check for consensus among group members, record the group's solutions, (d) Conciliator: resolve conflicts, ensure that members feel 'safe' to give opinions"
Others identify as leader	The variable is constructed using peers' input. We asked students to match each of the roles with one or more members of their team. Then, we constructed a binary variable that is equal to 1 if a student was mentioned as taking the Manager/Leader by at least one member of their team, and zero otherwise.	0,1	End of year survey	Match the following role with one or more members of your team. You can include yourself. Manager/Leader: provides leadership and direction for the group (2020-2021 only)
Predictions of academic performance				
Expected mark at end	This variable measures the student's expectation of their exam mark at the end of the course.	0-100	End of year survey	From 0 to 100, what do you expect your final mark in this course to be?
Change in mark's expectations (Start to end)	We construct this variable as the difference between the student's expected mark at the end and the start of the course (Expected Mark at end – Expected Mark at start)	(-100) - 100	Baseline and End of year survey	
Difference between actual mark and expected mark	We construct this variable as the difference between the student's exam mark and their expected mark at the end of the course (Mark - Expected Mark at end)	(-100) - 100	Administrative records and end of year survey	

Chapter 5

Conclusions

5.1 Main findings and contributions

In this thesis, I studied three main issues related to how imbalances in higher education affect disadvantaged groups. First, I analysed the role of admission systems in reproducing educational disparities. In particular, I exploited a national reform in Chile to test if university admissions rules are sex neutral. Second, I studied the suitability of affirmative action as a policy response to correct existing gender imbalances. To do so, I evaluated interventions that aimed to increase female participation in two Chilean engineering schools. I studied the effects of these policies on several outcomes, such as the number of applications, attendance, and the distribution of academic ability among incoming students. Third, I analysed the potential further consequences of interventions that improve student body diversity. For this purpose, I estimated the effects of gender composition on grades and drop-out rates among engineering school students. Finally, I implemented an experiment in a UK university to investigate further how peers can affect each other.

5.1.1 The role of admission systems on gender imbalances in higher education

University admission decisions are often made considering at least two types of measures of academic ability: one-shot admission exams and measures of continuous assessment over a longer time horizon. In Chile, admissions to degree-programmes only depend on students' application scores, which are a linear combination of scores in a high-stakes national admission exam and students' high school records. In the second chapter of this dissertation, I tested whether the importance we assign to each of these two factors affects gender imbalances in higher education. To do so, I exploited the implementation of a reform in the Chilean centralised admission system. The reform consisted of adding a new admission criterion based on the students ranking among their high school peers. The new "class ranking" criterion increases the weight of continuous assessment on students' application scores. I compared men's and women's college admission scores before and after the reform, and I found that the reform had gendered consequences. The findings on application scores showed that the reform helped to reduce an existing gender gap. The reform increased women's application scores relative to men by around 5 points.

Additionally, I evaluated if students who benefited from the reform[1] chose degree-programmes in more prestigious fields of study. I used a multinomial logit to evaluate the effect of the reform on choosing one of the following fields of study: agriculture, arts, business, engineering, humanities, healthcare, natural sciences, social sciences, and vocational degrees. I found that the policy change affected students' choices, as men and women who benefited from the reform were less likely to apply and enrol in programs in the lower-paid fields.

Finally, I presented results of a "quasi-experimental" analysis on the causal effect of the reform on applications and attendance to engineering degree-programmes. I found that the reform caused an increase in women applications to engineering

[1] Students who benefit from the reform are the ones who graduated in the top half of their high school class grade distribution.

degrees of about 1.5 percentage points. Taken together the results indicate that the reform allowed more women to apply and enrol in the more selective and better paid degrees.

These findings contribute to the literature on the causes of gender imbalances in higher education. Previous research on the determinants of gender segregation in fields of study had focused on two main issues: differences in preferences between women and men, and achievement gaps in mathematics and academic preparedness for STEM degrees.[2] Although strong evidence supports the hypothesis that these two points are consequential to the gender imbalances, I show that admission criteria also play a significant role.

5.1.2 Effectiveness of Affirmative action policies on improving women participation in STEM

In the third chapter, I tested the effectiveness of two distinct interventions aimed to increase the share of women in Chilean engineering schools. One of the schools, UCH, expanded its capacity and added 40 female exclusive seats to their regular offer. These seats are allocated at the admission stage to the women with the highest scores, among those who applied but did not get a seat through the regular admission process. The other policy, from PUC, consisted of changing the curriculum to make it more attractive to women and increasing female representation among members of staff. I studied the effectiveness of these two policies by comparing women's applications and enrolments to these two schools, with those of the other 28 engineering schools that are part of the centralised admission process. Additionally, I investigated the effect of the reform on the gender gap in application scores at these two schools.

Using the Chilean centralised admission system's data, and a difference-in-differences estimation strategy, I found that the UCH policy increased the percentage of women applying to their engineering degree programme in about 10%, and the

[2] See McNally (2020) for a review of explanations for the gender gap in STEM participation.

share of women enrolling in their programme in about 8.4%. The increase in women enrolling comes from applicants in the whole spectrum of the score distribution, rather than from the women just below the cut-off score. The extra female-exclusive seats explain about 4% of the increase, while women enrolling in regular admission process explain the further 4.4%. The PUC policy managed to increase women participation in their engineering school at similar levels. As the UCH policy lowers the admission criteria for women taking extra seats, I tested whether the intervention had a significant effect on increasing the gender gap in application scores among cohorts of incoming engineering students. Application scores of students admitted at the UCH engineering schools are in the range of 7 to 8.5,[3] I find that the policy increased the gender gap in 0.05 points, which is less than a quarter of a standard deviation of the application scores of admitted students.

