



Addressing Inequalities in Education: Need-Based Grants, Gender Differences and School Choice

José Montalban Castilla

► To cite this version:

José Montalban Castilla. Addressing Inequalities in Education: Need-Based Grants, Gender Differences and School Choice. Economics and Finance. École des hautes études en sciences sociales (EHESS), 2019. English. NNT: 2019EHES0155 . tel-03168305

HAL Id: tel-03168305

<https://theses.hal.science/tel-03168305>

Submitted on 12 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École des Hautes Études en Sciences Sociales

École doctorale n° 465 Économie Panthéon Sorbonne

Paris-Jourdan Sciences Économiques

DOCTORAT

Discipline: Analyse et Politique Économiques

Jose MONTALBAN CASTILLA

Identifier les inégalités dans le système éducatif: bourses sur critères sociaux, différences de genre et choix d'établissement scolaire

Thèse dirigée par : Julien GRENET

Date de soutenance : le 20 novembre 2019

Rapporteurs : Ghazala AZMAT, Professeur à la Sciences Po
Olmo SILVA, Professeur à la London School of Economics

Jury : Caterina CALSAMIGLIA, Professeur ICREA à IPEG
Gabrielle FACK, Professeur à Université Paris-Dauphine-PSL
Marc GURGAND, Directeur de recherche CNRS, Professeur à PSE

Directeur : Julien GRENET, Chargé de recherche CNRS, Professeur associé à PSE

École des Hautes Études en Sciences Sociales

Doctoral school n° 465 Économie Panthéon Sorbonne

Paris-Jourdan Sciences Économiques

PHD THESIS

Discipline: Economics

Jose MONTALBAN CASTILLA

Addressing Inequalities in Education: Need-Based Grants, Gender Differences and School Choice

Thesis supervised by: Julien GRENET

Date of defense: November 20th, 2019

Referees: Ghazala AZMAT, Professor at the Sciences Po
Olmo SILVA, Professor at the London School of Economics

Jury : Caterina CALSAMIGLIA, ICREA Research Professor at IPEG
Gabrielle FACK, Professor at Dauphine University-PSL
Marc GURGAND, Senior researcher CNRS, Professor at PSE

Supervisor: Julien GRENET, Full time researcher CNRS, Professor at PSE

Acknowledgements

I am incredibly grateful to my primary advisor, Julien Grenet. I first met Julien in his cozy office of the old PSE buildings to discuss my Master Thesis. After a couple of minutes, I was utterly convinced that I wanted him to be my primary supervisor. After more than four years of supervision, I can confidently claim that this is the best decision I have taken in my (still short) academic career. I am delighted that he accepted this challenging task. Since we met, he has guided me through the darkened path that every amateur researcher finds when it starts to face research for the first time. Shaping my mind for empirical analysis, he has taught me to overcome every problem that came along the way on the foggy path to achieve enlightenment. The way Julien does research has inspired and motivated me to the best of my level. He has a complete knowledge of the technical and theoretical methods that combines with an astonishing expertise on the empirical side. Guided by the academic curiosity that must govern the soul of every researcher, but only a few of them bear, he is always motivated to find both policy implications and academic contributions of every project, with the primary goal of improving society. Most importantly, he has shown to me his complete honesty in all research that he undertakes with the only interest of finding the truth. All of this and much more makes him my (one and only) role model for what I wish to become some say as a researcher. He has been amazingly devoted, patient, and kind in all his Ph.D. supervision. He has always been generous with his time, offering assistance whenever I needed it, and never asking anything in return. He has read my papers as the toughest referee, providing detailed comments and insightful suggestions. He has perfectly balanced in-depth monitoring with complete autonomy to undertake my research projects,

which was essential for my development as a researcher. We passed together through the uncertain path of the academic job market (a deeply complicated moment in which even the most consolidated friends, co-authors or marriages break up)...and surprisingly, after reading my job market paper thousands of times, he still wants to talk to me! He was extremely encouraging during all the job market process, offering the best of his knowledge and contacts to achieve the unique goal of helping me to get the placement that I wanted (and he made it). Besides, he has also been a good friend to chat, laugh, and talk about life. Wherever I will be, and whomever I will become, he will always be my master. I owe him all that you are about to read in this dissertation (bad jokes are entirely mine).

I am also indebted to Marc Gurgand, my co-supervisor and member of my Ph.D. committee. He has provided me with invaluable feedback and support throughout my Ph.D. In my papers, I always try to anticipate most of the questions that I would have, thinking about a valid answer for each of them. I must confess that Marc always makes brilliant comments that I have never thought in advance, pushing me to the limits of my mind (which he probably thinks that they are quite limited due to my high non-response rate) to improve all my papers dramatically. He also supported me through the job market with outstanding suggestions for my job market paper and unconditionally supporting me with a reference letter.

Moreover, my gratitude goes to the four other members of my thesis jury: Ghazala Azmat, Olmo Silva, Gabrielle Fack, and Caterina Calsamiglia. Ghazala has always been incredibly supportive and interested in my research ideas. The first thing she told me when she knew I was going to the job market was “congratulations, that’s great news”. In the context of the academic job market, in which people are extremely negative about this experience, she made me felt comfortable, motivated, and confident of myself. She provided me key feedback to significantly improve my job market paper. She also had the disgrace (or the privilege) to be the first person that listened to me in an improvised “mock interview” in her office. The fact that she agreed to be part of my dissertation jury after listening to this nonsense speech proves her endless constructive personality. Olmo has been one of my key mentors

during my time in London. Since the very first day, he was interested in my research agenda and went to my seminars at LSE. Apart from being one of the fastest eaters of “Chicken Katsu” that I have ever seen, Olmo has made me an incredible amount of outstanding suggestions to improve all my papers. He has greatly helped me through all the job market process with constant support. Gabrielle has been one of the people that I have disturbed the most during my Ph.D. She came to most of my seminars at PSE, providing me with guidance and excellent feedback. Gabrielle had the misfortune to read my ugly Stata codes during my Master Thesis. I hope that this dissertation compensates, at least partially, the unlucky moments that she suffered trying to understand my messy codes. Caterina has always been extremely positive and encouraging with all my research ideas. When I first met her at CEMFI, she was supportive of all my projects, and moreover, fully transparent and available to start a joint project someday. She has provided me with exceptional comments and incredibly valuable advice during my Ph.D.

I would also want to thank my referees in the academic job market: Antonio Cabrales and Steve Machin. Antonio has followed me throughout my professional career. We first met when I was working as a research analyst at FEDEA. We had also crossed our paths when he was the director of the Department of Economics at Carlos III University and University College of London. He has always been supportive and helpful in improving all my research ideas. He has followed me very carefully, and he has always been available to meet when I needed it. I believe that Antonio is one of the most talented economists that academia had, on the professional but also the personal side. Steve was the director of the CEP during my time at LSE. I am really grateful for all his fantastic help and feedback during my time at CEP and the job market. I am also indebted to my co-authors on my dissertation projects. Almudena Sevilla for her joint work on the second chapter of this thesis, and Lucas Gortázar and David Mayor for their fantastic collaboration in the third chapter.

During my Ph.D., I benefited from the financial assistance that was crucial to undertake my dissertation. I would like to especially thank Fundación Ramón Areces

and Bank of Spain from their financial support. I am extremely grateful to people that in many ways, have made this dissertation possible, helping me to get access to the different administrative data that are exploited. I am especially grateful to Ismael Sanz, Luís Pires, Gerardo Azor, Paloma Arnaiz Tovar, Elena Zamarro Parra, Gloria del Rey Gutiérrez, María Rosario Ronda and Luis Losada Romo for their invaluable help in collecting the data from the Region of Madrid and Carlos III University.

PSE is a wonderful environment to develop as a researcher. In particular, the quality of all the economists working on applied microeconomics is really inspiring and engaging. I cannot imagine a better place to grow as a Ph.D. student. I would like to make a special mention to those professors who get involved and engaged to Ph.D. students: Sylvie Lambert, Karen Macours, Jean Marc Tallon and David Margolis. Their kind dedication to improving Ph.D. students' lives is impressive and encouraging. I would also like to thank an invaluable pillar of PSE, Véronique Guillotin. She provides assistance of any kind and is always available when most needed. I am heartily grateful for her help and support, especially during the job market. Her office has been (and will always be) my home when all things go wrong.

I would like to especially thank four of my PSE best friends that have become part of my family during all these years. I would not have started my Ph.D. without the support of Clara Martínez-Toledano. We have shared the most valuable professional events: The MA at Carlos III University, Master in Public Policy and Development at PSE, Ph.D. at PSE, visiting LSE and Bank of Spain, plus getting grants from Fundación Ramón Areces and Bank of Spain. She has been the best college and friend that one can dream of. Talking to her every day and commenting on the daily news that happened in our lives was one of the most pleasant things of my Ph.D. Clara is a true friend, that has always been on my side in the good moments, but also the best support in the bad times. She is a rising star economist with a bright future. Martín Fernández-Sánchez is one of my best friends and most valuable co-author. He is one of the most talented economists that I know. He transmits his life passions in a way that spills over, making me feel energetic and eager to give my best. We

have shared many great moments, and I am looking forward to more. I am sure that he will accomplish every challenge that he will face in his life, except winning me at ping pong, for which I will always be his “black beast”. Laura Khury has been one of my main pillars in the job market and life. We have shared office since the very beginning, and being part of the same group of friends. All the experiences we have passed together, such as to share microphone for singing Tina Turner in a 24 hours karaoke in Atlanta, have made us unbreakable. Her talent as an economist will shine brightly on the wonderful NHH. Alessandro Tondini is a great friend that always takes care of me and has also been crucial in the job market. Even though he was a tough objector to my suits, he has proven to be one of the most helpful and amazing friends during my Ph.D. Apart from being a great econometrician, his passion for culture is probably about to give birth to an outstanding writer. To the four of them: Thanks for making this experience funny, pleasant, and enjoyable. I would like to make a special mention to my beautiful friends at PSE: Mariona Segú (*Puyola*), Clément Brébion (Clemente), Paul Brandily (worthy descendant of Charlemagne), Sofía Dromundo (guardian of Pepa), Marion Leroutier, Marion Monnet, Paul Dutronc, Francesco Filippucci, Miguel Uribe, Hubertus Wolff, Daniel Waldenström, and Paolo Santini.

I have benefited from a year-long stay at the CEP at LSE and a summer visiting at Bank of Spain. I have learned to the best of my level from these experiences. These institutions have improved me in both the research and the personal aspect. The quality of their economists makes these organizations enriching and dynamic places for every researcher that wishes to learn to the best of its abilities. I would like to thank Sandra McNally for hosting me at the CEP. She has always been extremely kind and provided me with fantastic feedback. Unique thoughts to my beloved friends: Jenny Ruiz-Valenzuela (the boss), Matteo Sandi (the *torero* of the CEP), Claudia Robles-García (skyrocket economist), Carmen Villa Llera (majority shareholder of the “George IV”), Maria Sanchez-Vidal and Rosa Sanchis-Guarner (map artist). I would also like to thank Kilian Russ, Guglielmo Ventura, Ria Ivandic, Vincenzo Scrutinio, Chiara Cavaglia, Nicolas Chanut, and Andres Barrios. I am

also grateful to my friends that organized the series of conferences entitled “What’s Next for Spain and Catalonia?” at LSE. I tremendously enjoyed working with them. I am sure that all of them have a bright professional future. Still, on top of everything, they are my friends: Lu s Cornago, Javier Padilla, Sergio Olalla, Enrique Chueca, Elisabet Vives, Javier Carbonell, and Berna Le n. In the Bank of Spain, I am grateful to my co-authors and friends, Brindusa Anghel, Laura Crespo, and Laura Hospido. Moreover, I am indebted to all great economists that helped me with their outstanding comments, making me feel at home: Aitor Lacuesta, Mario Alloza, Federico Tagliati, Sergio Puente, Enrique Moral, Ernesto Villanueva, Cristina Barcel , Juanfran Jimeno, Carlos Sanz, Olympia Bover, Yuliya Kulikova, and Elena Vozmediano.

I would also like to thank my true friends. I know some of them since I was a teenager: Rafa, Jorge, Aitor, Candela, Yendy, and Sonia. Some others I have met along these amazing years, and now they are best friends: Alex Buesa, Roberto Schiano, Sebasti n Franco, Raquel P ramo, Laia Varela, Borja Ortuzar, Irene L pez and Mar a Caparros.

I am indebted to my family. My father (Paco), my mother (Amelia), and my little (now not so little) brother, Pablo (currently enjoying the wonders that happen *sous le ciel de Paris*). They have always supported me in all the decisions I have made. They have offered me the assistance of a good friend, the honesty and advice of the wisest person, and the warm home that provides family. They have transmitted me with a passion for culture and knowledge. Wherever they are, I will always be at home. I would also like to mention other members of my family that have helped me, and gently took care of me along this way: Mari, Pepe, Mariajo, Ruf , Luc a, Yurema, Martita, and Manolo.

Finally, last but not least, I am incredibly grateful to Raquel Recio L pez. She has been my constant support during all these years. She always believed in my capacities and provided me with moral support when most needed. She has always been sympathetic to the tough particularities of the academic world, pushing me to the best of my abilities and always encouraging me to follow my dreams. She is the

girl from Ipanema, Layla, Marianne, *Il cielo in una stanza*, and Rachael (one more kiss, dear). La chica de ayer, de hoy y de siempre.

José Montalbán Castilla

October 3rd 2019,

Stockholm

Summary

This dissertation gathers evidence on three sources of education inequalities across different education levels (preschool, primary, secondary, and higher education) in the context of Spain. It revolves around the causal effects of large-scale educational policies on the efficiency and equity of educational systems.

The first chapter focuses on the effects of financial aid for disadvantaged students in the context of higher education. National financial aid programs for disadvantaged students cover a large fraction of college students and represent a non-negligible component of the public budget. These programs often have weak performance requirements for renewal, potentially leading to moral hazard and efficiency losses. Using a reform in the Spanish need-based grant program in higher education, this paper tests the causal effect of receiving the same amount of grant under different intensities of academic requirements on student performance, degree completion and student dropout. I use administrative micro-data on the universe of applicants to the grant in a large university. Exploiting sharp discontinuities in the grant eligibility formula, I find strong positive effects of being eligible for a grant on student performance when combined with demanding academic requirements, while there are no effects on student dropout. Students improve their final exam attendance rate, their average GPA in final exams, and their probability of completing the degree. They also reduce the fraction of subjects that they have to retake. The grant has no effects on student performance when academic requirements are low and typically comparable to those set out by national need-based student aid programs around the world. These results suggest that academic requirements in the context

of higher education financial aid can be an effective tool to help overcome moral hazard concerns and improve aid effectiveness.

The second chapter centers on the gender differences in academic performance due to the testing-environment, in the context of primary and secondary education. There is a substantial body of literature that focuses on measuring how gender differences in cognitive abilities and gender-stereotyping norms impact the gender gap in student performance. However, little attention has been devoted to investigating how the organization of student testing may influence the relative performance of male and female students. This paper analyzes the gender gap in test scores that arises as a result of differential responses by boys and girls to the testing environment. To that end, we exploit a unique randomized intervention on the entire population of students in the 6th and 10th grades in the Region of Madrid (Spain). The intervention assigned schools to either internally or externally administered testing. We find that girls do worse than boys in exams that are externally administered, especially in male-dominated subjects. Additional survey evidence on stress, self-confidence, and effort suggests that lower relative female performance in externally administered tests results from a lower ability to cope with stressful situations as a result from less familiarity with the testing environment.

The third chapter studies the relationship between school choice priorities and school segregation in the context of preschool education. Most of the empirical literature on market design has focused on the relative performance and strategic implications of alternative matching mechanisms, taking the inputs of school choice -preferences, priorities, and capacities- as exogenous. This work aims at broadening the scope of market design questions to school choice by examining how government-determined school choice priorities affect families' choices and pupil sorting across schools in the context of the Boston Mechanism. We use two large-scale school choice reforms in the school choice priority structure undertaken in the region of Madrid (Spain) as a source of variation. In particular, we exploit an inter-district school choice reform

that widely expanded families' choice set of schools. We combine an event study first difference across cohorts and a Difference-in-Difference design to identify the impact of the reforms. Using unique administrative data on parents' applications to schools, this paper shows that families reacted to the reform exerting higher inter-district choice and applying to schools located further away from home than before the reform. We find distributional effects of the reform concluding that parents from the highest education levels and parents of non-immigrant students were those who reacted the most in absolute terms. Interestingly, results support the idea of potential information gaps and the dynamic learning process across immigrant status groups. We find a decrease in school segregation by parental education and an increase in school segregation by immigrant status -though effects on the latest fade out when controlling for residential stratification. Results suggest that when parents' school choices exhibit a strong degree of polarization by social and immigrant background, priority structures need to be carefully designed to achieve diversity objectives.

Keywords: Economics of Education; Need-based Grants; Gender Differences; School Choice; School Segregation.

Résumé

Cette thèse rassemble des recherches sur trois sources d'inégalités éducatives en Espagne, à différents niveaux d'éducation (préscolaire, primaire, secondaire et supérieur). Ces recherches tentent d'évaluer l'impact causal de politiques de grande ampleur sur l'efficacité et l'équité des systèmes éducatifs.