This third chapter contributes to the literature on affirmative action policies and the policy debate by providing evidence from two innovative interventions. This is the first study to show proof of the effectiveness of an affirmative action policy that is integrated into a centralised, objective, and transparent admission system. Moreover, the intervention requires minimal accommodation from the rest of the system for its implementation. Thus, it shows a viable alternative for policymakers and university administrators concerned with increasing minority participation but in contexts where centralised rules limit the range of actions. Additionally, this paper adds to the literature on affirmative action by shedding light on how a more diverse body of students affects academic outcomes, which I discuss in more detail in the following section.

5.1.3 The role of diversity on student's outcomes

Finally, I presented findings on how the composition of a student body affects students' outcomes in Chapters 3 and 4. These two chapters differ in setting, focus, and scope. In Chapter 3, I study compositional effects on undergraduate

[3] Test scores are on a scale from 1.5 to 8.5 with a median of 5.

students at a Chilean engineering school. The setting is male-dominated, and the peer group is at the classroom level. I investigated gender peer effects on educational outcomes to explain further consequences of the gender affirmative action policy that the school had implemented. Meanwhile, in Chapter 4, my co-authors and I examined the role of peers in a UK postgraduate programme with a high percentage of women and international students. This last chapter focused on interactions at a smaller peer level, as we randomly assigned students into four-people study groups. In comparison to the setting in Chapter 3, peer-to-peer interactions are more relevant in this course. First, because students work in teams, and second, discussions between students are the fundamental part of the seminar classes. In this chapter, we investigated the effect of the share of native English speakers and gender composition in a wider variety of outcomes, including students' perceptions of their interaction with their peers. In both analyses, I take advantage of the exogenous variation in the demographic characteristics of the relevant peer group members to estimate the group composition's effects on the outcomes of interest. The findings show that gender of peers matters in both male-dominated and female-dominated environments. Additionally, Chapter 4 shows that the native language of peers is a determinant of student outcomes, especially in terms of team dynamics and skills.

In the final set of analyses of Chapter 3, I used data from a sample of first-year engineering students who are exogenously assigned to classrooms of 100 students each. I presented estimates for grades on four core courses: Algebra, Calculus, Physics, and Computer Science. These courses are assessed via standardised exams. Additionally, I analysed the effects of gender composition on grades in an Introduction to an engineering course in which students work in small teams on a collaborative project. I find that increasing the share of women does not affect men's or women's academic performance in the four core courses. However, in the case of the introduction to engineering course, the share of women does have an impact on students' marks. Men and women in classrooms with more women have, on average higher final marks in this course. Finally, I presented estimates of drop-out rates after the end of the first term. I found that, for women, an increase in the share of women in the student's relevant peer group (of the equivalent size of the increase caused by the intervention) more than halves the probability of dropping out after the end of the first term. This last finding is particularly relevant, especially when evaluated in the light of affirmative action interventions, as it indicates that policies that aid women's admissions also improve women's completion rates.

This analysis contributes to two strands of the literature: the one on affirmative action and the one on gender peer effects in higher education. Regarding affirmative action, chapter three's last set of analyses provides novel evidence on further consequences of affirmative action on the student body. I show that increasing minority participation can multiply the impact of the intervention by improving minority students' persistence. In respect of peer effects, the I show that gender composition has a significant role in academic outcomes when collaboration is involved.

In Chapter 4, we use data from a postgraduate course on International Social Policy. In the course, students are assigned to seminar groups from about 8 to 16 students each, and within seminar groups they are posteriorly randomised in small teams of 4. We presented results for the effects of gender composition and share of native speakers at these two peer group levels (seminar and team). We looked at two academic outcomes, exam marks, and classification in the course, and several non-academic outcomes, such as measures of "voice" in team interactions, leadership role, and expectations of academic performance.

The findings of this last chapter support those of Chapter 3 and also contribute to the broader peer effects literature. We demonstrate that gender and native languages of peers significantly affect the nature of group interactions in ways that are consequential to students' outcomes. For instance, we showed that women and non-native English speakers feel more heard in teams when there are more women. Moreover, we showed that the share of women positively affects exam marks in the course. It then adds to the findings of paper three by providing evidence that gender composition matters when peer-to-peer interactions are relevant in determining individual outcomes. Additionally, as the evidence of peer effects in adults is scarce, the paper sheds light on the role of gender and linguistics diversity on team dynamics at the workplace.

5.2 Discussion and implications

5.2.1 Reducing horizontal inequalities in university admissions

The idea that a transparent, objective, and standardised university admission systems can contribute to reproducing inequality is perhaps counterintuitive. However, in the second chapter of this thesis, I prove that centralised admission systems do play a role in explaining gender imbalances in higher education, even when solely relying on objective factors. The explanation for this crucial result lies in two aspects of admission decisions. First, there are average differences in performance across groups. Second, the size of the differences or even what group is at an advantage depends on the type of task. For instance, in the case of gender, evidence suggests that, on average, women perform better in low-stakes and non-competitive tasks than in one-shot standardised exams (Cai et al. (2019), Schlosser et al. (2019), Iriberri and Rey-Biel (2019), Montolio and Taberner (2021), Niederle and Vesterlund (2010)). Therefore, decisions to place weight on attainment in one task or another beget consequences to the distribution of achievement across groups.

However, for the most part, the purpose of using application scores in centralised admission systems is not to help equalise opportunities among advantaged and disadvantaged groups but to reflect the likelihood of success in higher education. So then, a natural question that might arise among policymakers is: Are interventions that shift weight towards high school measures of achievement aiding or hindering the fitness of matches between universities and applicants?

The literature on the relationship between achievement measures and university outcomes can help us shed light on this query. Until not so long ago, there was general agreement that standardised test scores were a fair measure of skills. For those who held this view, the correlation between achievement on admission tests and students' race, gender or socioeconomic status reflected inequality in academic preparedness. Naturally, reducing the weight placed on these exams would inevitably result in admitting students who were not sufficiently prepared (Barro, 2001). However, this standard view has been challenged by empirical evidence. In particular, Rothstein (2004) showed that the predictive power of the US SAT scores is smaller than previous estimations implied. Later, Bulman (2017) showed that academic performance in later grades is the best predictor of university and labour market outcomes, greatly exceeding entrance exam scores.

Thus, changes in admission criteria can be a strong policy tool. Putting less emphasis on competitive tests, while making room for alternative measures of achievement, can help to reduce imbalances in higher education while also improving the match between universities and students.

In the case of gender, where the problem is about segregation rather than transition rates, the small and virtually costless changes in the weights used to calculate composite scores help women by reducing the application scores gap and allowing them to pursue more degree programmes in the more selective fields. However, as I show in the second chapter, although it is true that women with higher application scores choose STEM degrees more, they still do it at a lesser rate than when men improve their application scores. These findings lead to the conclusion that correcting profound disparities in admissions to university requires stronger efforts.

In the third chapter of my dissertation, I show that the use of gender affirmative action interventions can significantly improve women's representation in STEM fields. Moreover, the two studied interventions caused no decrease in the average application scores of the post-reform cohorts. The results imply that institutions can implement even "stronger" interventions to correct gender imbalances without generating a harmful mismatch between applicants and universities.

The study showed that a small flexibilization in admission requirements for women brought effects beyond what was expected. By expanding the number of vacancies to include a few dozens of women with scores just below the cut-off, the intervention also managed to increase the percentage of women admitted above the cut-off. Hence, the intervention increased women's admission rates by far more than the mechanical effect of the capacity expansion. Second, the finding that women are more likely to stay in the degree programme when surrounded by more female peers implies that successful gender affirmative action policies bear a multiplier effect. The policy manages to improve minority representation not only by increasing admission but also by improving retention rates. Therefore, induced changes in the gender composition of first year students have direct implications for the efficacy of the intervention.

This policy evaluation paints a more detailed picture for policymakers concerned with the implications of this sort of intervention. The changes in admission requirements at the national level and the targeted interventions in engineering school proved to be successful in broadening women's access to the STEM fields. The interventions then, are contributing to reduce horizontal inequalities in higher education by addressing the problem of field segregation. However, the impact of the policies in reducing inequalities is likely to extend beyond the educational arena.

The earning premium for STEM degrees is well documented (Gerber and Cheung (2008), Arcidiacono (2004), Blau and Kahn (2017)). Thus, with women entering more lucrative fields, we can expect that gender gaps in labour market outcomes will narrow down. From the individual's point of view, this translates into women having better chances of accessing the wide range of benefits that higher earnings bring, such as better access to goods and services, and better health (Adler et al., 1994).

The implications of increasing women participation in STEM are not reduced to equality, or fairness of treatment. Indeed, the results presented in this dissertation are also relevant for policy makers concerned with economic growth and efficiency. Indeed, there is a growing branch of economic literature dedicated to the study of the relationship between gender inequality and macroeconomic outcomes.[4] For instance, studying entrepreneurship and labour market participation Cuberes and Teignier (2016) shows that gender gaps in participation affect negatively both income and aggregate productivity, since they reduce the entrepreneurs' average talent. Such results should easily extend to the STEM sector, which is key in determining productivity and growth. The premise is the following: if talent is distributed equally among men and women, increasing female participation will increase the pool of talent from which the STEM sector can choose from, which will in turn lead to a more efficient allocation of human capital.

The findings on how the rise in the share of female peers improves women's persistence in engineering degrees highlight the importance of considering the effects of a more diverse student body on student outcomes, which I explore more deeply in the last chapter of this dissertation. Moreover, the findings are a key input to the design of policies that aim to either reduce gender inequality or to foster the development of the STEM fields. In male dominated environments, a more balanced gender composition of peers can contribute to improve the science "leaky pipeline" problem.[5] In a related study with US data, Bostwick and Weinberg (2022) find that more female peers increase the probability of completing a STEM doctoral programme. Thus, although it requires further study, it is likely that the finding on increased persistence apply to the workplace too. Employers from the public and private sector might increase the retention of women in science by allowing them to work closely with other female peers.

5.2.2 The importance of understanding the role of peers

From the analysis of the importance of diversity on grades, we learn that peer effects, in terms of gender composition but also on linguistics diversity, are more relevant when interactions are intensive (as when working in small groups). From

[4] For a recent review of this literature see Kolovich et al. (2020).

[5] The leaky pipeline is a metaphor for the way girls and women opt out of STEM at different stages of their career path. See Clark Blickenstaff (2005) for a review of explanation to this phenomenon.

the study of engineering students, I find no evidence of gender peer effects in big classrooms (100 students), but I find them in the collaborative project where students work in small groups of 5 students each. Moreover, when studying postgraduate students, I find gender composition and share of native speakers relevant at the seminar level (8-16 students) and the team level (4 students).

The last chapter also provides some insight as to why the nature of the interaction is relevant for outcomes. By studying students' dynamics, we learn that the gender composition of the group is an important determinant of "voice". Students in groups with more women feel like their opinions are better heard. We find the same result when international students are in groups with more non-native speakers. This might lead to students speaking their views and opinions more and, in turn, contributing to their learning and the learning of their peers.