Le premier chapitre porte sur les effets de l'aide financière aux étudiants défavorisés, dans le contexte de l'enseignement supérieur. Les programmes nationaux d'aide financière (bourses) aux étudiants défavorisés couvrent une grande partie des étudiants et représentent une part non négligeable des budgets alloués à l'enseignement supérieur. Dans leur version la plus commune, ces programmes comportent des critères d'attribution peu exigeants en termes de performance académique, générant ainsi un risque d'aléa moral et de perte d'efficacité. En s'intéressant à une réforme du système de bourses en Espagne qui modifie les exigences académiques nécessaires pour bénéficier des aides, on tente d'identifier l'effet de ces exigences académiques, à niveau de bourse donné, sur les performances des étudiants, leur propension à abandonner ou réussir leurs études. Pour cela, on utilise des micro-données administratives sur l'univers des candidats aux bourses dans une grande université. En exploitant les discontinuités dans la formule d'éligibilité, on trouve des effets positifs marqués de l'association d'exigences académiques élevées aux bourses, sans aucun effet négatif sur le décrochage. Les étudiants sont plus fréquemment présents aux examens de fin d'année, leur moyenne générale comme leur taux de réussite y sont plus élevés, et ils ont moins souvent besoin de passer des rattrapages. En revanche, des exigences académiques plus faibles (et ainsi plus semblables aux critères généralement observés

dans le monde) n'ont aucun effet. Ces résultats suggèrent qu'associer des exigences académiques fortes peut permettre d'accroître l'efficacité des bourses et de lutter contre l'aléa moral.

Le deuxième chapitre s'intéresse à l'impact des conditions d'examen, dans le primaire et le secondaire, sur les performances académiques selon le genre. Il existe une littérature conséquente sur l'écart de performance entre les étudiants de genres différents, étudiant des causes aussi diverses que les capacités cognitives ou les normes stéréotypées. Les conditions d'examen académique elles-mêmes ont néanmoins fait l'objet de peu d'attention dans cette littérature. Ce chapitre s'intéresse précisément aux différences de performance entre garçons et filles qui apparaissent lorsque l'environnement de l'examen est modifié. Pour cela, ce chapitre analyse une intervention aléatoire concernant l'ensemble des élèves de 6e et 10e années dans la région de Madrid (Espagne). Les écoles de la région ont été assignées de façon aléatoire dans deux groupes devant administrer leurs examens de façon différente : en interne, où les élèves étaient évalués par des enseignants de l'école vs. en externe, où des enseignants extérieurs venaient effectuer les tests. Lorsque l'examineur est externe, les filles performant moins bien que les garçons, notamment dans les matières où elles performant généralement moins bien à la base. Les données d'une enquête additionnelle sur le stress, la confiance en soi et le degré d'effort indiquent que ces performances relatives découlent d'une moins bonne gestion du stress provoqué par un environnement non familier.

Le troisième chapitre étudie la relation entre choix de l'école et ségrégation scolaire, dans le contexte de l'éducation préscolaire. L'essentiel de la littérature empirique sur le market design a mis l'accent sur les performances relatives des différentes procédures dematching (ainsi que sur les comportements stratégiques qu'elles impliquent) en considérant les déterminants du choix de l'école (préférences, priorités, capacités) comme exogènes. Ce chapitre vise à élargir cette littérature en étudiant comment la régulation publique affecte les décisions des familles et la

ségrégation des enfants entre les écoles, dans le cadre du Mécanisme de Boston. Ce chapitre analyse deux réformes à grande échelle, entreprises dans la région de Madrid (Espagne), visant à modifier les mécanismes de choix des écoles (liste des vœux) par les familles. En particulier, on exploite ici une réforme interdistricts qui a largement élargi l'univers des choix possibles pour les familles. Son impact est mesuré en combinant une *event study first difference* entre cohortes avec une différence de différences. Grâce à des données uniques sur les candidatures des familles auprès des différentes écoles, ce chapitre montre que les familles ont candidaté à des écoles plus éloignées de leur domicile. La réforme produit des effets différenciés, avec les parents les plus éduqués et n'étant pas immigrés réagissant le plus fortement (en termes absolus). De façon intéressante, les résultats favorisent l'idée d'un écart de connaissance du système entre immigrés et non-immigrés se réduisant au fur et à mesure du temps à travers un apprentissage graduel. Les résultats indiquent un déclin de la ségrégation par niveau d'éducation des parents mais une croissance de cette ségrégation entre enfant de parents immigrés et non-immigrés (même si cet effet disparaît lorsque la stratification résidentielle est prise en compte). Ce chapitre suggère ainsi que, lorsque les choix d'école par les parents sont fortement polarisés par le niveau d'éducation ou les origines, les systèmes d'expression des préférences doivent être soigneusement conçus si l'on souhaite obtenir des résultats en termes de diversité.

Mots-clés: Economie de l'éducation; Subventions basées sur les besoins; Différences entre les sexes; École Choix; Ségrégation scolaire.

General Introduction

“Puesto que vivimos en pleno misterio, luchando contra fuerzas desconocidas, tratemos en lo posible de esclarecerlo. No nos desaliente la consideración de la pobreza de nuestro esfuerzo ante los magnos e innumerables problemas de la vida.”

– Ramón y Cajal.

The economics of education finds its roots on the “human capital revolution” of [Becker \(1962\)](#), in which education is modeled as a consumption good (wished for the consumer itself) and an investment (wished for the money it brings). The human capital theory emerged in the 60s with two parallel trends: The Growth Theory and the Microeconomic Theory of demand for education, finding their empirical origins on the Solow residual and wage differentials, respectively. Over the past decades, empirical and theoretical evidence has proven that education is a crucial path to upward socioeconomic mobility and economic growth. There are substantial non-pecuniary, and pecuniary returns to education. On the one hand, positive links (with at least a part of them causal) have been identified concerning health outcomes, fertility rates, occupational choice, consumption/saving patterns and participation in civic life ([Lance, 2011](#); [Oreopoulos and Salvanes, 2011](#)). On the other hand, both correlational and causal empirical evidence conveys a clear message: There are substantial pecuniary returns to invest in education, in particular to higher education ([Mincer, 1974](#); [Kane and Rouse, 1995](#); [Kirkeboen, Leuven and Mogstad, 2016](#); [Bhuller, Mogstad and Salvanes, 2017](#)).

Access to higher education has massively grown in the last decades. Over the period 1995 to 2014, the percentage of young adults who enter university in OECD

countries increased from 37 to 59 percent (OECD, 2016). Despite this general increase of participation in tertiary education, there are still substantial differences in college entry, persistence, and graduation across socioeconomic groups. Even though low-income students have benefited from the increase in college entry over the last decades, the absolute rise in college access and completion rates have been much lower for them in comparison with their peers of the general population. Young adults with college-educated parents are still more than double as likely to be in college compared with their peers with low-educated parents (OECD, 2016). College graduation trends have followed a similar pattern. Bailey and Dynarski (2011) shows that, in the US, *rates of college completion increased by only four percentage points for low-income cohorts born around 1980 relative to cohorts born in the early 1960s, but by 18 percentage points for corresponding cohorts who grew up in high-income families*¹. These differences in educational success translates into persistent future income inequalities.

The gap in educational attainment may be just the consequence of distinct predetermined levels of ability across different socioeconomic groups, or due to a combination of factors that emerge from various inequalities of opportunities in education. These sources of disparities may arise from pupils' characteristics (e.g., gender, race, immigrant status), family context (e.g., income levels, parents' human capital), or/and social environment (e.g., geographic space, neighborhood amenities or peers).

This dissertation gathers evidence on three sources of education inequalities across different education levels (pre-school, primary, secondary, and higher education) in the context of Spain. It revolves around the causal effects of large-scale educational policies on the efficiency and equity of educational systems. I gather large administrative datasets, design and collect survey data, to exploit these together with experimental and quasi-experimental econometric techniques to answer policy-relevant research questions. The first chapter studies to what extent financial aid for disadvantaged students attached to performance standards for renewal can improve student academic

¹Of the adults with at least one college-educated parent, 67 percent attained a tertiary qualification, compared to only 23 percent among those with low-educated parents (OECD, 2016).

achievement, make national financial aid programs more effective, and help to overcome moral hazard concerns in higher education. The second chapter examines the role of the test-taking environment on gender differences in academic performance, in the context of standardized testing in primary and secondary education. The third chapter focuses on how school choice priorities impact families' choices and pupil sorting across schools in pre-school education.

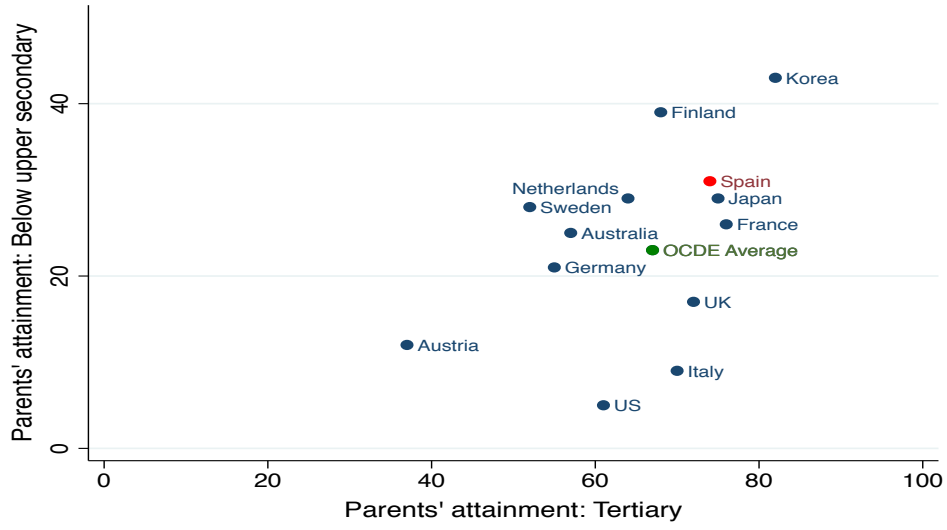
The three chapters of this dissertation share a similar context. They are focused on Spain. In particular, I use various sources of administrative micro-data of the Spanish region of Madrid and Carlos III University. The institutional context is highly attractive since there is relatively very little evidence that focused on this framework using rich administrative data. Besides, several large-scale education policy reforms were undertaken in the last years, offering a unique opportunity to study relevant research questions that are under-explored in the academic literature. To put into perspective this dissertation, the context is crucial. In the next few lines, I describe the main features of the Spanish higher education system, the gender differences in education performance, school choice, and school segregation. The Spanish higher education institutional framework is similar to France, Italy, Belgium, or Austria. Most of the post-secondary system is public, and tuition fees are relatively low. Panel A of Figure 1c plots the cross-country inter-generational mobility of educational attainment in higher education on a selected sample of countries. In Spain, of the young adults with parents' education below upper secondary education, 31% attained a tertiary qualification, eight percentage points more than the OECD average (23%). Spain is far above countries such as the US, UK, Italy, or Austria (below the OECD average), and similar to Netherlands, Sweden, Japan or France in this relative measure. Panel B of Figure 1c plots the cross-country gender differences in Mathematics and Reading on the PISA exam for 15-years old students. In most of the selected countries, girls outperform boys in Reading and boys outperform girls in Mathematics. Spain places among countries with the highest gender differences in Math and the lowest differences in Reading. Figure 1c plots the cross-country dissimilarity index for disadvantaged students and the proximity-based criterion

to assign pupils to schools. Spain reports slightly higher school segregation by socioeconomic background than OECD average and places among countries with the highest weight for residence-based assignment.

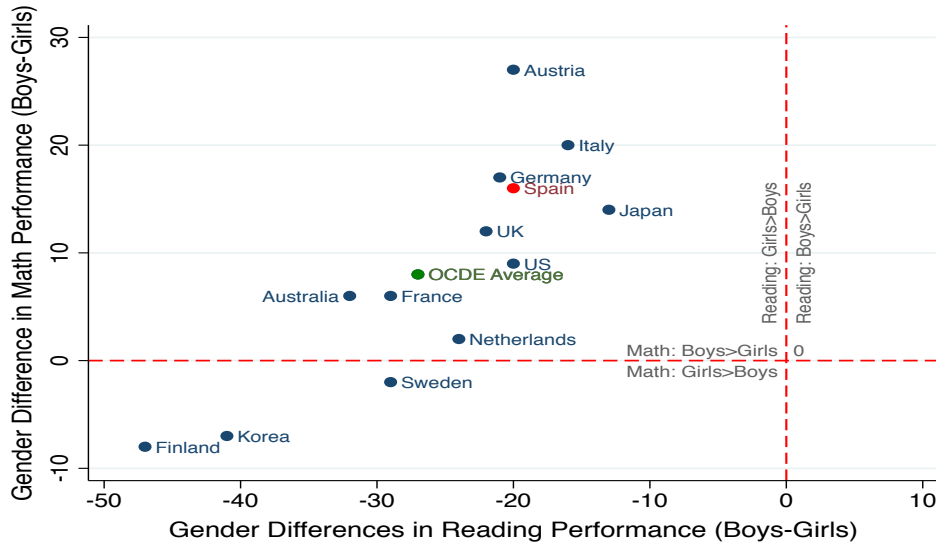
This dissertation examines the first source of education inequality that focuses on low-income family context. Specifically, it centers on the financial difficulties that low-income students face to access higher education. Empirical evidence has found that their main deterrent to entering college education is the financial barriers (Ellwood, Kane et al., 2000; Baum, Ma and Payea, 2013; Berg, 2016). To make higher education more accessible for disadvantaged students, many countries have implemented different policies, such as affirmative action programs, differential tuition fees rates, or financial aid (e.g., grants or loans). Several countries provide means-tested grants that cover tuition fees and award cash transfers to alleviate students' budget constraints. Examples of such programs are the Pell Grant in the US, the Maintenance Grant in the UK, the *Bourses sur critères sociaux* in France, or the *Becas de Carácter General* in Spain. Those programs are central in education policy debates since large-scale financial aid covers a large fraction of college students and represents a non-negligible share of the public budget. For instance, the US Pell Grant benefited over a third of college students (7 million) and accounted for 18% of the total federal student aid (28.2\$ billion) in 2017/18 (Board, 2017). Most of the existing literature focuses on the effects of need-based grant programs on college enrollment (Dynarski, 2003; Castleman and Long, 2016), with fewer papers looking at other outcomes such as college persistence (Bettinger, 2015; Fack and Grenet, 2015; Goldrick-Rab et al., 2016), graduation (Murphy and Wyness, 2016; Denning, 2018), and earnings (Angrist, 1993; Stanley, 2003; Denning, Marx and Turner, 2017). Existing studies have documented the positive influence of such programs on low-income students' enrollment, persistence, graduation, and earnings, in particular for the sub-population of students who would not have entered university without financial support ("marginal" students). Hence, the available evidence suggests that need-based grants are effective in expanding higher education opportunities for low-income students.

Figure 1: Inter-generational Mobility, Gender Differences, and School Choice.

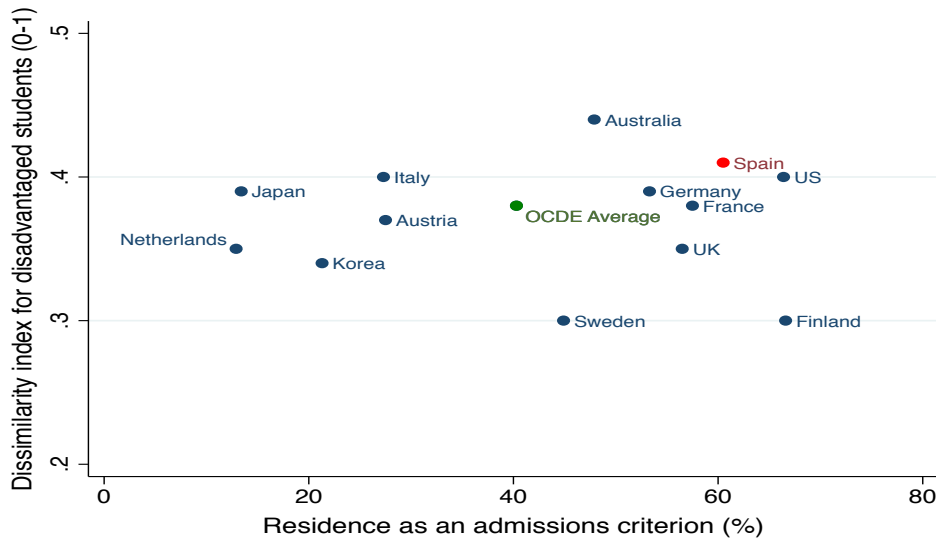
(a) Tertiary Educational Attainment as a Function of Parent's Attainment.



(b) Gender Differences in Math and Reading.



(c) School Choice vs. School Segregation.



Note: Sources PISA 2015, OECD (2016) and OECD (2019a).

In addition to the need-based criteria, most of the programs request applicants to meet minimum performance-based standards for renewal.² The principal-agent theory suggests that financial incentives that are not attached to performance for renewal may encourage enrollment and persistence of students who underperform in college and who eventually may not be able to graduate, creating moral hazard concerns.³ Introducing instruments of student accountability, such as grants linked to minimum academic requirements to renew them (i.e., having passed a certain number of credits), can serve as an effective tool to better monitor student effort and potentially align social and private incentives. However, there is a salient potential trade-off. While academic requirements may mitigate moral hazard concerns by helping students to reduce failure rates on exams and time to graduation, they can have the unintended side effect of inducing some students to drop out. Whether academic requirements improve the effectiveness of large-scale financial aid programs remains an empirical question.