Additionally, studying a more female-dominated student body helps us better understand gender dynamics. An important finding of this study is that men are more likely to perceive themselves as leaders when they have more female peers. Then, while increasing gender diversity in male-dominated environments can improve the learning experiences of female students, evidence from the last chapter suggests otherwise in the inverse case. As men's outcomes do not suffer when they are a minority, we can infer that gender attitudes or stereotypes explain the problem for women in male-dominated environments.

Overall, the findings have implications for policymakers, university leaders, and employers concerned with diversity in classrooms and teams. In terms of gender, policy suggestions vary depending on context. In male-dominated environments, increasing the proportion of women might be beneficial to improve women's persistence, and individual and group performance. However, when using affirmative action is not possible or desirable, deliberate group design can also improve various outcomes. In male-dominated environments, allocation groups of only males would allow for a better gender balance in the groups where women participate. In female-dominated environments, the same idea can be pushed further, perhaps even allocating women to only-women groups. Booth et al. (2018) study in undergraduate economic courses in the UK provides some evidence that also support this policy suggestion, as they find that young women assigned to all-female classes in their first year of university obtain better academic outcomes than women in coed classes.[6]

The findings on linguistic diversity alleviate concerns about the potential negative impact of the increase of international students on the academic achievement of native students. Moreover, the findings inform the efficient design of teams, which can be helpful for employers dealing with the internationalisation of the workforce. For instance, as non-native speakers feel more listened to when working closely with other non-native speakers, including several international team members might help make all views and ideas heard.

5.3 Avenues of future research

There is scope for future advances in the evaluation of admission policies and their impact on inequality. A straightforward direction to follow is to explore the impact of admission weights on mobility. With the Chilean admission reform and data on applicants' socioeconomic background, it would be possible to study the effect of the reform on the distribution of socioeconomically disadvantaged students across fields of study and better and lower ranked institutions. Moreover, as the first cohort of students who benefited from the reform have already graduated and are probably part of the labour market, it is now possible to study the longer-term effects of the reform on quality of employment and income inequality.

Another interesting extension of this dissertation would be to implement and evaluate interventions that aim to improve the retention rates of women in STEM. An almost costless intervention is to widen the range of variation in the gender composition of the peer group. For example, in settings where men are the major-

[6] The literature studying primary and secondary school offer further support to the hypothesis that single-sex environments can improve women outcomes. For instance, Dustmann et al. (2018) find causal evidence that girls in single-sex schools outperform their counterparts in coed schools.

ity, the range of the gender composition of smaller groups (classroom or seminar group) can be widened by intentionally allocating women to groups with more women while leaving some single-sex groups of men. A wider range of variation that includes a sufficiently high number of groups with women in the majority can help to understand if there is any non-linearity in the effect of the share of women on retention rates found in the second chapter of this dissertation. Moreover, studying similar interventions or quasi-experiments in STEM industries can shed light on how to address the leaky pipeline problem in further stages of women's career paths.

Finally, the study of team composition in work settings needs much further exploration. The study of peers' influence is relevant not only in STEM but in businesses, politics, public enterprises, and any field where collaborative work is important for productivity. Up to now, we know from a few studies in educational settings about the effects of team composition on individual outcomes but there is a lack of causal evidence on its effects on team's outcomes. Moreover, incentives for employees and students might differ. Thus, although it might be difficult to find a good opportunity, setting experiments at the workplace can generate big contributions to the literature on peer's effects, and on diversity and team productivity.

Bibliography

Adler, N. E., T. Boyce, M. A. Chesney, S. Cohen, S. Folkman, R. L. Kahn, and S. L. Syme (1994). Socioeconomic status and health: The challenge of the gradient. *The American psychologist 49*(1), 15–24.

Adler, T. J. and M. Ben-Akiva (1976). Joint-choice model for frequency, destination, and travel mode for shopping trips. *Transportation Research Record* (569).

Anelli, M. and G. Peri (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal 129*(618), 553–602.

Angrist, J. D. and A. B. Krueger (1999). Empirical strategies in labor economics. In *Handbook of labor economics*, Volume 3, pp. 1277–1366. Elsevier.

Anil, B., D. Guner, T. Delibasi, and G. Uysal (2016). Does classroom gender composition affect school dropout? *IZA Discussion Paper*.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1-2), 343–375.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica 73*(5), 1477–1524.

Arcidiacono, P. and M. Lovenheim (2016). Affirmative action and the quality-fit trade-off. *Journal of Economic Literature 54*(1), 3–51.

Arias, O., A. S. Mizala, and F. Meneses (2016). Brecha de género en matemáticas: el sesgo de las pruebas competitivas (evidencia para chile). *Universidad de Chile*.

Athey, S. and G. W. Imbens (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica 74*(2), 431–497.

Aughinbaugh, A., O. Robles, and H. Sun (2013). Marriage and divorce: patterns by gender, race, and educational attainment. *Monthly Labor Review 136*(10), D1–.

Avery, C., O. Gurantz, M. Hurwitz, and J. Smith (2018). Shifting college majors in response to advanced placement exam scores. *Journal of Human Resources 53*(4), 918–956.

Bagde, S., D. Epple, and L. Taylor (2016). Does affirmative action work? Caste, gender, college quality, and academic success in india. *American Economic Review 106*(6), 1495–1521.

Barro, R. J. (2001). Why colleges shouldn't dump the sat. https://www.bloomberg.com/news/articles/2001-04-08/why-colleges-shouldnt-dump-the-sat#xj4y7vzkg. Accessed: 03-09-2022.

Battaglini, M., J. Harris, and E. Patacchini (2022). Interactions with powerful female colleagues promote diversity in hiring. *Journal of labor economics (forthcoming).*

Belley, P., M. Frenette, and L. Lochner (2014). Post-secondary attendance by parental income in the u.s. and canada: Do financial aid policies explain the differences? *Canadian Journal of Economics/Revue canadienne d'économique 47*(2), 664–696.

Belley, P. and L. Lochner (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital 1*(1), 37–89.

Ben-Akiva, M. and S. R. Lerman (1974). Some estimation results of a simultaneous model of auto ownership and mode choice to work. *Transportation 3*(4), 357–376.

Bernhofer, J. and M. Tonin (2022). The effect of the language of instruction on academic performance. *Labour Economics*, 102218.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics 119*(1), 249–275.

Bertrand, M., R. Hanna, and S. Mullainathan (2010). Affirmative action in education: Evidence from engineering college admissions in India. *Journal of Public Economics 94*(1-2), 16–29.

Birch, S. H. and G. W. Ladd (1997). The teacher-child relationship and children's early school adjustment. *Journal of school psychology 35*(1), 61–79.

Björklund, A. and K. G. Salvanes (2011). Education and family background: Mechanisms and policies. Volume 3 of *Handbook of the Economics of Education*, pp. 201–247. Elsevier.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2013). Under pressure? the effect of peers on outcomes of young adults. *Journal of Labor Economics 31*(1), 119–153.

Blanden, J., M. Doepke, and J. Stuhler (2022). Educational inequality. Working Paper 29979, National Bureau of Economic Research.

Blandin, A. and C. Herrington (2022). Family heterogeneity, human capital investment, and college attainment. *American Economic Journal: Macroeconomics (forthcoming)*.

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

Bleemer, Z. (2021). Affirmative action, mismatch, and economic mobility after california's proposition 209. *The Quarterly journal of economics 137*(1), 115–160.

Blundell, R. and M. C. Dias (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources 44*(3), 565–640.

Bobbitt-Zeher, D. (2007). The gender income gap and the role of education. *Sociology of education 80*(1), 1–22.

Boneva, T., M. Golin, and C. Rauh (2022). Can perceived returns explain enrollment gaps in postgraduate education? *Labour Economics 77*, 101998. European Association of Labour Economists, World Conference EALE/SOLE/AASLE, Berlin, Germany, 25 – 27 June 2020.

Booth, A. and E. Yamamura (2018). Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in japan. *Review of Economics and Statistics 100*(4), 581–593.

Booth, A. L., L. Cardona-Sosa, and P. Nolen (2018). Do single-sex classes affect academic achievement? an experiment in a coeducational university. *Journal of Public Economics 168*, 109–126.

Born, A., E. Ranehill, and A. Sandberg (2018). A man´s world?–the impact of a male dominated environment on female leadership. *Working Papers in Economics*.

Bostwick, V. K. and B. A. Weinberg (2022). Nevertheless she persisted? gender peer effects in doctoral stem programs. *Journal of labor economics 40*(2), 397–436.

Braakmann, N. and S. McDonald (2018). Student exposure to socio-economic diversity and students' university outcomes–evidence from english administrative data. *MPRA Paper No. 90351*.

Brown, J. (2011). Quitters never win: The (adverse) incentive effects of competing with superstars. *Journal of Political Economy 119*(5), 982–1013.

Bulman, G. (2017). Weighting recent performance to improve college and labor market outcomes. *Journal of Public Economics 146*, 97–108.

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Buser, T., N. Peter, and S. C. Wolter (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American economic review 107*(5), 125–30.

Cai, X., Y. Lu, J. Pan, and S. Zhong (2019). Gender gap under pressure: Evidence from china's national college entrance examination. *Review of Economics and Statistics 101*(2), 249–263.

Campbell, S., L. Macmillan, R. Murphy, and G. Wyness (2022). Matching in the dark? inequalities in student to degree match. *Journal of labor economics 40*(4), 807–850.

Card, D. and A. A. Payne (2021). High school choices and the gender gap in stem. *Economic inquiry 59*(1), 9–28.

Carrell, S. E., M. E. Page, and J. E. West (2010). Sex and science: how professor gender perpetuates the gender gap. *The Quarterly journal of economics 125*(3), 1101–1144.

Caucutt, E. M. and L. Lochner (2020). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy 128*(3), 1065–1147.

Cavaglia, C., S. Machin, S. McNally, and J. Ruiz-Valenzuela (2021, 01). Gender, achievement, and subject choice in English education. *Oxford Review of Economic Policy 36*(4), 816–835.

Celhay, P. and S. Gallegos (2015). Persistence in the transmission of education: Evidence across three generations for chile. *Journal of Human Development and Capabilities 16*(3), 420–451.

Chavatzia, T. (2017). Cracking the code: girls' and women's education in science, technology, engineering and mathematics (STEM).

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2017). Mobility report cards: The role of colleges in intergenerational mobility. Technical report, national bureau of economic research.

Chetty, R., J. N. Friedman, E. Saez, N. Turner, and D. Yagan (2020, 02). Income Segregation and Intergenerational Mobility Across Colleges in the United States. *The Quarterly Journal of Economics 135*(3), 1567–1633.

Chetty, R. and N. Hendren (2018, 02). The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects. *The Quarterly Journal of Economics 133*(3), 1107–1162.

Chevalier, A., I. E. Isphording, and E. Lisauskaite (2020). Peer diversity, college performance and educational choices. *Labour Economics 64*, 101833.

Chiswick, B. R. (2003). Jacob mincer, experience and the distribution of earnings. *Review of economics of the household 1*(4), 343–361.

Chiswick, B. R. and P. W. Miller (2005). Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of Multilingual and Multicultural Development 26*(1), 1–11.