The first chapter of this dissertation is devoted to investigating the causal effect of financial aid attached to minimum academic requirements on low-income students' academic performance, degree completion, and drop out from higher education. I use a Regression Discontinuity Design (RDD) that exploits the sharp discontinuities induced by family income eligibility thresholds to estimate the impact of being eligible for different categories of allowances on these student outcomes. Besides, I can explore the extent of the trade-off mentioned before analyzing a reform in the Spanish large-scale need-based grant program undertaken in 2013. This reform raised the academic requirements from a setting that is relatively comparable to those found in other national programs around the world (*weak* henceforth), such as the Satisfactory Academic Progress (SAP) in the US, to a more demanding one (*strong* henceforth). I take advantage of this natural experiment to estimate the causal

²In the US, such academic criteria are the Satisfactory Academic Progress (SAP) requirements for federal need-based aid programs, which commonly require students to maintain a cumulative grade point average (GPA) of 2.0 or higher and to complete at least two thirds of the course credits that they attempt (Scott-Clayton and Schudde, 2019)

³This concern has been particularly vivid in the US since college attendance rates have risen substantially, while undergraduate degree completion has been stable over the last two decades (Deming, 2017; Deming and Walters, 2017). This seems to be particularly salient for low-income students (Bailey and Dynarski, 2011).

effect of financial aid under *weak* and *strong* academic requirements. I use linked administrative micro-data, covering the universe of Carlos III University of Madrid students applying for the Spanish national grant program over the period 2010–2015. The dataset includes a comprehensive set of outcomes (e.g., GPA, dropout, final exam attendance, or selection of courses).

The main contributions of this chapter are two-fold. First of all, one of the main challenges in identifying the role that academic requirements play on the impact of financial aid on student performance is that the empirical evidence is usually only able to capture the combined impact of the awarded cash amounts and the impact of academic standards. Generally, the lack of reforms on large-scale national programs and data availability make it difficult to address to what extent these requirements contribute to the total effect of financial aid.⁴ The first contribution of this chapter is that this is the first paper that is able to isolate the specific contribution of academic requirements from the total effect of financial aid. Another main obstacle to identification is the difficulty to isolating the impact of grants on the intensive margin response of student performance, since most programs affect both the extensive and intensive margins simultaneously. The vast majority of the literature has focused on the extensive margin of enrollment, with several papers finding a statistically significant impact (Seftor and Turner, 2002; Fack and Grenet, 2015; Denning, Marx and Turner, 2017). This makes it difficult to interpret the intensive margin effect on performance due to the potential selection bias on those who access higher education. Few papers have been able to isolate the effect of financial aid on the intensive margin (Murphy and Wyness, 2016; Goldrick-Rab et al., 2016; Denning, 2018). The second contribution of this paper is to isolate the impact of the grant on the intensive margin performance, taking advantage of the specific timing of grant application in Spain. In this unique framework, students are already enrolled in higher education when they apply for the grant, allowing to capture the

⁴In a detailed summary of the lessons taken from the literature of financial aid, Dynarski and Scott-Clayton (2013) claim that “for students who have already decided to enroll, grants that link financial aid to academic achievement appear to boost college outcomes more than do grants with no strings attached”. Recent papers have raised doubts about this statement, finding mixed evidence. Goldrick-Rab et al. (2016) find that grants with no strings seem to increase college persistence of low-income students using a randomized experiment in several public universities in Wisconsin.

effects on the intensive margin response since the extensive margin is essentially muted due to the timing of grant applications.

I find that academic requirements turn out to be crucial, as their intensity plays an important role in stimulating low-income students' performance and degree completion when combined with financial aid. In addition, I show that the increase in their stringency does not necessarily have an impact on student drop out of higher education. The results show that being eligible for an average grant of 825 euros (relatively to being eligible for only a fee waiver) under *strong* academic requirements increases student average GPA and fraction of credits earned by 0.45 points (on a 0 to 10 scale) and 6 percentage points respectively, which corresponds to an increase of approximately 7.3 and 7.6 percent with respect to the baseline mean. These effects correspond to about 25 percent of the standard deviation of the dependent variable. Results persist over time, enhancing the student cumulative average GPA and fraction of credits earned over two consecutive years and increasing degree completion. Academic requirements in the context of higher education financial aid seem to be an effective tool to overcome moral hazard problems, though the optimal intensity of those requirements may be institutional context-specific.

The second source of education inequality that is analyzed in this dissertation focuses on the pupils' characteristics. In particular, it centers on the gender differences in academic performance that are due to the testing environment. Gender differences in academic attainment and achievement have dramatically reversed over the past 60 years. Whereas several decades ago men graduated from college at much higher rates than women, the situation is now reversed ([Goldin, Katz and Kuziemko, 2006](#)). As Panel B of Figure [1c](#) shows, in most industrialized countries boys continue to outperform girls in Science, Technology, Engineering, and Mathematics (STEM) ([Fryer Jr and Levitt, 2010](#); [Ellison and Swanson, 2018](#)). Women participation rates in STEM fields is lower than those of their male peers ([Card and Payne, 2017](#)), and the increasing female participation trend has leveled off in most OECD countries ([OECD, 2016](#)).⁵ This gender difference has implications for future labor market outcomes since

⁵See [Kahn and Ginther \(2017\)](#) for a review of the literature on the gender gap in STEM fields.

STEM occupations have been proven to offer higher wages ([Brown and Corcoran, 1997](#); [Black et al., 2008](#); [Blau and Kahn, 2017](#)). There is a long-standing debate on whether the sources of these gender differences in performance are biological (brain functioning) or due to the established social and cultural norms ([Baron-Cohen, 2004](#); [Nollenberger, Rodríguez-Planas and Sevilla, 2016](#)), with more evidence for the second hypothesis. The empirical evidence finds that in countries in which gender stereotypes are stronger have greater gender differences in STEMs' performance in favor of boys.

Most industrialized countries regularly undertake standardized testing of school children at different stages of the education path. In most countries, standardized test results determine an individual's educational choice set later on in life. However, measuring student ability through standardized testing may be problematic. Test scores depend not only on cognitive ability, but also in other factors such as non-cognitive skills ([Cunha and Heckman, 2007](#)), the ability to cope with competition ([Niederle and Vesterlund, 2010](#)), and the student's stereotypical beliefs ([Ehrlinger and Dunning, 2003](#)), among others. All these factors may artificially distort the underlying distribution of cognitive skills, making it difficult to infer a student's true ability. There is evidence that dealing with high-pressure environments such as an exam competes with the resources in the working memory that would otherwise be used to solve the test instead, leading to lower performance than expected given an individual's cognitive ability ([Beilock, 2010](#)). Experimental evidence from the psychology literature shows that pressure resulting from an emphasis on the importance of the positive consequences of good performance and/or the negative consequences of bad performance (the so-called importance of the process) can lead individuals to perform more poorly than expected given their skill level ([Beilock, 2008](#)). This is particularly so for women in male-dominated subjects, because of the "stereotype threat" about how one's social group should perform (e.g., "girls can't do math") produces less-than-optimal execution ([Steele, 1997](#); [Carr and Steele, 2009](#)).

There is substantial experimental literature showing that women underperform relative to men in competitive environments ([Niederle and Vesterlund, 2007](#); [Ors,](#)

[Palomino and Peyrache, 2013](#)). Several studies focus on the competitive nature of exams. Competitiveness is usually measured by the number of competitors and the final score being dependent on relative performance. These studies usually find that increases in competitive pressure favors men over women ([Iriberry and Rey-Biel, 2011](#); [Ors, Palomino and Peyrache, 2013](#)). A recent study looks at girls underperformance in a non-competitive environment where scores depend on absolute, rather than relative performance. They show that women perform worse when under pressure, defined as higher stakes associated with the exam ([Azmat, Calsamiglia and Iriberry, 2016](#)). Despite the relatively extensive literature on the gender differences in student performance, it remains unclear to what extent the organization of student testing influences the relative performance of male and female students in standardized tests.

The second chapter of this dissertation is devoted to investigate the causal effect of the testing environment on student standardized test scores. This chapter is jointly written with Almudena Sevilla. We use a randomized control trial which introduced changes to the test-taking environment implemented on the full population of students in 6th and 10th grades in the Region of Madrid (Spain). Some schools were randomly allocated to either an internally or an externally administered standardized testing procedure. We estimate the causal effect on student performance under familiar versus non-familiar test-taking environments using unique linked administrative student-level information on test performance for the population of students in 6th and 10th grades in 2016/2017 and 2017/2018. We use survey data collected specifically for this chapter to test the mechanism at play.

The main contributions of this chapter are three-fold. First, the randomized nature of the intervention allows us to focus on the effect of test-taking pressure on student's scores as a result of an exogenous variation in the familiarity with the testing environment, while other factors such as the gender of markers and teachers, the competitiveness of the environment, and the skills being tested are held constant. Second, our results are less likely to suffer from external validity and generalization. We exploit a randomized intervention rather than a quasi-natural experiment or a lab experiment as in previous studies. We also use population-level data for the

entire region of Madrid, rather than survey-based data or small populations based on a single school or a selected sample of high-ability individuals. Third, additional evidence from students' questionnaires allows us to test the mechanism directly at play. We rule out effort as the driver behind the increased gender differences in scores under externally administered testing environments. Rather, increased stress and lower confidence under externally administered testing environments seems to be the major factor at play.

We find that girls underperform in externally administered testing environments relative to boys. In particular, whereas under an internally administered testing environment boys in 6th Grade outperform girls by 0.1 standard deviations in Mathematics, and girls outperform boys in Science, Spanish and English by 6.8, 38 and 24 percent of a standard deviation, the gender gap widens in about 0.05 of a standard deviation for Mathematics (an effect size of 50 percent), and narrows in Science, Spanish and English (an effect size of 58, 13 and 18 percent respectively) under an externally administered testing environment. Similarly, whereas under an internally administered testing environment boys in Grade 10th outperform girls in Mathematics and Social Sciences by about 16 and 7 percent of a standard deviations, and girls outperform boys in Spanish and English by 15 and 11 percent of a standard deviations, we find that the gender gap in student performance widens by about 0.08 and 0.05 of a standard deviation in Mathematics and Social Sciences (an effect size of 52 and 62 percent respectively) under an externally administered test. Survey data shows that across all subjects, girls report more stress levels during the test and lower levels of confidence. However, girls studied more for all tests, and put more effort during all tests. We find that gender differences along these dimensions are exacerbated during externally administered testing environments, although the coefficients are less precisely estimated.

Empirical evidence has shown that the higher the stakes, the larger the increase in the level of pressure ([Azmat, Calsamiglia and Iriberry, 2016](#)). The estimates of this paper are based on a low stakes test. Therefore, it seems likely that the estimates from our paper are a lower bound of the potential effects in high stakes

testing. Most of the education systems rely on high stakes external standardized testing to access different education tracks. We show that girls are more sensitive than boys to the testing environment (independently of their true level of ability), decreasing their performance when the latest is less familiar to students. Result imply that a non-familiar testing-environment in standardized test exacerbates the gender differences in performance, which may translate into differential carrier choice, and thus persistent gender inequalities over the earnings life-cycle.

This dissertation examines a third source of education inequality that focuses on the social environment. In particular, it targets the relationship between school choice and school segregation. A substantial body of research has documented that the social environment understood as the geographical space, such as neighbourhood amenities or peer exposure, plays an important role in the economic, health, and educational outcomes. Literature finds that those individuals that live in low-income areas report worse economic, health, and educational outcomes than those who live in high-income neighbourhoods ([Jencks and Mayer, 1990](#); [Leventhal and Brooks-Gunn, 2000](#); [Sampson, Morenoff and Gannon-Rowley, 2002](#)). Moving from high-poverty housing projects to low-poverty neighbourhoods increases college attendance, earnings and reduces single-parenthood rates ([Chetty, Hendren and Katz, 2016](#)). In addition, the Moving to Opportunity Project (MTO) experiment finds that the duration of exposure to a better environment during childhood is a crucial determinant for children’s long-term outcomes ([Chetty, Hendren and Katz, 2016](#); [Chetty and Hendren, 2018a,b](#)). There is a large body of research that studies how the exposure to different peers have an impact on student performance ([Coleman, 1968](#); [Hoxby, 2000](#); [Mora and Oreopoulos, 2011](#); [Lavy, Silva and Weinhardt, 2012](#); [Bursztyn and Jensen, 2015](#)) looking at classroom peer effects. However, other papers have looked at the peer effects at the neighborhood ([Goux and Maurin, 2007](#); [Gibbons, Silva and Weinhardt, 2017](#)), and the family level ([Goodman et al., 2015](#); [Joensen and Nielsen, 2018](#)). Another strand of the literature has been devoted to analyzing the relationship between school choice policies and school segregation.

The relationship between school choice reforms and student segregation has

generated significant policy interest. During the last three decades, there has been a clear pattern of educational authorities have increased the degree of school choice in their educational systems [Musset \(2012\)](#). In the US, many school choice reforms were complemented by busing programs (e.g., Seattle in 1999 or North Carolina in 2002). In particular, school choice reforms involve, among others, zoning and de-zoning policies, changes in admission criteria, and changes in the system of assignment of students to schools. Most of the school choice literature has focused on the market design questions to school choice, mainly devoted to analyzing the relative performance and strategic implications of alternative matching allocation mechanisms, taking the inputs of school choice -preferences, priorities, and capacities- as exogenous ([Abdulkadiroglu and Sönmez, 2003](#)). The potential effects of school choice policies on school segregation are not straightforward. Although several papers have been devoted to study the impact of increasing the level of choice on segregation ([Epple, Romano and Urquiola, 2017](#); [Böhlmark, Holmlund and Lindahl, 2016](#); [Söderström and Uusitalo, 2010](#)) with most of them finding a positive relationship, the extent through which school choice priorities affect school segregation remains an open question.

The third chapter of this dissertation broadens the scope of the market design questions to school choice by investigating how government-determined school choice priorities affect households' choices and pupil sorting across schools. This chapter is jointly written with Lucas Gortázar and David Mayor. We use two large-scale school choice reforms in the school choice priorities structure as a source of variation. First, the low-income priorities' to the top-ranked school were reduced and granted an additional point to alumni family members of the school in 2012/2013. Second, the resident-based priorities to assign pupils to schools were almost completely abolished in 2013/2014. The city of Madrid counts 21 school districts that were almost merged *de facto* into a unique single district.⁶ We use unique administrative data on the universe of applicants to the public school system from 2010 to 2016

⁶In the Region of Madrid, the Regional Government, enlarged the choice zone to the municipal level, granting a larger set of options to all households. The reform moved from around 2,000 within-municipality catchment areas to 179 single municipal zones. We focus on the city of Madrid for our main analysis.

in the Region of Madrid, along with detailed data on school supply, household socioeconomic characteristics and standardized test scores. We combine two different empirical strategies to identify the impact of the reforms. First, we use an event study first difference approach. We compare families entering the educational system for the first time (pre-school age 3) before and after the reforms. Second, we use a Difference-in-Difference Analysis (DID) in the spirit of treatment intensity, focusing on parents that are closer to the school district boundaries (“treatment group”), and comparing them with those whose main residence is located at the geometric center (centroid) of each the school district boundaries (“control group”).

The main contributions of the third chapter to the currently existing literature is three-fold. First, we can compute the contemporaneous effect of increasing choice on school segregation at the earliest schooling stage (pre-school for three years old students). Most of the literature has focused on secondary education that entails two different aspects: (i) Segregation may be the result of a combination of factors that are shaped in earlier educational stages; (ii) Priority bonus in secondary education are typically based on student grades, while those of primary education are usually centered on socio-demographic indicators which may potentially have a more direct impact on school segregation. Second, this paper is able to closely relate changes in school choice priorities with the immediate impact on school segregation. Most of the related literature has been focused either on broader contexts or on the impacts of early-stage policy reforms (e.g., primary education) of later phases of the educational career (e.g., secondary education) -which results may be potentially biased by time-variant confounding factors-. Third, this paper explores variables that some of the previous literature does not consider, such as families’ choices or the precise geo-location of the household’s main residence and schools. This allows computing variables that are determinants of school segregation, such as the residential segregation or accurate measures of families’ willingness to commute.

We find that the inter-district school choice reform increased the out-of-district choice and assignment (with heterogeneous effects). We find distributional effects of the reform, concluding that parents from the highest education levels and parents

of non-immigrant students were those who reacted the most in absolute terms. Interestingly, results support the idea of potential information gaps and the dynamic learning process across immigrant status groups. We measure school segregation using the Mutual Information Index, which satisfies several desirable properties ([Frankel and Volij, 2011](#)). We find a decreasing trend in school segregation by parental education over time (mostly driven by the decrease in the within school district segregation), but an increasing trend in school segregation by immigrant status-though effects on the latest fade out when controlling for residential stratification. This chapter draws an important policy takeaway: When strong polarization in parents' school choice, priority structures need to be carefully designed to achieve diversity objectives. Levels of segregation prior to the reform potentially matter to predict the effects of increasing school choice.