Clark, D. and H. Royer (2013, October). The effect of education on adult mortality and health: Evidence from britain. *American Economic Review 103*(6), 2087–2120.

Clark Blickenstaff, J. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and education 17*(4), 369–386.

Conlon, J. J. (2019). Major malfunction: A field experiment correcting undergraduates' beliefs about salaries. *Journal of Human Resources*, 0317–8599R2.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American journal of Sociology 106*(6), 1691–1730.

Cortes, K. E. and J. A. Lincove (2016, May). Can admissions percent plans lead to better collegiate fit for minority students? *American Economic Review 106*(5), 348–54.

CRUCH (2019). Ranking de notas. http://sistemadeadmision.consejoderectores.cl/ranking-de-notas. Accessed: 17-04-2019.

Cuberes, D. and M. Teignier (2016). Aggregate effects of gender gaps in the labor market: A quantitative estimate. *Journal of human capital 10*(1), 1–32.

Cullen, J. B., M. C. Long, and R. Reback (2013). Jockeying for position: Strategic high school choice under texas' top ten percent plan. *Journal of public economics 97*, 32–48.

Davies, S. and N. Guppy (1997). Fields of study, college selectivity, and student inequalities in higher education. *Social forces 75*(4), 1417–1438.

Davison, A. C., D. V. Hinkley, et al. (1997). *Bootstrap methods and their application*, Volume 1. Cambridge university press.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

De Zwart, F. (2000). The logic of affirmative action: Caste, class and quotas in India. *Acta Sociologica 43*(3), 235–249.

Deming, D. J. and K. L. Noray (2018). Stem careers and technological change. *NBER Working Paper No. 25065*.

DEMRE (2018). Information for english speakers. https://portaldemre.demre.cl/informacion-para/english-speaker. Accessed: 07-12-2018.

DEMRE (2019). Puntaje ranking. https://demre.cl/psu/proceso-admision/factores-seleccion/puntaje-ranking. Accessed: 01-09-2021.

Diette, T. M. and R. Uwaifo Oyelere (2012). Do significant immigrant inflows create negative education impacts? lessons from the north carolina public school system. *IZA Discussion Paper*.

Durlauf, S. N. and Y. M. Ioannides (2010). Social interactions. *Annual review of economics 2*(1), 451–478.

Dustmann, C., H. Ku, and D. W. Kwak (2018). Why are single-sex schools successful? *Labour Economics 54*, 79–99.

Dynarski, S. M. (2003, March). Does aid matter? measuring the effect of student aid on college attendance and completion. *American Economic Review 93*(1), 279–288.

Edwards, D., H. Coates, and T. Friedman (2012). A survey of international practice in university admissions testing. *Higher Education Management and Policy 24*(1), 1–18.

Epple, D., R. Romano, and H. Sieg (2008). Diversity and affirmative action in higher education. *Journal of Public Economic Theory 10*(4), 475–501.

Epple, D. and R. E. Romano (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*, Volume 1, pp. 1053–1163. Elsevier.

Fischer, M. J. and D. S. Massey (2007). The effects of affirmative action in higher education. *Social Science Research 36*(2), 531–549.

Francis, A. M. and M. Tannuri-Pianto (2012a). The redistributive equity of affirmative action: Exploring the role of race, socioeconomic status, and gender in college admissions. *Economics of Education Review 31*(1), 45–55.

Francis, A. M. and M. Tannuri-Pianto (2012b). Using Brazil's racial continuum to examine the short-term effects of affirmative action in higher education. *Journal of Human Resources 47*(3), 754–784.

Friedman, S. (2020). *The class ceiling : why it pays to be privileged.* Bristol, UK: Policy Press.

Frisancho, V. and K. Krishna (2016). Affirmative action in higher education in India: targeting, catch up, and mismatch. *Higher Education 71*(5), 611–649.

Fuller, W. C., C. F. Manski, and D. A. Wise (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources*, 477–498.

Geay, C., S. McNally, and S. Telhaj (2013). Non-native speakers of english in the classroom: what are the effects on pupil performance? *The Economic Journal 123*(570), F281–F307.

Gerber, T. P. and S. Y. Cheung (2008). Horizontal stratification in postsecondary education: Forms, explanations, and implications. *Annual Review of Sociology 34*, 299–318.

Glaeser, E. L., G. A. Ponzetto, and A. Shleifer (2007). Why does democracy need education? *Journal of economic growth 12*(2), 77–99.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The quarterly journal of economics 118*(3), 1049–1074.

Gould, E. D., V. Lavy, and M. Daniele Paserman (2009). Does immigration affect the long-term educational outcomes of natives? quasi-experimental evidence. *The Economic Journal 119*(540), 1243–1269.

Grossman, M. (2006). Education and nonmarket outcomes. Volume 1 of *Handbook of the Economics of Education*, pp. 577–633. Elsevier.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American Economic Review 84*(3), 622–641.

Guiso, L., F. Monte, P. Sapienza, and L. Zingales (2008). Diversity: Culture, gender, and math. *Science (American Association for the Advancement of Science) 320*(5880), 1164–1165.

Hamermesh, D. S. and S. J. Trejo (2000, 02). The Demand for Hours of Labor: Direct Evidence from California. *The Review of Economics and Statistics 82*(1), 38–47.

Hanushek, E. A., J. Hanna, P. E. Peterson, L. M. Talpey, and L. Woessmann (2019). The achievement gap fails to close: Half century of testing shows persistent divide between haves and have-nots. *Education next 19*(3), 9–17.

Hanushek, E. A. and L. Woessmann (2011). The economics of international differences in educational achievement. Volume 3 of *Handbook of the Economics of Education*, pp. 89–200. Elsevier.

Hearn, J. C. (1991). Academic and nonacademic influences on the college destinations of 1980 high school graduates. *Sociology of education 64*(3), 158–171.

Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The economics and econometrics of active labor market programs. In *Handbook of labor economics*, Volume 3, pp. 1865–2097. Elsevier.

Hernán, M. A. and J. M. Robins (2016). *Causal inference.* Boca Raton.

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *Review of Economics and Statistics 94*(3), 712–722.

Hoogendoorn, S., H. Oosterbeek, and M. Van Praag (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science 59*(7), 1514–1528.

Hoogendoorn, S. and M. Van Praag (2012). Ethnic diversity and team performance: A field experiment. *Tinbergen Institute Discussion Paper*.

Howell, J. S. (2010). Assessing the impact of eliminating affirmative action in higher education. *Journal of Labor Economics 28*(1), 113–166.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

Iriberri, N. and P. Rey-Biel (2019). Competitive pressure widens the gender gap in performance: Evidence from a two-stage competition in mathematics. *The Economic Journal 129*(620), 1863–1893.

Jensen, P. and A. W. Rasmussen (2011). The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review 30*(6), 1503–1515.

Joensen, J. S. and H. S. Nielsen (2016). Mathematics and gender: Heterogeneity in causes and consequences. *The Economic Journal 126*(593), 1129–1163.

Justman, M. and S. J. Méndez (2018). Gendered choices of stem subjects for matriculation are not driven by prior differences in mathematical achievement. *Economics of education review 64*, 282–297.

Kellogg, R. and H. Wolff (2008). Daylight time and energy: Evidence from an australian experiment. *Journal of Environmental Economics and Management 56*(3), 207–220.

Kolawole, E. (2008). Effects of competitive and cooperative learning strategies on academic performance of nigerian students in mathematics. *Educational Research and Reviews 3*(1), 033–037.

Kolovich, L., V. Malta, M. Newiak, and D. Robinson (2020). Gender equality and macroeconomic outcomes: evidence and policy implications. *Oxford review of economic policy 36*(4), 743–759.

Kuhn, P. and C. Weinberger (2005). Leadership skills and wages. *Journal of Labor Economics 23*(3), 395–436.

Lance, L. (2011). Nonproduction benefits of education: Crime, health, and good citizenship. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of The Economics of Education*, Volume 4 of *Handbook of the Economics of Education*, pp. 183–282. Elsevier.

Lavy, V. and E. Sand (2018). On the origins of gender gaps in human capital: Short- and long-term consequences of teachers' biases. *Journal of public economics 167*, 263–279.

Lavy, V. and A. Schlosser (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

Lazear, E. P. (1999). Culture and language. *Journal of political Economy 107*(S6), S95–S126.

Lechner, M. et al. (2011). *The estimation of causal effects by difference-in-difference methods*. Now Hanover, MA.

Li, Q., W. Zang, and L. An (2013). Peer effects and school dropout in rural china. *China Economic Review 27*, 238–248.

Lihamba, A., R. Mwaipopo, and L. Shule (2006). The challenges of affirmative action in tanzanian higher education institutions: A case study of the University of Dar es Salaam, Tanzania. In *Women's Studies International Forum*, Volume 29, pp. 581–591. Elsevier.

Lleras-Muney, A. (2005, 01). The relationship between education and adult mortality in the united states. *The Review of Economic Studies 72*(1), 189–221.

Lloyd, K. M., K. T. Leicht, and T. A. Sullivan (2008). Minority college aspirations, expectations and applications under the Texas top 10% law. *Social Forces 86*(3), 1105–1137.

Long, M. C. (2007). Affirmative action and its alternatives in public universities: What do we know? *Public Administration Review 67*(2), 315–330.

Long, M. C. and M. Tienda (2008). Winners and losers: Changes in texas university admissions post-hopwood. *Educational evaluation and policy analysis 30*(3), 255–280.

Machin, S. and S. McNally (2007). Tertiary education systems and labour markets. *Education and Training Policy Division, OECD 6*.

Machin, S. and P. A. Puhani (2003). Subject of degree and the gender wage differential: evidence from the uk and germany. *Economics Letters 79*(3), 393–400.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

McGrath, C. H. and M. Frearson (2016). Admission systems and student mobility: a proposal for an eu-wide registry for university admission. *Foro de Educación 14*(21), 167–195.

McGuigan, M., S. McNally, and G. Wyness (2016). Student awareness of costs and benefits of educational decisions: Effects of an information campaign. *Journal of human capital 10*(4), 482–519.

McNally, S. (2020). Gender differences in tertiary education: what explains stem participation? Technical report, IZA Policy Paper.

Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics 13*(2), 151–161.

Milligan, K., E. Moretti, and P. Oreopoulos (2004). Does education improve citizenship? evidence from the united states and the united kingdom. *Journal of Public Economics 88*(9), 1667–1695.

Mincer, J. (1958). Investment in human capital and personal income distribution. *The Journal of political economy 66*(4), 281–302.

Montolio, D. and P. A. Taberner (2021). Gender differences under test pressure and their impact on academic performance: a quasi-experimental design. *Journal of Economic Behavior & Organization 191*, 1065–1090.

Nie, N. H., J. Junn, K. Stehlik-Barry, et al. (1996). *Education and democratic citizenship in America*. University of Chicago Press.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The quarterly journal of economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives 24*(2), 129–44.

Niu, S. X., M. Tienda, and K. Cortes (2006). College selectivity and the texas top 10% law. *Economics of education review 25*(3), 259–272.

Nollenberger, N., N. Rodríguez-Planas, and A. Sevilla (2016, May). The math gender gap: The role of culture. *American Economic Review 106*(5), 257–61.