Even though this dissertation makes important academic and policy contributions, there are still many interesting research questions that need to be explored:

- In the first chapter, it is still unclear which would be the distribution of the effects of a grant on performance and drop out along the different possible academic requirements (from zero strings attached to full accountability). Understanding how the impact of a grant changes when attached to all different possible academic requirements turns out to be a crucial research question, to improve the effectiveness of large-scale financial aid programs. Establishing the optimal line of academic standards and the amount of financial aid that is socially optimal remains a topic for future research.
- In the second chapter, further research needs to explore potential heterogeneity in the invigilation of the test (e.g., characteristics of the invigilator) and/or the characteristics of the testing environment (e.g., disposition of tables or size of classes) that may contribute to the observed gender differences in student performance. It is also worth investigating whether some subgroups of the population may be affected by the testing environment more than others. Answering these types of questions can help to improve the design of testing environments to elicit students' true level of ability better.

- In the third chapter, further research needs to be undertaken to understand under which conditions, such as the predetermined levels of school (and residential) segregation, subgroups of the population considered, school choice allocation mechanism or parents' preferences, school priorities may contribute to either reduce or increase school segregation.

Contents

Acknowledgements	v
Summary	xiii
Résumé	xvii
General Introduction	1
1 Countering moral hazard in higher education : The role of performance incentives in need-based grants	21
1.1 Introduction	23
1.2 Institutional Background	31
1.2.1 Higher Education in Spain	31
1.2.2 <i>The Becas de Carácter General</i> Need-Based Grant Program	32
1.3 Data	37
1.4 Empirical Strategy	38
1.5 Theoretical Framework	42
1.6 Results	46
1.6.1 Internal Validity of the Empirical Strategy	46
1.6.2 Discontinuities in Grant Amounts	48
1.6.3 Impact on Dropout	49
1.6.4 Impact on Student Performance	49
1.6.5 Robustness Checks	52
1.6.6 Heterogeneous Effects of Grants on Student Performance	54
1.6.7 Mechanisms	56
1.6.8 Impact on Degree Completion	58
1.7 Discussion	58
1.7.1 External Validity	59
1.7.2 Interpretation of Results: Efficiency and Equity	60
1.7.3 Potential confounding factors	65
1.8 Conclusion	67
Appendix 1.A Appendix: Main Figures	69
Appendix 1.B Appendix: Main Tables	74
Appendix 1.C Additional Figures and Tables	84
1.C.1 Low-Income Students' Performance in Higher Education	84
1.C.2 Validity of the Research Design: McCrary (2008) Test	87
1.C.3 Discontinuities in Awarded Grants	91
1.C.4 Robustness Checks	93
1.C.5 RDD-DID	104
1.C.6 Comparability between Period I and Period II	106

1.C.7	Equity Effects	107
1.C.8	Academic requirements for BCG grant	112
2	The Gender Gap in Student Performance: The Role of the Testing Environment	115
2.1	Introduction	117
2.2	Institutional Background	122
2.2.1	Standardized Testing in Spain	122
2.2.2	A Randomized Intervention in the Administration of Standardized Testing	124
2.3	Data and Descriptive Statistics	125
2.4	Empirical Strategy	127
2.5	Results	128
2.5.1	Baseline Results	128
2.5.2	Robustness Checks and Heterogeneity Analysis	130
2.5.3	Ruling out alternative explanations	131
2.6	Identifying the mechanisms through which the testing environment affects the gender gap in math	135
2.7	Conclusion	137
	Appendix 2.A Main Figures and Tables	139
	Appendix 2.B Additional Figures and Tables	154
2.B.1	Additional Balancing Tests	154
2.B.2	Description of Variables	155
2.B.3	Additional Heterogeneous Effects	156
3	School Choice Priorities and School Segregation: Evidence from Madrid	159
3.1	Introduction	161
3.2	Institutional Background	168
3.3	Data and Summary Statistics	175
3.3.1	Data	175
3.3.2	Summary statistics of applicants	178
3.4	Empirical Strategy	180
3.4.1	Out of School District Choice and Assignment	180
3.4.1.1	Event Study: First Difference Approach	180
3.4.1.2	Difference-in-Difference Analysis	183
3.4.2	School Segregation	185
3.4.3	Identification Threats	187
3.5	Results	189
3.5.1	Out of District Choice and Assignment	189
3.5.1.1	Event Study: First Difference Approach	189
3.5.1.2	Difference-in-Difference Analysis	191
3.5.2	School Segregation	192
3.6	Robustness Check: Phasing-in of the Reform in other Municipalities	195
3.6.1	Out-of-Municipality Assignment and Distance to Assigned School	196
3.6.2	School Segregation by Immigrant Status	197
3.7	Conclusion	198
	Appendix 3.A Main Figures	200
	Appendix 3.B Main Tables	210
	Appendix 3.C Additional Figures and Tables	213

3.C.1	Description of priority criteria	213
3.C.2	Section Example	214
3.C.3	Years of Schooling	214
3.C.4	Population trends in the Region of Madrid	216
3.C.5	Housing Prices and School Average Performance.	217
3.C.6	To which schools are pupils assigned?	218
3.C.7	Sample Restrictions	221
3.C.8	Theoretical Properties of the Boston Mechanism	221
3.C.9	Difference-in-Difference Approach	222
3.C.10	Alternative measure of segregation by parental education . . .	224
3.C.11	School Classification	227
<i>Bibliography</i>		246
<i>List of Tables</i>		250
<i>List of Figures</i>		252

Chapter 1

Countering moral hazard in higher education : The role of performance incentives in need-based grants *

*I would like to especially thank Julien Grenet for his excellent, devoted and generous advice. I would like to express my gratitude to Ghazala Azmat, José E. Boscá, Antonio Cabrales, Natalie Cox, Gabrielle Fack, Javier Ferri, Martín Fernández-Sánchez, Martín García, Jesús Gonzalo, Marc Gurgand, Xavier Jaravel, Juan Francisco Jimeno, Clara Martínez-Toledano, Sandra McNally, Guy Michaels, Chrstopher A. Neilson, Fanny Landaud, Michele Pellizzari, Steve Pischke, Olmo Silva, Daniel Reck, Mariona Segú, Alessandro Tondini, Carmen Villa Llera and Ernesto Villanueva for their useful and wise comments. I am also grateful to Paloma Arnaiz Tovar, Gloria del Rey Gutiérrez, Luis Losada Romo and Elena Zamarro Parra, for their invaluable help in collecting the data. I thank seminar participants at Paris School of Economics, London School of Economics, University College London, Queen Mary University, Bank of Spain, European University Institute, Institute of Social Research (SOFI) at Stockholm University, Melbourne Institute of Applied Economic and Social Research (University of Melbourne), Universidad Autónoma de Madrid and DISES-CSEF (University of Naples Federico II), as well as conference participants at the EEA-ESEM 2018, IZA World Labor Conference 2018, EALE 2018, IWAE 2018, 16th “Brucchi Luchino” Labor Economics Workshop, SAE 2018, XXVI AEDE Meeting, VIII Workshop on Economics of Education and Impact of Grants on Education and Research Seminar at AIRF. Finally, I am thankful to Fundación Ramón Areces and Bank of Spain for its financial support. This paper has been awarded with the *Nada es Gratis* Prize for Job Market Papers in Economics in 2018, and with the María Jesús San Segundo Second Best Paper Award for Young Researchers in 2017. An earlier version of this work was circulated under the title: “The role of performance incentives in need-based grants for higher education: Evidence from the Spanish *Becas*.”

Abstract

National financial aid programs for disadvantaged students cover a large fraction of college students and represent a non-negligible component of the public budget. These programs often have weak performance requirements for renewal, potentially leading to moral hazard and efficiency losses. Using a reform in the Spanish need-based grant program in higher education, this paper tests the causal effect of receiving the same amount of grant under different intensities of academic requirements on student performance, degree completion and student dropout. I use administrative micro-data on the universe of applicants to the grant in a large university. Exploiting sharp discontinuities in the grant eligibility formula, I find strong positive effects of being eligible for a grant on student performance when combined with demanding academic requirements, while there are no effects on student dropout. Students improve their final exam attendance rate, their average GPA in final exams, and their probability of completing the degree. They also reduce the fraction of subjects that they have to retake. The grant has no effects on student performance when academic requirements are low and typically comparable to those set out by national need-based student aid programs around the world. These results suggest that academic requirements in the context of higher education financial aid can be an effective tool to help overcome moral hazard concerns and improve aid effectiveness.

JEL Codes: I21, I22, I23, I28, H52

Keywords: Need-based grants; performance incentives; moral hazard; college achievement.

1.1 Introduction

To make higher education more accessible for disadvantaged students, many countries have implemented different policies, such as affirmative action programs, differential tuition fees rates or financial aid (e.g., grants or loans). Financial incentives that are not attached to performance for renewal may encourage enrollment and persistence of students who underperform in college and who eventually may not be able to graduate, creating moral hazard concerns.¹ Introducing instruments of student accountability, such as grants linked to minimum academic requirements to renew them (i.e., having passed a certain number of credits), can serve as an effective tool to better monitor student effort and potentially align social and private incentives. However, while academic requirements may mitigate moral hazard concerns by helping students to reduce failure rates on exams and time to graduation, they can have the unintended side effect of inducing some students to drop out. Whether academic requirements improve the effectiveness of large-scale financial aid national programs remains an empirical question. This issue is central in education policy debates, since large-scale financial aid covers a large fraction of college students and represents a non-negligible share of the public budget. For instance, the US Pell Grant benefited over a third of college students (7 millions) and accounted for 18% of the total federal student aid (28.2\$ billion) in 2017/18 (Board, 2017).² Despite the relatively extensive literature on the effects of financial aid, it remains unclear to what extent grants tied to academic requirements are more effective than those without.

This paper investigates the causal effect of financial aid attached to minimum

¹This concern has been particularly vivid in the US, since college attendance rates have risen substantially, while undergraduate degree completion has been stable over the last two decades (Deming, 2017; Deming and Walters, 2017). This seems to be particularly salient for low-income students (Bailey and Dynarski, 2011).

²These programs typically provide fee waiver and award cash transfers to students based on their family income – other examples are the *Bourses sur critères sociaux* in France or the *Becas de Carácter General* in Spain. The debate regarding SAP has focused in the academic (Dynarski and Scott-Clayton, 2013; Goldrick-Rab et al., 2016; Patel and Rudd, 2012) and the public (Reauthorization (2013) and Gates Foundation’s Reimagining Aid Design and Delivery project) sphere. Questions and concerns on financial aid policy are summarized in Scott-Clayton (2017b), and some of the proposals to change the Pell Grants are Baum and Scott-Clayton (2013) and Scott-Clayton (2017a).

academic requirements on low-income students' academic performance, degree completion and dropout from higher education. I use a Regression Discontinuity Design (RDD) that exploits the sharp discontinuities induced by family income eligibility thresholds to estimate the impact of being eligible for different categories of allowances on these student outcomes. There is a potential trade-off in tightening minimum academic requirements that is analyzed based on a principal-agent model building on [Bénabou and Tirole \(2000, 2002\)](#) with academic standards (i.e., requirements), and financial aid ([Scott-Clayton and Schudde, 2019](#)). The model predicts that, if the social planner increases the level of standards, some agents would become “more ambitious” exerting higher effort. However, the net effect in equilibrium is ambiguous, since some of the weaker agent types would “give up”, exert zero effort and potentially drop out.³ I explore the extent of this trade-off exploiting a reform in the Spanish large-scale need-based grant program undertaken in 2013. This reform raised the academic requirements from a setting that is relatively comparable to those found in other national programs around the world (*weak* henceforth), such as the Satisfactory Academic Progress (SAP) in the US, to a more demanding one (*strong* henceforth).⁴ I take advantage of this natural experiment to estimate the causal effect of financial aid under *weak* and *strong* academic requirements. I use linked administrative micro-data, covering the universe of Carlos III University of Madrid students applying for the Spanish national grant program over the period 2010–2015. The dataset includes a comprehensive set of outcomes (e.g., GPA, dropout, final exam attendance or selection of courses). In a nutshell, I find that academic requirements turn out to be crucial, as their intensity plays an important role in stimulating low-income students' performance and degree completion, when combined with financial aid. In addition, I show that the increase in their stringency does not necessarily have an impact on

³Note that from the principal's perspective, increasing standards would be worthwhile if the rise in the value due to those who are induced to exert higher effort overcomes the loss of value induced by those who shirk. When financial aid is incorporated into the model, [Scott-Clayton and Schudde \(2019\)](#) show that if the social value of those who shirk is lower than the value of financial aid, then aid with high standards seems to be unambiguously better than aid without. Nonetheless, the optimal line of standards and amount of financial aid that is socially optimal remains a question for future research.

⁴The SAP generally requires students to maintain a GPA of 2.0 or higher, and to complete at least two thirds of the course credits they undertake to renew the Pell Grant ([Scott-Clayton and Schudde, 2019](#)).

student drop out from higher education.

This is the first paper that is able to isolate the specific contribution of academic requirements from the total effect of financial aid. One of the main challenges in identifying the role that academic requirements play on the impact of financial aid on student performance is that the empirical evidence is usually only able to capture the combined impact of the awarded cash amounts and the impact of academic standards. Generally, the lack of reforms on large-scale national programs and data availability make it difficult to address to what extent these requirements contribute to the total effect of financial aid.⁵ Accordingly, previous empirical evidence presents limitations. Most of the existing literature focuses more on narrowly defined rather than large-scale national programs, implemented in the US for a specific university or state ([Deming and Dynarski, 2009](#); [Dynarski and Scott-Clayton, 2013](#)). Such programs are usually performance-based, the vast majority of which are targeting high or medium achieving low-income students ([Dynarski, 2008](#); [Scott-Clayton, 2011](#)). Hence cannot be easily generalized to the population of low-income students. In addition, they often include additional components, such as academic and support services, which make it difficult to isolate the specific role of academic requirements ([Angrist, Lang and Oreopoulos, 2009](#); [Angrist, Oreopoulos and Williams, 2014](#)). The first contribution of this paper aims to fill this gap, by identifying how the impact of being eligible for equal grant amounts differs when combined with different intensities of academic standards.

The second contribution of this paper is to isolate the impact of the grant on the intensive margin performance, taking advantage of the specific timing of grant application in Spain. In this unique framework, students are already enrolled in higher education when they apply for the grant, allowing to capture the effects on the intensive margin response since the extensive margin is essentially muted due to the timing of grant applications. Another main obstacle to identification is the

⁵In a detailed summary of the lessons taken from the literature of financial aid, [Dynarski and Scott-Clayton \(2013\)](#) claim that “for students who have already decided to enroll, grants that link financial aid to academic achievement appear to boost college outcomes more than do grants with no strings attached”. Recent papers have raised doubts on this statement, finding mixed evidence. [Goldrick-Rab et al. \(2016\)](#) find that grants with no strings seem to increase college persistence of low-income students using a randomized experiment in several public universities in Wisconsin.

difficulty to isolate the impact of grants on the intensive margin response of student performance, since most programs affect both the extensive and intensive margins simultaneously. Ex-ante, financial aid may induce new students to access college (“marginal” students), incentivizing enrollment. Average student performance may improve due to the increase in initial college enrollment, called the extensive margin. Ex-post, after enrollment, financial aid may enhance the performance of those students whose initial enrollment was not affected by the grant. This potential channel is the intensive margin (these students are referred as “inframarginal” students). The vast majority of the literature has focused on the extensive margin of enrollment, with several papers finding a statistically significant impact (Seftor and Turner, 2002; Fack and Grenet, 2015; Denning, Marx and Turner, 2017).⁶ This makes it difficult to interpret the intensive margin effect on performance due to the potential selection bias on those who access higher education.⁷ Few papers have been able to isolate the effect of financial aid on the intensive margin (Murphy and Wyness, 2016; Goldrick-Rab et al., 2016; Denning, 2018). However, they cannot look at more comprehensive measures of student performance, such as fraction of subjects to retake or final exam attendance rate, since they are mainly limited to graduation outcomes, earnings and college persistence. The second contribution of this paper aims to identify the effect of being eligible for a grant on the intensive margin of an exhaustive set of student performance measures.

The results show that being eligible for an average grant of 825 euros (relatively to being eligible for only a fee waiver) under *strong* academic requirements increases student average GPA and fraction of credits earned by 0.45 points (on a 0 to 10 scale) and 6 percentage points respectively, which corresponds to an increase of approximately 7.3 and 7.6 percent with respect to the baseline mean. These effects correspond to about 25 percent of the standard deviation of the dependent

⁶Empirical evidence looking at the Pell Grant aid program seems to be mixed, as several studies find no effect on college enrollment (Kane, 1995; Turner, 2017; Marx and Turner, 2018; Denning, 2018).

⁷Furthermore, the vast majority of papers looking at the impacts of merit-based and need-based allowances on student achievement may not be entirely representative of all the population of college students, due to the fact that they focus on non-enrolled or freshmen students, who report the highest probability of dropout.

variable. The magnitudes of the effects are similar to those found for the West Virginia’s PROMISE merit-based scholarship (Scott-Clayton, 2011), larger than those of Project STAR (Angrist, Lang and Oreopoulos, 2009) and slightly lower than Opportunity Knocks’s merit-based experiments in a large Canadian university (Angrist, Oreopoulos and Williams, 2014). Results persist over time, enhancing the student cumulative average GPA and fraction of credits earned over two consecutive years and increasing degree completion. In contrast, I find no effects of similar cash amounts in a regime of *weak* academic requirements. These results are consistent with papers finding limited effects of similar Pell Grant’s cash amount on student GPA under a relatively similar context as the *weak* academic requirements setting (Denning, Marx and Turner, 2017; Denning, 2018). The baseline results are robust to an exhaustive number of specifications and sample selection criteria.