OECD (2018). *Education at a glance 2018.* OECD, Paris, France.

OECD (2019). *PISA 2018 Results (Volume I).* OECD, Paris, France.

OECD (2020a). *Education at a Glance 2020.* OECD, Paris, France.

OECD (2020b). *International Migration Outlook 2020.* OECD, Paris, France.

OECD (2021a). *Employment Outlook 2021.* OECD, Paris, France.

OECD (2021b). *What are the earnings advantages from education?* OECD, Paris, France.

Onsongo, J. (2009). Affirmative action, gender equity and university admissions–Kenya, Uganda and Tanzania. *London review of education 7*(1), 71–81.

Oosterbeek, H. and R. Van Ewijk (2014). Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review 38*, 51–63.

Oreopoulos, P. and K. G. Salvanes (2011, March). Priceless: The nonpecuniary benefits of schooling. *Journal of Economic Perspectives 25*(1), 159–84.

Oswald, A. J. and S. Wu (2011, 11). Well-being across america. *The Review of Economics and Statistics 93*(4), 1118–1134.

Papay, J. P., R. J. Murnane, and J. B. Willett (2016). The impact of test score labels on human-capital investment decisions. *Journal of Human Resources 51*(2), 357–388.

Patterson, R. W., N. G. Pope, and A. Feudo (2021). Timing matters: Evidence from college major decisions. *The Journal of human resources*, 820–.

Powdthavee, N., W. N. Lekfuangfu, and M. Wooden (2015). What's the good of education on our overall quality of life? a simultaneous equation model of education and life satisfaction for australia. *Journal of Behavioral and Experimental Economics 54*, 10–21.

QS-World-Ranking (2019). Qs world university rankings. `https://www.topuniversities.com/university-rankings/world-university-rankings/2019`. Accessed: 03-08-2021.

Rapoport, B. and C. Thibout (2018). Why do boys and girls make different educational choices? the influence of expected earnings and test scores. *Economics of Education Review 62*, 205–229.

Reimer, D., C. Noelke, and A. Kucel (2008). Labor market effects of field of study in comparative perspective: An analysis of 22 european countries. *International journal of comparative sociology 49*(4-5), 233–256.

Rothstein, J. M. (2004). College performance predictions and the sat. *Journal of Econometrics 121*(1), 297–317. Higher education (Annals issue).

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education*, Volume 3, pp. 249–277. Elsevier.

Schleicher, A. (2019). Pisa 2018: Insights and interpretations. *OECD Publishing*.

Schlosser, A., Z. Neeman, and Y. Attali (2019). Differential performance in high versus low stakes tests: evidence from the gre test. *The Economic Journal 129*(623), 2916–2948.

Schneeweis, N. (2015). Immigrant concentration in schools: Consequences for native and migrant students. *Labour Economics 35*, 63–76.

Schneeweis, N. and M. Zweimüller (2012). Girls, girls, girls: Gender composition and female school choice. *Economics of Education review 31*(4), 482–500.

Shan, X. (2022). The minority trap: Minority status drives women out of male-dominated fields. *Unpublished manuscript*.

Sidanius, J. and R. C. Veniegas (2013). Gender and race discrimination: The interactive nature of disadvantage. In *Reducing prejudice and discrimination*, pp. 57–80. Psychology Press.

STEM Women (2021). Women in stem — percentages of women in stem statistics. `https://www.stemwomen.co.uk/blog/2021/01/women-in-stem-percentages-of-women-in-stem-statistics`. Accessed: 01-09-2021.

Stoddard, O., C. Karpowitz, and J. Preece (2020). Strength in numbers: A field experiment in gender, influence, and group dynamics. *IZA Discussion Paper*.

Teigen, M. (2000). The affirmative action controversy. *NORA : Nordic journal of women's studies 8*(2), 63–77.

Turner, S. E. and W. G. Bowen (1999). Choice of major: The changing (unchanging) gender gap. *Industrial and labor relations review 52*(2), 289.

UCH (2019). Programa de equidad de género. `http://www.uchile.cl/portal/presentacion/asuntos-academicos/pregrado/admision-especial/96722/ingreso-prioritario-de-equidad-de-genero-peg`. Accessed: 17-04-2019.

UCH Engineering (2019). Engineering school facts and numbers. `https://www.topuniversities.com/university-rankings/world-university-rankings/2019`. Accessed: 03-08-2021.

UNESCO (2022a). *Higher education global data report.* UNESCO, Paris, France.

UNESCO (2022b). *Scoping Progress in Education (SCOPE).* UNESCO, Paris, France.

Valenzuela, J. P., C. Bellei, and D. d. l. Ríos (2014). Socioeconomic school segregation in a market-oriented educational system. the case of chile. *Journal of education Policy 29*(2), 217–241.

Van de Werfhorst, H. (2008). Educational fields of study and european labor markets: Introduction to a special issue. *International journal of comparative sociology 49*(4-5), 227–231.

Wiswall, M. and B. Zafar (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies 82*(2), 791–824.

Wiswall, M. and B. Zafar (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics 133*(1), 457–507.

World Bank (2022). Gender statistics. `https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS?locations=OE`.

Xie, Y., M. Fang, and K. Shauman (2015). Stem education. *Annual review of sociology 41*, 331–357.

Yelowitz, A. S. (1995, 11). The Medicaid Notch, Labor Supply, and Welfare Participation: Evidence from Eligibility Expansions. *The Quarterly Journal of Economics 110*(4), 909–939.

Zafar, B. (2013). College major choice and the gender gap. *Journal of Human Resources 48*(3), 545–595.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review 109*(1), 1–47.