Interestingly, the effects do not seem to depend on the amount of financial aid awarded, as no further improvements in student performance are found for increments in the amount of awarded aid when students are already entitled to some positive amount of aid. The results suggest that student performance is mostly responsive to being awarded a grant with *strong* academic requirements, but consistent with the findings of Gneezy and Rustichini (2000): the boost in performance does not depend linearly on the amount received. To explore potential equity effects beyond the RDD estimates, I perform a Difference-in-Difference model (DID), a RDD-DID specification, and a Logit analysis to measure the propensity to dropout before and after the change in requirements. I find that combining financial aid with *strong* academic requirements does not necessarily lead to an increase in drop out from higher education. This result seems to contradict the heterogeneous discouragement effects of SAP failure in the US Pell Grant, and similar effects of being placed in academic probation for students (regardless of financial aid) at one large baccalaureate institution in Canada found by Scott-Clayton and Schudde (2019) and Lindo, Sanders and Oreopoulos (2010) respectively. The institutional context, especially the cost of college, may potentially be a relevant factor affecting the elasticity of dropout with respect to academic requirements.

This paper explores several possible mechanisms. The main ways in which financial aid may influence low-income students' performance are the cost-of-college and performance-incentives effects. First, the relaxation of budget constraints may prevent financially constrained students from working part-time, inducing them to devote more time to study. Second, if students lack sufficient motivation, have high time preferences, or are not aware of the exact returns to schooling, performance-based incentives may increase their motivation to exert higher academic effort and self-improve. Nevertheless, [Fryer \(2011\)](#) remarks that the direction of the effects may be ambiguous if students lack the structural resources or knowledge to convert effort into a measurable achievement, or if financial rewards (or any kind of external reward) undermine intrinsic motivation and lead to negative outcomes.⁸ I find that students attend final exams more often, an indirect proxy of student effort. Students enhance their GPA on final exams taken and reduce the fraction of subjects that they have to retake, suggesting a genuine increase in student performance. In addition, I investigate potential confounding factors that seem to rule out an increase in the probability of dropout from higher education and student selection of courses that may bias the baseline results. I am able to distinguish between subjects that are mandatory (i.e., those where students do not have the right to self-select) and elective (i.e., where students can choose a subset from a certain degree-specific set). The effect seems to be particularly robust for those courses which are compulsory, and hence cannot be avoided by students. Finally, I cannot test the cost-of-college channel directly in this paper, but I find that the student effort responds positively to positive amounts of aid combined with academic requirements, and the effort response seems unrelated to the amount awarded, potential indirect evidence that may rule out this hypothesis.

Related literature. This paper is closely related to the literature on the effects of financial aid, in which relatively little is known about the effectiveness of large-scale need-based grant programs in enhancing low-income students' educational outcomes.

⁸See, for instance, [Deci \(1972\)](#), [Kohn \(1996\)](#) and [Gneezy and Rustichini \(2000\)](#) to observe some pieces of the extensive debate in psychology on whether extrinsic rewards crowd out intrinsic motivation.

Most of the existing literature focuses on the effects of need-based grants programs on college enrollment (Dynarski, 2003; Fack and Grenet, 2015; Castleman and Long, 2016), with fewer papers looking at other outcomes such as college persistence (Bettinger, 2015; Goldrick-Rab et al., 2016), graduation (Murphy and Wyness, 2016; Denning, 2018) and earnings (Angrist, 1993; Stanley, 2003; Denning, Marx and Turner, 2017). Existing studies have documented the positive influence of such programs on those low-income students' outcomes, specially for the sub-population of students who would not have entered university without financial support. Beyond the need-based programs, the vast majority of the literature has focused on merit-based grants that are awarded to students who meet certain academic requirements, and typically do not target low-income students.⁹ These papers point out the importance of performance-based incentives on stimulating student performance, but effects seem to be small (Angrist, Lang and Oreopoulos, 2009; Angrist, Oreopoulos and Williams, 2014) and mixed.¹⁰

This paper is also related to the literature on how extrinsic incentives affect performance in the labor market (Lazear, 2000) and in non-employment contexts (Gneezy, Meier and Rey-Biel, 2011). In the latter, a number of studies have provided empirical support for the claim that stronger monetary incentives tend to lead to higher levels of effort, but the effect of monetary compensation on performance does not seem to be monotonic (Gneezy and Rustichini, 2000). If incentives are sizable enough, their direct price effects will be larger than the crowd-out effect, but if they are too high, individuals can “choke under pressure” and incentives can backfire (Ariely et al., 2009). Empirical evidence that has evaluated extrinsic incentives

⁹These grants have a long tradition in the US post-secondary system. Traditional programs such as the US National Merit program and Canadian Excellence Awards, were originally targeted to top-performers. In the 1990s, several programs such as Georgia's Helping Outstanding Pupils Educationally (HOPE) were introduced for non-top students in different US states (e.g., Florida or Arkansas). Empirical evidence devoted to investigating the effects of the numerous HOPE-style programs have found positive results for key students' outcomes (see, e.g. Cornwell, Mustard and Sridhar (2006) for an evaluation of Georgia's HOPE program, Dynarski (2008) and Sjoquist and Winters (2012) for an investigation of Georgia and Arkansas HOPE-like programs, and Castleman et al. (2014) for an analysis of the Florida Bright Futures Scholarship.

¹⁰The Manpower Demonstration Research Corporation (MDRC) has performed several randomized evaluations of performance-based scholarship (several of them targeted to low-income students) finding mixed results (e.g., Mayer et al. (2015), Barrow et al. (2014), Cha and Patel (2010), Miller et al. (2011), Richburg-Hayes et al. (2009) or Richburg-Hayes, Sommo and Welbeck (2011).

from large-scale field experiments in educational contexts show that they increase attendance and enrollment, have mixed results on effort and achievement, and seem to work for some students but not for others (Gneezy, Meier and Rey-Biel, 2011). In addition, this paper also contributes to the literature on the effects of performance standards. There is a vast literature that has examined the impact of academic standards in the educational context, especially in the form of high school exit exams. Empirical evidence has focused on the effects of the presence of high school exit exams, and the impact of failing high school exams on dropout and performance, finding mixed results (Muller, 1997; Amrein and Berliner, 2002). In higher education, Lindo, Sanders and Oreopoulos (2010) examined the students' responses of being placed in academic probation at the end of the first year, showing that some students are discouraged to return to university while those who stay improve their GPAs. The authors highlight the potential trade-off between the increased effort and dropout, induced by the academic standards. However, it remains unclear to what extent results from these papers may be generalized to the case when extrinsic incentives and performance standards are combined. This paper contributes to these strands of the literature showing that combining financial aid and *strong* academic requirements has positive effects on student performance, while it does not necessarily lead to an increase in drop out, using a large-scale program targeted to low-income students.

Organization of the Paper. The remainder of this paper is structured as follows. Section 2 provides institutional background on the Spanish higher education system and on the national need-based grant program. Section 3 describes the data used in the paper. Section 4 explains the empirical strategy. Section 5 shows a theoretical framework to describe the fundamental trade-offs involved in setting performance standards in need-based student aid. Section 6 discusses the internal validity of the research design, analyzes the main results, explores heterogeneous effects and examines the different mechanisms that could explain the results. Section 7 discusses the external validity, efficiency and potential equity costs, and potential additional confounding factors that may affect the results. Section 8 concludes.

1.2 Institutional Background

1.2.1 Higher Education in Spain

The Spanish educational system is organized as six years of primary schooling (from the age of 6 to the age of 12), four years of secondary education (from the age of 13 to the age of 16), and two years of non-compulsory education, which is divided into a vocational track (*Ciclos Formativos*) and an academic track (*Bachillerato*). After graduating from high school, students choose whether to pursue into higher education. The vast majority decide to enroll in college education, leading to vocational undergraduate degrees (*CFGS*), academic undergraduate degrees (four-year degree called *Grado*), graduate degrees (*Master*) and doctoral studies. To access higher education, students must pass the standard access to university test (*PAU*),¹¹ which consists of two-year college preparation courses and a standardized entry exam (*Selectividad*).¹² If the demand for a specific program in the Spanish public universities exceeds the number of available seats, students are admitted in the order of their *PAU* grades until all seats are filled. Outside of these two main tracks, a minority of students enroll in artistic education (e.g., arts, music, dance, dramatic arts), which offers undergraduate and graduate degrees.

The cost of higher education in Spain is mainly composed of tuition fees and living expenses. Tuition costs vary depending on the region where the university is located, the degree program undertaken, and the repeated subjects failures. The national average tuition fees for a full academic year was 1,100 euros for undergraduate students in 2015 and between 1,634 and 2,347 euros for graduate students.¹³ Given the fact that most of the universities are located in large urban areas, students face relatively high living costs. In 2011, a survey on living conditions of Spanish college students indicated that the majority were living with their parents, and that only 6.3

¹¹The name has changed from 2017 onward to *Evaluación de Bachillerato para el Acceso a la Universidad* (EBAU). 92 percent of the students who took the test in 2015 passed it.

¹²The final grade of *PAU* is composed by a preponderated average with weights 0.6 for *Bachillerato* and 0.4 for *Selectividad*.

¹³Public prices are detailed in *Estadísticas de precios públicos universitarios del MECD*.

percent were living in university residence halls.¹⁴ Furthermore, according to current estimations for the first semester of 2015, the average cost of living expenses in Spain for a nine-month period was 5,069 euros,¹⁵ which represent a significant financial barrier to emancipate from their family home and to access higher education.

1.2.2 *The Becas de Carácter General* Need-Based Grant Program

Grant Program. The *Becas de Carácter General* (BCG hereafter) is the Spanish national financial aid program for low-income students in post-secondary education. BCG is the most ambitious program for college students in Spain, since it represents about 86 percent of the total budget for grants in higher education. About a quarter of the academic undergraduates and 15 percent of graduate students enrolled received this grant in 2014, for a total cost of 829 millions of euros. The official objectives assigned to this grant program by the Ministry of Education are the equality of opportunities and to improve educational efficiency by promoting low-income students' potential.

The program consists on three main levels of grant: (i) the Fee Waiver (Threshold 0) exempts eligible applicants from paying tuition fees; (ii) the Residence Grant (Threshold 1) provides cash allowance which is intended to cover home expenses of students who live away from their family home by reasons of college distance; and (iii) the Compensate Grant (Threshold 2) provides cash allowance to compensate the student's lack of family income. Students who qualify for the Residence Grant (T1 grant hereafter) receive an average annual cash allowance of approximately 1,068 euros (or about 2,300 euros) for those living inside (outside) the family home. When students fulfill the Compensate Grant (T2 grant hereafter) requirements, the average amount increases on an additional 3,000 euros (3,500 euros) for those who live in (away from) their parents' home. Before 2013, there was an additional level of grant, the Displacement and Other Needs Grant (Threshold 3). This level of allowance

¹⁴See Ariño (2011).

¹⁵These estimates are based on the CJE (2015), using the rent prices offered by *Idealista.com* and the *Censo de Población y Viviendas de 2011*.

provided students with different cash endowments as displacement to the university, urban transport, academic material or final undergraduate degree project. A student who received this grant (T3 grant hereafter) could obtain only one or a combination of those different endowments.

Eligibility Rules. Students are eligible to the BCG grant if they are citizens of member states of the European Union, are enrolled in a Spanish higher education institution, and do not hold a degree of equivalent or higher level than the one they are applying for.¹⁶ Students can receive a BCG grant for at most one year more than the official length of the program which they are enrolled in, and for a maximum of two additional years of the program length for students who are enrolled in STEM (Science, Technology, Engineering and Mathematics) degrees.

Grant eligibility is assessed on the basis of student needs and academic performance. The need condition is evaluated on the basis of the applicant's annual household income the year before application, which is based on the after-tax household income. Qualification for a grant and the amount awarded depend on the students' household taxable income, as well as the number of household members.¹⁷ The applicant's annual household income is computed as the household taxable income minus specific quantities to which student's may qualify (such as large family or disability).¹⁸ The grant can be denied based on household income as well as when household wealth, family business activity and capital returns exceed certain thresholds.

Family income thresholds determine applicant's eligibility to different levels of grant depending on the number of household members. The fact that income eligibility thresholds change with the number of family members creates multiple

¹⁶From 2013 onward, students from post-compulsory degrees (such as college preparation or vocational track) in the educational system are also eligible. Detailed information about the students' eligibility rules is provided in *Real Decreto 1721/2007 de 21 de diciembre*, *Boletín Oficial del Estado (BOE)*. Furthermore, each courses application specifics are detailed in the *BOE: Orden EDU/1781/2010 de 29 de junio*, *EDU/2098/2011 de 21 de julio*, *Resolución de 2 de agosto de 2012*, *Resolución de 13 de agosto de 2013*, *Resolución de 28 de julio de 2014*, and *Resolución de 30 de julio de 2015*.

¹⁷The definition of a student's household includes the student's father, mother, siblings under the age of 25, grandparents, and the applicant. All of them are counted only if they live in the same family dwelling. In case of parental divorce, only the household members who live with the applicant are considered.

¹⁸For instance, if income sources are coming from any other household member but student's parents, the household is classified as large family, or there is a family disabled member, among others.

discontinuities, which are graphically displayed in Figure 1.A.1. To be eligible to the first and second levels of grant (fee waiver and T1 grant), for a household with four members (which is the average size in the sample), the annual family income must fall behind 38,831 and 36,421 euros respectively, which corresponds approximately to the fourth and top quintiles of the Spanish income distribution.¹⁹ To be eligible to the highest level of grant (T2 grant), the same household must earn less than 13,909 euros, which roughly corresponds to the bottom quintile of the income distribution in Spain.

The grant's academic requirement is met conditional on having passed a minimum fraction of credits the year before application. Applicants must be enrolled in at least 60 ECTS credits, which corresponds to the number of credits obtained in a typical academic year.²⁰

Application Process. The allowance is set up on a yearly application process that is common to all applicants. A summary of the application procedure follows:

- *July-early August:* the official call is made public in the Official State Gazette.
- *Mid August-Mid October:* applications are submitted to the Ministry of Education. The application form consists of an online questionnaire. No document transfer is needed since the Ministry contacts directly the institutions concerned, i.e., the Tax Authority and the university where the student is enrolled.
- *Mid November/December:* applications start to be answered for non-eligible students. Application outcomes are not necessarily disclosed at the same time for all applicants and answers are distributed along the academic year. Awarded and denied grants are notified in February–March of the academic year on average. Usually, the total amount granted is transferred to the students one month after the notification.

¹⁹Computations based on [de España et al. \(2017\)](#).

²⁰There are some special exceptions where students are allowed to be enrolled in less than 60 credits, e.g. when the attended program is made of less than 60 credits per year or when the student is affected by a disability.

The unique application process of this program allows to estimate the cash allowance effect on student performance with no concerns of enrollment effects that may bias the results. This is due to the fact that students are already enrolled in a higher education institution when they apply, and the vast majority of grant decisions are not notified before the end of the first term of the academic year. Hence, estimations are based on “intramarginal” students (students who would have enrolled in university irrespective of whether or not they receive financial support) and measure “intensive” rather than “extensive” margin responses.

The potential manipulation of information by applicants could be a concern for this type of allowances. It should be noted, however, that the authority handling the grant applications directly contacts the Tax Agency and the university in order to check applicants’ household income and academic status. Hence, students have only limited ability to misreport this information. A more serious concern is that students may be more likely to apply if they are below the income family thresholds, generating a discontinuity in application rates at the cutoffs. Before 2009, income eligibility thresholds changed every year, thus complicating applicant’s knowledge of their accurate situation, but over 2010–2015, income thresholds remained unchanged. Discontinuities in application rates would be more likely to occur at the Fee Waiver grant cutoff, since, at other levels of grants, students remain eligible for at least some form of aid (e.g., tuition fees) and hence have strong incentives to apply even if they anticipate being below the corresponding cutoffs. Moreover, the existence of multiple income reductions that affect the computation of students’ annual household income, makes it difficult for students to accurately evaluate their relative distance to the grant eligibility cutoffs. The complexity of the eligibility rules may encourage students to apply even in cases when they are unsure on whether they meet the criteria. Potential manipulation on eligibility threshold is discussed extensively in Section 5.1.

Changes to the BCG Grant: Period I vs. Period II. From now on I will refer to the three academic year terms of 2010–2012 as Period I, and the years 2013–2015 as Period II, concerning two different BCG frameworks. In 2013, academic requirements

were modified and a variable component was included.²¹ Overall, the changes between the two periods can be summarized as: (i) an increase in the minimum fraction of credits earned; (ii) the inclusion of the average GPA as a requirement; (iii) the incorporation of the grant's variable component which award students with different grant amounts depending on their GPAs and family income. For simplicity, from now on I will refer to academic requirements in Period I as *weak*, and these of Period II as *strong*.²²

First, in addition to the three main cash allowances, the framework in Period I included the T3 grant, which was based on a number of criteria such as distance to university, or educational material. In 2013, these different fixed amounts were merged into a single individual variable element, with conceded allowance if the student's family income was below the T1 threshold. The variable component of the grant is set at a minimum amount of 60 euros, and is computed as a deterministic function of the student's average GPA, the average GPA distribution of grant holders, the applicant's income, and the income distribution of all applicants in the year before application.²³ Therefore, the main difference across periods is the fact that students were receiving a large fixed amount of grant in Period II, and smaller fixed amount plus a variable allowance in Period II. Overall, the average amount of T1 grant was statistically the same across periods, but it was lower in Period II than in Period I for T2 grant (discontinuities in average grant amounts are discussed extensively in Section 5).

Second, the academic requirements set to be eligible for a grant became more stringent in Period II. Freshmen students must show an average grade in *PAU* of: (i) 5/10 points (corresponding to having passed the university entrance exam) to qualify for all grant levels in Period I,²⁴ (ii) 6.5/10 to qualify for all grant levels, and

²¹Detailed information about the change in BCG setup is provided at the end of this section and in the online appendix, section H.

²²A detailed summary of the policy change regarding academic performance requirements is provided in the online appendix, section H.

²³The exact formula of the variable component of the grant is provided in the online appendix (section H). The Ministry of Education offers an online simulator for the variables amounts at the following address: <http://www.mecd.gob.es/educacion-mecd/mc/becas/2016/estudios-universitarios/simulador.html>

²⁴In Spain the GPA can take values between 0 (the minimum grade) and 10 (the maximum). GPA's equivalence is the following: less than 5 points corresponds to a D grade, 5 points to a C

5.5/10 to be only eligible for the fee waiver allowance in Period II. Students who are not in their first year of higher education, must provide evidence on have passed a certain fraction of credits the year before applying:

- *Period I (2010–2012)*: 60 percent if the student is enrolled on a STEM degree, and 80 percent in non-STEM.
- *Period II (2013–2015)*: 65 (90) percent if enrolled on a STEM (non-STEM) degree to be only eligible for the fee waiver endowment. In order to qualify for all grant types, the student must have passed either: (i) 85 (100) percent if enrolled on a STEM (non-STEM) degree; or (ii) 65 (90) percent if enrolled on a STEM (non-STEM) degree, plus have obtained an average GPA of 6/10 (6.5/10) respectively for STEM (non-STEM) degree.

1.3 Data

Data. The data used in this paper are a combination of different administrative micro-data of BCG grant applicants over the six academic-year period 2010–2015, who were enrolled in Carlos III University of Madrid. I exploit the *SIGMA* database which consists of four administrative data files, which can be matched on the basis of an encrypted student identifier: (i) household information; (ii) awarded grants; (iii) university enrollment; (iv) grades in university. The household information database contains the set of variables that determine grant eligibility (household taxable income, number of family members, household wealth, family business, large family condition and whether a family member suffers a disability), the administrative status of the scholarship (grant final status, reason for denial and type of scholarship), and parent’s occupation. The awarded grants database provide details about the BCG grant amounts, the type of allowance and the date of award. The university enrollment database embrace information about grant applicants at the time they enter university, such as gender, nationality, postal code and the score in the *PAU* entrance exam. The database on university grades covers all academic curricula of grade, 7 points to a B grade, 8 to a B+, 9 to an A, and 10 to a A+.

students who have applied once or more to the BCG grant between 2004 and 2015, providing information on the department, degree and subjects in which each student is enrolled, as well as detailed information of each subject's course undertaken (e.g., final grade obtained, attendance of the final exam, retake).

Sample Restrictions. On average, 5,300 Carlos III students apply for a BCG grant in a given year. Table 1.B.1 displays the number of BCG applicants by year and degree program. The analysis is restricted to undergraduate students, who represent 93 percent of all applicants. Graduate students are not included in the analysis due to the small sample size. Moreover, I focus on students who were not denied the grant due to problems with the Tax Agency, were declared non-eligible due to excess wealth or business income, and meet with the minimum academic requirements in order to make the regression discontinuity design sharper.²⁵

1.4 Empirical Strategy

The goal is to estimate the causal effect of being eligible for a need-based grant on student performance and degree completion under two different grant settings. The estimates of a simple OLS regression of college achievement on a dummy variable indicating whether the student receives a grant would be subject to omitted variable bias, even after controlling for observable characteristics such as parental income, gender or predetermined ability measures, since the investigation would not account for unobservable determinants of student performance that are likely to be correlated with financial aid status (e.g., motivation).

To identify the treatment effect of being eligible for a need-based grant, I exploit the sharp discontinuities in the amount of cash allowances awarded using a regression

²⁵Students excluded from the sample of analysis represented 25 percent of the total applicants over the six-year period covered by the study (16 percent corresponds to problems with the Tax Agency and exceed the wealth and business thresholds, and 9 percent for not meeting the academic criteria). Excluding such students would be a problem if the probability of being denied a grant due to the reasons explained above was discontinuous at the grant eligibility cutoffs, thus leading to sample selection. This potential threat to identification is not a concern here, since rejection probabilities are continuous on either sides of the cutoffs (results available upon request). Moreover, discarding students who did not meet the minimum academic requirements do not change the statistical significance and magnitude of results (results available upon request).

discontinuity design (RDD). The BCG grant generates two different discontinuities at the T1 and T2 grant eligibility thresholds. Let $E_{i,k,t}$ denote a dummy variable that takes value one if applicant i is eligible for a grant of level k ($k = 1, 2$) at year t , and zero otherwise. Eligibility for a level k grant is a deterministic function of the applicant's net household taxable income c_{it} , and the number of family members, m_{it} :

$$E_{i,k,t} = \mathbb{1}\{c_{it} \leq \bar{c}_k(m_{it})\}; \quad (1.1)$$

where $\mathbb{1}\{\cdot\}$ is the indicator function and $\bar{c}_k(\cdot)$ is a deterministic function that returns the household taxable income threshold when the number of family members is m_{it} .

Let A_{it} denote the amount of conditional aid awarded to student i at time t . The total amount granted is determined as the sum of the different allowances increments $\alpha_{k,p}$ for which students are eligible at k level of grant in period p , where $p = 1$ for Period I (2010–2012) and $p = 2$ for Period II (2013–2015):

$$A_{it} = \mathbb{1}\{t \leq 2012\} * \sum_{k=1}^2 \alpha_{k,1} E_{i,k,t} + \mathbb{1}\{t > 2012\} * \sum_{k=1}^2 \alpha_{k,2} E_{i,k,t}; \quad (1.2)$$

The allowance increments in Period I and Period II are:

$$\alpha_{k,1} = \gamma_{k,1}; \quad (1.3)$$

$$\alpha_{k,2} = \gamma_{k,2} + z_i(c_i, \bar{c}_i, g_i, \bar{g}_i); \quad (1.4)$$

where $\gamma_{k,1}$ and $\gamma_{k,2}$ are period-specific fixed amounts, and $z_i(\cdot, \cdot, \cdot, \cdot)$ is a deterministic function that returns the amount of the variable component granted to applicant i with household where income is c_i and grades are g_i when average household income and average grades among applicants are \bar{c}_i and \bar{g}_i respectively.

The reduced-form equation capturing the relationship between the eligibility

formula and the outcome variable is:

$$Y_{it} = \alpha + \mathbb{1}\{t \leq 2012\} * \sum_{k=1}^2 \beta_{k,1} E_{i,k,t} + \mathbb{1}\{t > 2012\} * \sum_{k=1}^2 \beta_{k,2} E_{i,k,t} + \epsilon_{it}; \quad (1.5)$$

where Y_{it} is the outcome variable of student i at time t and ϵ_{it} are residuals of individual i at time t . In equation (9), the parameters $\beta_{k,p}$ are the treatment effects of being eligible for a grant k at period p .

Several identification assumptions are needed in order to identify a causal effect. I assume that the conditional distribution function is smooth in the forcing variable, and that there is no observed jump in the conditional probability of the outcome variable at every point in the household income distribution. Absent treatment, the outcome variable is a smooth function of parental income. Under this assumption, the causal effect of being eligible for a BCG grant of level k is identified by comparing outcomes for applicants who are close but below the eligibility income threshold (treatment group) with students who are near but above (control group). Thus, the local average treatment effect of being eligible for a BCG grant of level k relative to a grant of level $k - 1$, in period p , is:

$$\beta_{k,p} = \lim_{c \uparrow \bar{c}_k(m)} E[Y | c, m, p] - \lim_{c \downarrow \bar{c}_k(m)} E[Y | c, m, p]; \quad (1.6)$$

A specific feature of the BCG design is the existence of multiple income eligibility thresholds. In total, there are 22 distinct eligibility cutoffs for the T1 and T2 grants, depending on the applicant's household size (see Figure 1.A.1). To have sufficient statistical power, I pool all thresholds that are associated to a given level of grant²⁶. I use the relative distance to the income-eligibility threshold as forcing variable.

²⁶Note that the fee waiver eligibility threshold is close to the eligibility cutoff the the T1 grant (as observed in Figure 1.A.1) making difficult to construct two treatment samples (with sufficient number of observations) between T1 grant and fee waiver which do not overlap. The discontinuity induced by the tuition fee eligibility cutoff is therefore ignored in the main analysis. However, as a robustness check, I conduct a separate analysis of the treatment effect of tuition fee eligibility. The results (reported in the online appendix, section D) show no evidence of statistically significant effects on student outcomes at this income-eligibility threshold.

The treatment samples are defined as follows: (i) The first sample combines the eleven household taxable income cutoffs of the T1 grant. In this sample, I identify the treatment effects $\beta_{1,1}$ and $\beta_{1,2}$ of being eligible for an approximate average cash allowance of 675 euros and 825 euros respectively (the difference is not statistically significant) in Period I and II, relatively to being eligible to fee waiver only. (ii) The second sample combines the eleven parental income thresholds of T2 grant. In the second treatment sample, I identify the $\beta_{2,1}$ and $\beta_{2,2}$ treatment effects of being eligible for an approximate additional average cash allowance of 2,955 euros and 1,240 euros respectively (the difference is statistically significant) in Period I and II, relative to being eligible for about 1,400 euros on average. Figure 1.A.2 shows the amount of annual cash amounts awarded to applicants with 4 family members as a function of their parents' taxable income across periods.

Notice that the treatment effects are measuring the causal effect of a change in the amount of grant on student performance under certain academic requirements:

- $\beta_{1,1}$ and $\beta_{1,2}$ measure the causal effect of equivalent change in the average grant amounts under weak and stronger academic requirements. The change in the cash amount is the same but interacted with two different levels of performance standards. Comparing both estimates I can investigate the direct effect of two different academic requirements holding constant the change in the cash amount awarded. The comparison group in both estimates is awarded with fee waiver.
- $\beta_{2,1}$ and $\beta_{2,2}$ measure the causal effect of different changes in the average grant amounts under weak and stronger academic requirements. The common factor between both parameters is the fact that the comparison group is already awarded with non-zero the baseline average grant (approximately 1,400 euros). However, the changes in the amount of grant (1,700 euros of difference) and academic requirements are different.

The treatment effects are estimated using a triangular kernel.²⁷ The standard

²⁷Results are robust to using a rectangular instead of a triangular kernel. Results are available upon request.

errors are computed using standard least squares methods (robust standard errors) clustered at the student level.²⁸ The bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#).²⁹

Summary Statistics. Table 1.B.2 shows summary statistics on the sample of BCG grant applicants who are considered in the analysis. I split the estimation sample into two sub-groups: (i) the T1 grant sample (Threshold 1) includes applicants who are in the vicinity of the T1 grant threshold; (ii) the T2 grant sample (Threshold 2) includes applicants whose relative household income is close to the T2 grant threshold. Most of the applicants are Spanish, live with their parents when they entered university, and are enrolled in non-STEM degrees. The average household taxable income is approximately 32,000 euros for the T1 grant sample, and approximately 14,000 euros for the T2 grant sample. The average household size is four people and 11 to 17 percent of applicants in the treatment samples qualify for the large family bonus. The majority of applicants' family head member work as blue collars.

1.5 Theoretical Framework

In this section, I adapt [Bénabou and Tirole \(2000\)](#) principal-agent model with standards (i.e. academic requirements), and the extension of [Scott-Clayton and Schudde \(2019\)](#) introducing financial aid, to the different scenarios that this paper is able to test. These models are useful to rationalize the intuitions behind this paper.³⁰

²⁸Standard errors are clustered at the student level due to the fact that the same student may be observed several times in the same treatment sample if she applied more than once in the period studied.

²⁹A potential concern with the RDD regards the presence of treated (untreated) students for complying (not meeting) the academic requirements on both sides of the income-eligibility thresholds. I perform a robustness check testing the significance of the baseline results on a treatment sample that include those students. The results are robust to this test and are available upon request. An alternative potential empirical analysis to account for it may be to develop a two-dimensional RDD, with two running variables: relative distance to income-eligibility thresholds, and distance to the academic requirement thresholds. Two problems arise to implement this type of RDD. First, due to sample size limitations, separate estimations at each academic requirements threshold would be imprecise. Second, there are multiple academic requirement thresholds, since in the second period additional thresholds were incorporated in order to combine the fraction of credits earned and average GPA on the year before application. The presence of multiple dimensions of academic cutoffs reduces the sample size even more and complicate the identification. A normalization for all academic cutoffs may be a solution but results would be difficult to interpret.

³⁰[Lindo, Sanders and Oreopoulos \(2010\)](#) also use this model to relate it with academic probation.

Consider an agent facing a choice between three effort levels: low-effort (Task 1) with private benefit to the agent V_1 , effort cost c_1 , and principal's benefit W_1 ; high-effort (Task 2) with private benefit to the agent V_2 , effort cost c_2 , and principal's benefit W_2 ; zero effort with zero private benefit to the agent, effort cost, and principal's benefit. In the context of this paper, low-effort and high-effort level may be understood as the required effort to meet the *weak* and *strong* academic requirements respectively. Either effort level yields 0 to both subjects in case of failure such that:

$$0 < V_1 < V_2; 0 < W_1 < W_2; \text{ and } 0 < c_1 < c_2; \quad (1.7)$$

I further assume that:

$$\theta_1 \equiv \frac{c_1}{V_1} < \theta_2 \equiv \frac{c_2 - c_1}{V_2 - V_1} < 1; \quad (1.8)$$

Agent's ability is expressed as the probability of success in either option (θ), in which higher ability translates into higher probability of success. The probability of success is the same in both tasks. I assume that the agent knows θ but the principal does not. The prior cumulative distribution of θ on $[0, 1]$ is denoted as $F(\theta)$ with density $f(\theta)$. The agent solves:

$$\max \{0, \theta V_1 - c_1, \theta V_2 - c_2\}; \quad (1.9)$$

The agent chooses to shirk if $0 \leq \theta < \theta_1 \equiv \frac{c_1}{V_1}$, low effort level if $\theta_1 \equiv \frac{c_1}{V_1} \leq \theta < \theta_2 \equiv \frac{c_2 - c_1}{V_2 - V_1}$, and high effort level if $\theta_2 \equiv \frac{c_2 - c_1}{V_2 - V_1} \leq \theta < 1$.

Suppose now that the principal forbids the low-effort level task, such that the agent has to choose between shirking or exerting the high-effort level. The intuition behind this action is that the principal increases the academic requirements from *weak* to *strong*. With *strong* standards, the agent exerts effort if and only if:

$$\theta < \frac{c_2}{V_2} \equiv \theta^*; \quad (1.10)$$

Notice that the *strong* standards make types in $[\theta^*, \theta_2]$ to exert the high-effort level (*more ambitious*), but at the same time makes types in $[\theta_1, \theta^*]$ to shirk (*give up*). From the principal perspective, including *strong* standards would be worthwhile if the raise in the value due to those who are induced to exert higher effort overcome the loss of value induced by those who shirk:

$$S(\theta_1, \theta_2) = \left(\int_{\theta^*}^{\theta_2} \theta f(\theta) d\theta \right) (W_2 - W_1) - \left(\int_{\theta_1}^{\theta^*} \theta f(\theta) d\theta \right) W_1 > 0; \quad (1.11)$$

In this paper considers I test four different scenarios:

Scenario 1: No aid vs. low amount of aid under *weak* academic requirements.

I build on [Scott-Clayton and Schudde \(2019\)](#) model, which introduces financial aid to [Bénabou and Tirole \(2000\)](#). The principal provides the agent with financial aid $\phi(a)$, that is granted based on student enrollment and does not depend on other outcomes (i.e., it is available for agents who exert low and high effort levels). $\phi(\cdot)$ denotes a strictly concave function of financial aid, which can be interpreted as the agent's perception of having the grant level a . This scenario is similar to T1 Discontinuity under *weak* academic requirements (Period I). At the vicinity of this threshold, there is uncertainty on whether students will get the grant, and $\phi(a)$ is devoted to capture that. In this scenario, there is a new $\theta_1^a < \theta_1$, but $\theta_2^a = \theta_2$. The grant induces more individuals to exert low effort levels but none additional agent to high effort level. From the principal perspective, financial aid may be desirable relative to no aid under *weak* academic requirements if:

$$S(\theta_1^a, \theta_1) = \left(\int_{\theta_1^a}^{\theta_1} \theta f(\theta) d\theta \right) W_1 - \left(\int_{\theta_1}^{\bar{\theta}} f(\theta) d\theta \right) \phi(a) > 0; \quad (1.12)$$

Scenario 2: High vs. low amount of aid under *weak* academic requirements.

This scenario is similar to T2 Discontinuity, in which the agent's perception of getting some amount of grant is almost certain, under *weak* academic requirements (Period I). Without lose of generality, I assume that when students are at T2 Discontinuity, they get $\phi(a) \approx a$. It will end up with a new $\theta_1^{a'} < \theta_1^a$, but $\theta_2^{a'} = \theta_2^a = \theta_2$. Higher level

of grant induces more individuals to exert low effort levels but none additional agent to high effort level. From the principal perspective, high amount of financial aid may be desirable relative to low level of grant under *weak* academic requirements if:

$$S(\theta_1^a, \theta_1) = \left(\int_{\theta_1^{a'}}^{\theta_1^a} \theta f(\theta) d\theta \right) W_1 - \left(\int_{\theta_1^{a'}}^{\theta_1^a} f(\theta) d\theta \right) a > 0; \quad (1.13)$$

Scenario 3: No aid vs. low amount of aid under *strong* academic requirements.

The principal increases the academic standards from *weak* to *strong*. Then, the threshold for high effort option with financial aid declines to:

$$\frac{c_2 - a}{V_2} \equiv \theta^{*a} < \theta^* < \theta_2; \quad (1.14)$$

Notice that *strong* standards with low amount of financial aid make types in $[\theta^{*a}, \theta^*]$ to exert the high-effort level (*more ambitious*), and interestingly, do not make any additional agent to shirk (*give up*). This scenario is similar to T1 Discontinuity under *strong* academic requirements (Period II). From the principal perspective, financial aid may be desirable relative to no aid under *strong* academic requirements if:

$$S(\theta^{*a}, \theta^*) = \left(\int_{\theta^{*a}}^{\theta^*} \theta f(\theta) d\theta \right) W_2 - \left(\int_{\theta^{*a}}^{\theta^*} f(\theta) d\theta \right) \phi(a) > 0; \quad (1.15)$$

Scenario 4: High vs. low amount of aid under *strong* academic requirements.

This scenario is similar to T2 Discontinuity under *strong* academic requirements (Period II). It will end up with a new $\theta^{*a'} < \theta^{*a}$. Higher level of grant induces more individuals to exert high effort levels but none additional agent to shirk. From the principal perspective, high amount of financial aid may be desirable relative to low level of grant under *strong* academic requirements if:

$$S(\theta^{*a'}, \theta^{*a}) = \left(\int_{\theta^{*a'}}^{\theta^{*a}} \theta f(\theta) d\theta \right) W_2 - \left(\int_{\theta^{*a'}}^{\theta^{*a}} f(\theta) d\theta \right) a > 0; \quad (1.16)$$

Overall, financial aid reduce the agent's private cost of effort, inducing students to exert higher levels of effort than a non financial aid setting. Introducing financial aid

would make some students types *more ambitious* and exert either low or high effort levels depending on whether academic standards are *weak* or *strong* respectively. Increasing academic requirements would induce some agent's types to exert higher levels of effort, but at the same time, it may induce weaker agent's types to shirk and potentially drop out from higher education. It seems that a minimum level of standards is desirable, but determining whether it is too high or rather low would require weighting the value of encouragement and discouragement effects.

The testable implications of this theoretical framework are the following: (i) Financial aid attached to *strong* academic requirements increases the overall level of effort relative to financial aid under *weak* standards; (ii) Higher additional levels of financial aid increases the overall level of effort relative to lower aid, irrespective of academic requirements; (iii) Financial aid attached to *strong* academic requirements increases the overall probability of students dropping out relative to financial aid under *weak* standards; (iv) *Strong* academic requirements induce relatively high-ability students to work harder and increase their level of effort, while relatively low-ability students to dropout. The direction and significance of the effects of an increase in the academic requirements would depend on four dimensions: the magnitude of a , the shape of the ability distribution, the relative benefits of high effort versus low effort, and the relative benefits of low effort versus shirk.

1.6 Results

1.6.1 Internal Validity of the Empirical Strategy

The internal validity of the RDD requires that there is no endogenous sorting on either side of grant eligibility cutoffs. The forcing variable is the relative distance to the household after-tax income cutoff. This type of endogenous sorting is more likely to occur in the common case where the treatment assignment rule is public knowledge [Imbens and Lemieux \(2008\)](#), as in this paper. As precise thresholds are public information and have not changed since 2010, a concern of manipulation at the cutoff arises especially for the first income-eligibility threshold (fee waiver). In

contrast, manipulation is less likely at higher cutoffs, since students have incentives to apply on either side given the fact that students on both sides are eligible for a positive amount of aid. [Montalvo \(2018\)](#) highlight the fact that after-tax income is more difficult to manipulate than income, and in Spain the changes in the tax code are frequent.

Figure [1.A.3](#) shows the graphical representation of the density estimates in the vicinity of the cutoffs, displaying that it does not seem to be systematic manipulation of household parental income around the thresholds. The density of applicants increased as parental income decreased in T1 grant, given the fact that more students may be encouraged to apply as they were closer to the cutoff. Density estimates at T2 grant were roughly constant, since applicants have incentives to apply on both sides as they would be awarded with a positive cash allowance. The test statistics proposed by [McCrary \(2008\)](#) fail to reject a statistically significant jump at the eligibility cutoffs for any of the treatment samples used in this paper (i.e. period, gender, or predetermined ability).³¹

An additional test for local random assignment is to check whether applicants baseline characteristics are “locally” balanced on either side of the thresholds. If some groups of students are more likely to sort on the “high” side of a threshold may indicate endogenous sample selection, and treatment assignment cannot influence variables that are predetermined with respect to the treatment. Local linear regressions are performed for each of the applicants’ observable characteristics (i.e., gender, nationality, parental income, *PAU* score, or parents’ occupation) as dependent variable. Panel A of Table [1.B.3](#) presents the regression results, showing that the observable characteristics of applicants are well balanced, since less than 10 percent of the variables do change discontinuously at income eligibility thresholds. Furthermore, a chi-squared test based on a system of seemingly unrelated regression with as many equations as baseline characteristics is performed. Panel B indicates that the null hypothesis that the discontinuity gaps are jointly equal to zero cannot be rejected.

An additional concern is that parental income, at constant prices of 2015, is

³¹See online appendix (section B) for details of McCrary test’s estimates for all treatment samples.

highly correlated over time (regressing applicants' income in a given year on income the year before leads to a coefficient estimate of 0.73), which may lead to a persistent sorting of applicants on either side of eligibility cutoffs and may confound the effects of current year discontinuities in grant amounts with those from previous years. In fact, there is variation in income, since the fraction of applicants who reported the same parental income than the one registered the year before is only 3.2 percent. Students' who were awarded a grant in a given year might be more likely to re-apply the next year. It might be that impacts would be driven by this group, with no density break for applicants at the cutoffs but so for re-applicants. A robustness check testing the discontinuity in the density of re-applicants cannot reject the null hypothesis of zero discontinuity in the density of re-application.³²

1.6.2 Discontinuities in Grant Amounts

In this subsection, I examine the discontinuities in average grant amounts awarded of the income-eligibility thresholds, which is a necessary condition for the empirical design to identify the causal effects of grants on student outcomes.

Figure 1.A.4 shows the average fraction of applicants who were awarded either a T1 or T2 grant plotted against the relative income-distance to the relevant eligibility thresholds. The figure indicates that approximately 98 percent of the theoretical eligible applicants received the grant.

Figure 1.A.5 presents the average conditional grant amount for all treatment samples as a function of applicants' relative distance to the thresholds separately for the two periods under study. The results indicate a clear discontinuity in the average conditional cash allowance for T1 and T2 grants in both periods, which is confirmed by the statistically significant results showed in Table 1.B.4 (Panel A). T1 grant provides not statistically significantly different average grant amount for both periods, with an average cash amount of 675 euros in Period I and 825 euros in Period II (relatively to been awarded with fee waiver). T2 grant reports a drastic decrease in the average grant amounts awarded across periods, with an average increment in

³²See online appendix, section B.

the cash amount of 2,955 euros in Period I and 1,240 euros in Period II (relatively to T1 grant cash awards), and both estimates are statistically significantly different.

1.6.3 Impact on Dropout

Most of the literature focuses on the extensive margin effect of grants on enrollment. When there is an effect on enrollment (which is often stronger for freshmen students), disentangling the intensive margin response on performance is challenging (due to the potential selection bias that the enrollment effect provides). An advantage of the setting under study is the specific timing of grant applications in Spain, which allows to estimate the effects of grants on students who are already enrolled, and for whom dropout rates are relatively small. Table 1.B.5 displays the RDD estimates on dropout from higher education. The null hypothesis of a zero effect of cash allowance on dropout cannot be rejected for all types of grants and periods. The results are suggesting that the effect of grant on student performance do not seem to be biased by dropout effects. This result is consistent with Montalvo (2018), finding no effect on dropout of a sharp increase in tuition fees in a highly comparable setting.

1.6.4 Impact on Student Performance

I focus on the average GPA, which in Spain can take values between 0 (the minimum grade) and 10 (the maximum)³³ and the fraction of credits that the student passed among the total attempted credits as measures of student performance.

Figure 1.A.6 plots the average GPA for all treatment samples as a function of applicants' relative income-distance to the thresholds separately for both periods studied. The solid black lines are the fitted values from a linear projection. The average GPA is slightly different across periods for the two samples of applicants (around the T1 and T2 grant thresholds respectively). The average GPA was around 5.9 points in Period I and 6.15 points in Period II. Table 1.B.4 presents the non-parametric RDD estimates.

³³GPA's equivalence is the following: less than 5 points corresponds to a D grade, 5 points to a C grade, 7 points to a B grade, 8 to a B+, 9 to an A, and 10 to a A+.

Result #1: No effect of neither small nor large discontinuous change in grant amounts when interacted with *weak* academic standards. I find no effect of relatively large cash allowance (neither for 675 nor for 2,955 euros) on student performance in Period I, when performance requirements were comparable to the performance incentives that characterize the typical need-based grant programs around the world (*weak*). This result is significant, since 2,955 euros of increase in the amount of cash allowance is a large grant amount in comparison with the average amounts awarded in other related papers under similar frameworks. Students do not seem to be particularly sensitive to large changes in cash amounts when interacted with academic requirements that are *weak*.

Result #2: Positive effect of relatively small discontinuous change in grant amounts when interacted with *strong* academic standards. I find that being eligible for an average of 825 euros grant (relatively to being eligible for only fee waiver) interacted with *strong* academic requirements, increases student average GPA and fraction of credits earned by 0.45 points and 6 percentage points respectively, which corresponds to an increase of approximately 7.3 and 7.6 percent with respect to the baseline mean. The T1 grant threshold offers a unique opportunity to analyze the role of academic requirements in the total effect of financial aid on performance. Since average grant amounts are not statistically significantly different between Period I and Period II, I can test the effect of the same grant amount under two different intensities of academic requirements. The effect of a change in the grant amount is not statistically significant when academic requirements are *weak*, but statistically significant when *strong*. I use a Difference-in-Difference-RDD to test whether the estimates are not statistically significantly different across periods. The null hypothesis that both effects are equal across periods is rejected at the 1 percent confidence level. The results suggests that academic incentives augment the effectiveness of aid in improving student performance for those students who are changing from zero to some positive cash amount.³⁴

Result #3: Results seems to persist over two consecutive academic years.

³⁴Results of the Difference-in-Difference RDD are displayed in the online appendix, section E.

Being eligible for a grant may have dynamic effects over a student's academic career. Grants awarded in a given year may produce long-lasting effects, impacting students' outcomes in several subsequent years. I compute the effect of being eligible for a grant on applicants' cumulative performance over several academic years: conditional on applying for a grant at time t (with certain level of household income), it is possible to compute the cumulative average GPA and fraction of credits earned over subsequent years. This method would provide potential unbiased estimates and less concern for sample selection, but potentially the first stage would decrease over time due to the variability of students' application status and household income.³⁵ Local linear regression estimates indicate that being eligible for a grant under *strong* performance-based incentives increases the cumulative average GPA and fraction of credits earned over two years by 0.5 points and 6.3 percentage points per year respectively, which corresponds to an increment of about 8 and 7 percent per year with respect to the baseline mean. The eligibility for a grant interacted with strong performance standards seems to have a positive impact on student performance that last for two consecutive years.

Result #4: Large discontinuous change in grant amounts when students are already awarded with non-zero financial aid has no differential effect, irrespective of the academic requirements. Students do not seem to react differentially to discontinuous changes in the amount of grant when they are already receiving certain amount of cash, neither when academic requirements are *weak* nor *strong*. The interpretation of the non-significant results at the T2 threshold is ambiguous. It may be due to the fact that academic requirements affect differently students who are entitled to different levels of grants (students behavioral response to incentives may be more powerful when they start receiving some amount of grant), though the incapacity of students in the lowest part of the income distribution (T2 grant applicants) to react to incentives due to their potential lack of ability for developing effective study strategies, or the fact that positive effect of stronger academic requirements may be offset by the potential negative effect of large decrease

³⁵First stages and a test for discontinuity in the density function of the running variable at the cutoff are presented in the online appendix, section C.

in the cash allowance awarded between Period I and Period II.³⁶

1.6.5 Robustness Checks

In this subsection, I perform a number of tests in order to check the robustness of the baseline estimates: (i) investigate the sensitivity of estimates to the choice of bandwidth; (ii) perform the local polynomial regression with robust bias-corrected confidence intervals proposed by [Calonico, Cattaneo and Titiunik \(2014\)](#); (iii) run the baseline regressions adding student individual predetermined variables and year fixed effects; (iv) test for jumps at non-discontinuity points by running placebo regressions; (v) check for an effect of being eligible for only fee waiver; (vi) investigate the effect of being eligible for T3 grant in Period I; (vii) analyze potential effects in 2012 where academic requirements were slightly modified. Overall, baseline results are robust to all different specifications and vary from an effect of 0.27 to 0.5 points, which corresponds to about 4.5 to 8.3 percent with respect to the baseline mean. Although the magnitude of estimates varies across specifications due to the limited sample size, the direction of the effects hold over the different specifications, indicating a robust impact of grant eligibility on student performance when the academic standards are *strong*. The differences between the robustness checks and the baseline estimates are not statistically different. In addition, the null effect of the grant under the other different thresholds (T2 and fee waiver grant) and periods is also robust and persistent for each sensitivity check performed.³⁷

Sensitivity to the Choice of Bandwidth. I analyze the sensitivity of the non-parametric estimates to the choice of the bandwidth changing the bandwidth size to half and twice the value of the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#). Results are very similar to those obtained in the baseline estimates, but larger when using half the optimal bandwidth than double.

Local Polynomial Regression with Robust Bias-Corrected Confidence Intervals.

To test for the variability of the results under local polynomial regression and a

³⁶The interpretation of the results is discussed extensively in the Discussion section of the paper.

³⁷See online appendix (section D) to see the details of robustness checks.

different computation for confidence intervals (robust bias-corrected), I test the effects using the optimal bandwidth proposed by [Calonico, Cattaneo and Titiunik \(2014\)](#). The point estimates are very similar to the baseline estimates and the difference is not statistically significantly different.

Individual Control Variables and Year Fixed Effects. I investigate the volatility of baseline results when adding individual predetermined students' characteristics (e.g. *PAU* percentile rank, gender, or being enrolled in a STEM degree) and year fixed effects that capture time trends in the outcome variable. Results are statistically significant at the 5 percent confidence level with the point estimate smaller than baseline estimates, though the difference is not statistically significantly different.

Testing for Jumps at Non-Discontinuity Points. To test for jumps at non-discontinuity points, I run a placebo regression in which the income thresholds are artificially set at the midpoint between the actual eligibility thresholds by period. Since these midpoint do not correspond to any change in applicants' grant eligibility status, I should expect to find no significant jumps in average GPA. Points estimates are close to zero and non-significant in all specifications.

Fee Waiver Grant (Threshold 0). The fee waiver is the first type of grant that students may receive, and covers the tuition fees but does not award with an amount of cash. This eligibility threshold (T0) is very close to the eligibility cutoff of T1 grant. Hence, it is difficult to construct two treatment samples between T1 grant and fee waiver which do not overlap. The discontinuity induced by the tuition fee eligibility cutoff was ignored in the analysis, in order to focus on grants types where students were awarded with certain cash amounts (in T1 and T2 grant). I conduct a separate analysis of the treatment effect of being eligible for only tuition fee. Results shows no evidence of statistically significant effects on awarded cash amounts and student performance at this income-eligibility threshold.

Displacement and Other Needs Grant (T3): The importance of performance standards intensity. Figure [1.A.5](#) and Figure [1.A.6](#) display the fact that with equal average cash amount granted, an allowance setting with strong performance-based components seems to be more effective to enhance student performance, as opposed

to a setting with weak incentives. An additional robustness check is testing whether there is some statistically significant effect at T3 threshold (working only in Period I) where the amount of cash awarded was similar to the one awarded for non-movers in Period II. When academic requirements are weak, there is no statistically significant effect of being eligible for a grant in every income-eligibility thresholds (T3, T2 and T1), which reinforce the baseline results of zero effect in the absence of strong requirements.³⁸

Differential effects on 2012. This section test for specific effects of the grant in 2012. While academic incentives in Period II were homogeneous throughout the three academic years, Period II reported a change in 2012. In 2012 the requirements rose to 65 (90) percent for students enrolled in STEM (non-STEM) degrees. The null hypothesis of zero effect of the grant on student performance and dropout cannot be rejected. The results suggest that a single increase in the fraction of credits earned is not enough to affect student performance. This points toward the apparent important role that design rather than the existence of certain academic incentives have on stimulating student performance.

1.6.6 Heterogeneous Effects of Grants on Student Performance

Despite of the robust baseline estimates, investigating the existence of heterogeneous results for academic term (Fall vs. Spring grades) and different subgroups of population (gender, predetermined academic ability and residence status) is necessary to understand the implications of the estimated effects.

Student Performance by Academic Term. Students are already enrolled at the higher education institution when they apply, and the vast majority of grant decisions are not notified before the end of the first term of the academic year. In addition, conceded grants were divulged between February-March on average. This unique process may create an unclear view of when the grant incentives are created to improve student performance, due to the fact that students faced a different timing of acceptance/rejection disclosure. An ideal way to test whether there is heterogeneous

³⁸For further analysis of this robustness check see online appendix, section D.

effects of the grant on student performance before or after the student received the notification, it is to compare the effect on students who received it before the term exams versus those who were informed after. Unfortunately, this sample split creates endogenous selection at the eligibility cutoffs, since denied grants were disclosed before accepted grants on average, leading to a significant break in the density at income thresholds.³⁹

An alternative way to test it is to look at the impact of the grant on student performance by academic term (Fall and Spring). Students had a higher probability to get an answer on the second rather than on the first term. Then, it is reasonable to believe that student reaction to the allowance would be stronger for Spring than Fall grades. Table 1.B.6 presents the non-parametric estimates by academic term, confirming this hypothesis. The effect sizes are larger for the second than first term. Although the difference is not statistically significant, the estimates for Fall Term seem to be more sensitive to the functional form, as can be seen in Figure 1.A.7 and Figure 1.A.8.

Student Performance by Subgroups of Population. Table 1.B.7 and Table 1.B.8 presents the RDD estimates for T1 and T2 grants by period and subgroups of applicants. The positive effect of the T1 grant on student performance coupled with more stringent performance incentives are found for both males and females, but the magnitudes differ (Panel A). The point estimate is statistically significant for males, but it is not statistically different of female's.

Panel B explores heterogeneous effects by level of academic ability, dividing the samples into two groups based on applicants' percentile rank on *PAU* exam. Being eligible for a T1 grant in a setting with high incentives has significantly positive effects on student performance for students above the median percentile rank of the *PAU* exam score distribution. The null hypothesis that both effects are equal across groups below and above the median cannot be rejected.

Panel C presents the results divided by the different applicants' residence conditions. It seems that the positive impacts are driven by both applicants who were

³⁹Notice that student disclosure time is not a perfect continuous variable, but rather discrete, since groups of students were receiving notification at the same time as they were sent in blocks.

living with their parents in the Region of Madrid (non-movers hereafter) when they enter university and students who were living away from their family home (movers hereafter). The null hypothesis of equality of coefficients cannot be rejected. While students living away from their family home receive positive amounts decreasing in the second period (from 2,858 to 1,673 euros on average), applicants who live with their parents earned a zero amount in the first period and 445 euros on average in the second period.⁴⁰

The grant structure allows to consider the first period of non-movers as a placebo test, when this group of applicants did not receive a positive average amount and incentive components were weaker. A change from zero to a positive cash allowance of 445 euros, interacted with strong academic incentive components leads to a positive impact on students' average GPA of 0.45 points (7 percent with respect to the baseline mean). In addition, despite of the reduction on the amount of grant awarded for movers across periods, being eligible for an average grant of 1,673 euros (relatively to being eligible for fee waiver) under strong academic requirements, increases students' average GPA by 0.52 points (8 percent with respect to the baseline mean). The null hypothesis that both effects are equal across movers and non-movers cannot be rejected.

1.6.7 Mechanisms

This section investigates whether the performance increments observed are due to an actual rise in student success and not to other confounding factors (e.g. students taking easier subjects). Table 1.B.9 presents the results of the RDD estimates on different outcomes for T1 and T2 by period, and Figure 1.A.9 plots these different channels, reinforcing the results than the non-parametric estimates.

Final exam attendance rate. Panel A shows that being eligible for an average cash allowance of 825 euros (relative to fee waiver only) increases the final exam attendance rate when performance incentives were more demanding. Although the average fraction of grant applicants who attended final exams was already high (93

⁴⁰See online appendix, section C.

percent), qualifying for such level of grant enhances this average by 3.2 percentage points, which corresponds to an increase of about 3 percent with respect to the baseline mean.

Fraction of Subjects in Retake. Eligible students may increase their attendance rate to final exams through attending more often to retakes. Panel B present that being eligible for an average cash allowance of 825 euros (relative to fee waiver only) decreases the fraction of subjects that students have to retake by 4.6 percentage points, which corresponds to an increase of about 25 percent with respect to the baseline mean when interacted with *strong* academic requirements. Students seem to be increasing their fraction of credits earned in final exams directly, and not through an increase in the fraction of retakes.

GPA on final exams taken. Students who received the BCG grant may attend to final exams with higher frequency, while their performance on them may remain unchanged. The fact that students show-up more often to final exams may enhance their total average GPA due to the less frequent inclusion of subjects graded as zero points (grade given to students who fail to attend) in the total average GPA computation, rather than to an actual improvement in their performance. To test this hypothesis, I examine the discontinuity in the average GPA on final exams taken, in order to capture the increase in performance over the subjects that students took. Results in Panel C shows that students who were eligible for an average cash allowance of 825 euros (relative to fee waiver only) under a setting with strong performance-based incentives improved their average GPA in final exams taken by 0.35 points, which corresponds to an increase of about 5 percent with respect to the baseline mean. The null hypothesis of zero effect is rejected at the 1 percent confidence level, indicating that the observed improvement in average student performance is genuine.⁴¹

GPA on mandatory and elective subjects. I investigate the differential results on the average GPA for mandatory and elective courses as an additional test. Students are required to pass a certain number of elective courses chosen from a determined

⁴¹However, this result seems to be slightly sensitive to the functional form used.

set of subjects which are specific to every degree, and several mandatory courses which are compulsory and degree specific. If applicants were to self-select in easier subjects, it is reasonable to expect that it would for elective courses, due to the fact that such courses are those for which students can select different subjects. Panel E and F analyze the effects of the grant on the average GPA for mandatory and elective courses. Applicants who were eligible for an average cash allowance of 825 euros (relative to fee waiver only) under strong performance-based incentives, increases their average GPA on those subjects which are mandatory by 0.46 points, which corresponds to an increase of 7.5 percent with respect to the baseline mean. Despite the higher average GPA on elective courses compared with mandatory, the null hypothesis of zero grant effect on average GPA in elective courses cannot be rejected.⁴²

1.6.8 Impact on Degree Completion

Table 1.B.10 expands the analysis by investigating the impact of financial aid on degree completion. The table focuses on students who applied for the grant in the final year of a degree program, i.e., in their fourth year of a bachelor's degree. The non-parametric estimates indicate that being eligible for 825 euros (relatively to the tuition waiver) in the period when performance requirements were more stringent, increases student's chances of obtaining a degree in 12.5 and 11 percent with respect to the baseline mean for all applicants and if the applicant is on the graduation year respectively. In contrast, the null hypothesis of zero effect on degree completion under a setting with *weak* academic requirements cannot be rejected.

1.7 Discussion

The Spanish national need-based grant program provides a unique design to analyze the causal effect of this form of financial aid on student performance and graduation.

⁴²Investigating the point estimate and the standard errors, it seems plausible that there is no statistically significant effect on the average GPA of elective courses due to sample size limitations in order to identify a smaller minimum detectable effect.

I find no effect of relatively high cash amounts on student performance (average GPA and fraction of credits earned) and degree completion in a setting with *weak* performance incentives comparable to the typical need-based grant programs around the world. These results are consistent with papers finding no effect of similar Pell Grant’s cash amount on student GPA interacted with relatively similar academic requirements (Denning, Marx and Turner, 2017; Denning, 2018). In contrast, I find that an average provision of 825 euros cash allowance (relatively to receiving only fee waiver) increased student performance and probability of degree completion in a setting with more demanding performance-based incentives, but zero effect at a higher level of grant. The size of the effects is similar to Scott-Clayton (2011), the most related paper. Effects are larger than Angrist, Oreopoulos and Williams (2014), and slightly lower than those found only for women in Angrist, Lang and Oreopoulos (2009). Students enhance their final exam attendance rate and their GPA on final exams taken. Eligible students reduce the fraction of subjects that have to retake and the total amount of credits in which they enroll. I ruled out the hypothesis that results are driven by student selection of less demanding courses or dropout from higher education.

1.7.1 External Validity

A potential important concern of this paper is the fact that results might be difficult to generalize to other settings. The estimates are based on a sample of low-income high school graduates enrolled in Carlos III University who applied to a BCG grant to start or to continue undergraduate college studies. Carlos III University is a public higher education institution. An analysis comparing the educational attainment of Carlos III students with the rest of collegians enrolled in Spanish public universities presents that these students scored higher in the standardized university access exam, reported higher graduation rates and presented lower dropout rates (among other measures) than their counterparts in the rest of Universities in Spain. However, the group of students who drive these differences are non-BCG grant recipients. In contrast, BCG grant recipients in Carlos III are highly comparable with

BCG grant recipients in Spain, reporting similar GPAs, number of credits passed (over the total credits enrolled and the final exams taken), and time to graduation.⁴³ The sample of grant applicants is reasonably representative of the general population of low-income students in Spain. This group can be considered as comparable to the typical targeted population of most large-scale need-based grant programs around the world (e.g., students graduated from high school and admitted to college). The results cannot be directly extrapolated to the population of low-income students who fail in high school graduation and might respond differently to financial aid, whereas they can be comparable to non-traditional students.⁴⁴

The institutional features of higher education systems are decisive for the external validity of the results. Spain is part of the group of countries (along with France, Italy, Belgium or Austria) where post-secondary systems are mainly public and tuition fees charged are relatively low (OECD, 2016). In these countries, the student level of debt is considerably low, and the need-based programs cover tuition fees and part of the living expenses for low-income students. The results of this paper cannot be immediately compared with the US students who are not eligible to fee waivers (e.g. the Pell Grant), due to the fact that they present substantially larger levels of debt, higher tuition fees, and greater probabilities of working to pay for college. On the other hand, the effects of this paper are potentially comparable to the population of US students who are entitled to both fee waiver and need-based grants.

1.7.2 Interpretation of Results: Efficiency and Equity

Efficiency. This paper points out the importance of minimum academic requirements on need-based grants' cost-effectiveness. From the efficiency point of view, a program with strong performance standards presented a differential performance improvement for those students whose comparison group was receiving only fee waiver and zero cash amount. In contrast, the grant did not seem to affect differentially those students whose comparison group was awarded with a certain cash amount, neither under

⁴³Details of the analysis are provided in the online appendix, section A.

⁴⁴The Spanish BCG grant eligibility criteria does not impose any upper age limit, neither does the US Pell Grant.

a setting with *weak* nor *strong* academic standards. Three hypothesis may help to interpret the results.

First, the null impact of the grant could alternatively be due to the fact that students at the bottom of the family income distribution may not be as able as their peers to respond to performance incentives, even under fairly large cash amounts granted. This result is consistent with [Fryer \(2011\)](#), which found no effect of financial incentives on student achievement on a sample of urban schools in the U.S. Perhaps, the muted effectiveness of the grant on those students may partly reflect the trouble struggling students have developing effective study strategies ([Angrist, Oreopoulos and Williams, 2014](#); [Daly and Lavy, 2009](#); [Fryer, 2011](#)).

Second, students behavioral response to incentives may be more powerful when they start receiving some amount of grant. Effects seem to be large with small amounts of grant when changing from zero to certain positive amounts, but no differential effect is found when students are entitled to additional amounts. This result is consistent with [Fack and Grenet \(2015\)](#), a study in which the largest effects of the French national need-based grant program were concentrated on those students receiving the first endowment of cash allowance and not in students entitled to higher levels of grant. In addition, several studies find large behavioral responses to small-scale interventions ([Bettinger et al., 2012](#); [Hoxby, Turner et al., 2013](#)). Incentives seem to work when there is some entitlement to the grant, but their effect is not clear with additional grant amounts. Interestingly, additional large changes in grant amounts when students are already awarded with non-zero financial aid, have no differential effect neither when interacted with *weak* nor *strong* academic requirements. Specially relevant is the case of the period with *weak* academic requirements, in which students increase their cash amount in 3,000 euros from 1,400 euros approximately, a large differential amount in relative terms to what is found in the literature of financial aid. Thus, student performance does not seem to be particularly sensitive to monetary incentives, and this result may potentially rule out the cost-of-college channel, though I cannot test this hypothesis directly in this

paper.⁴⁵

Third, the natural experiment is not exactly the same for those students at the bottom of the parental income distribution, since both cash amounts and academic standards changed across periods. Then, it is not possible to investigate whether those students would react differently under the same amount of grant interacted with *weak* and *strong* academic requirements. I cannot reject the hypothesis that the large decrease in average amounts across periods may offset the positive effect induced by stronger academic standards.

Equity. The equity implications of increasing academic requirements on grants devoted to low-income students may be ambiguous. The intuitions of the [Bénabou and Tirole \(2000\)](#) model, increasing standards may encourage additional students to exert higher effort, but other lower ability students may give-up and dropout from higher education. An extension of this model including financial aid ([Scott-Clayton and Schudde, 2019](#)) show that discouragement (dropout) would be concentrated among those in the lower part of the ability distribution, while encouragement (improved GPA) may be concentrated among those who are close to the performance requirement threshold. The significance and magnitude of these effects may be student (e.g., distribution of parental financial constraints, student ability) and country specific institutional dependent (e.g., cost of college).

First of all, it is important to elucidate whether the increase in academic requirements was binding. Unfortunately, I cannot directly measure the actual stringency of the policy, given that it is unlikely that students apply for a grant if they know that do not meet the academic requirements. It would be reasonable to expect that the annual fraction of applicants who are not eligible for performing below the academic requirements grows when those increase, but this number was about 26 percent and was fairly constant over the period.⁴⁶ However, an approximation may be undertaken in order to proxy the tightness of academic requirements. For income-eligible applicants who apply for a grant in 2010 and 2011, I measure that 23

⁴⁵This channel implies that the relaxation of budget constraints may prevent financially constrained students from working part-time, inducing them to devote more time to study.

⁴⁶Note that this is not a problem of internal validity, but to address how binding the policy was.

and between 35–43 percent would not meet the academic requirements if they were *weak* and *strong* respectively. For first year students, I compute that 30 and 47–55 percent would not be eligible to renew the grant in those two scenarios. Although the previous estimates contains measurement error, they seem to indicate that the policy was binding.⁴⁷

To determine the net size of this increase, the positive effects in performance induced by the grant with *strong* academic requirements should be taken into account, which potentially make some students to meet the new threshold (inducing them to be *more ambitious*). I compute that 8 (11) percent of all (first year) students would have meet the requirements as a consequence of being affected by the average effect of the grant with *strong* standards on fraction of credits earned and average GPA, under the assumption that the average effect would be constant on the distribution of income-eligible students. This measure raises to 11 (14) percent of all (first year) students, when looking at the counterfactual of those students who start receiving some cash amount (students who were eligible to only fee waiver and not cash amounts). Then the increase in the fraction of potentially eligible students who would not keep the grant does not seem to be drastic, since the positive effect of the grant with *strong* requirements counteract the negative impact of the raise in requirements by making some students *more ambitious*.

In order to analyze whether the increase in the academic requirements have statistically significant effects on student dropout from higher education (whether induced some students to *give up*), a more thoughtful investigation is required. I perform a RDD to observe differential changes in dropout at the income-eligibility cutoffs, a Difference-in-Difference model (DID), a RDD-DID specification, and a Logit analysis to measure the propensity to dropout before and after the change in requirements. I do not find significant equity effects of an increase in the academic requirements. At the vicinity of income-eligibility cutoffs, there is neither evidence of a differential effect of the grant on dropout from higher education (RDD estimates), nor a change in the differential effect before and after the reform (RDD-DID). Results

⁴⁷[Schudde and Scott-Clayton \(2016\)](#) calculate that between 25-40 percent of first year Pell recipients at public institutions were placed at risk of losing financial aid.

present no heterogeneous effects by gender, predetermined measure of student ability and residence status. These results show the local average treatment effects estimated with the RDD, but do not provide average effects for the total distribution of grant applicants. To address whether students would dropout as a consequence of not meeting the *strong* academic requirements, I perform a DID analysis comparing students who performed higher to the *weak* academic requirements but not overcame the *strong* (Treatment Group), with students who do not meet the *weak* academic requirements (Control Group), before and after the reform. I find no statistically significant effect of the reform on drop out from higher education, showing that the group of students who would not meet the requirements due to their increase do not significantly give up and dropout. An additional Logit analysis on the full population of grant applicants over 2010–2015 shows that the probability of dropout from higher education does not statistically significantly change after the increase of requirements. These results seem to be robust for first and non-first year students. In addition, the ability of students who apply for a grant (measured as the percentile rank in the access to university exam) does not change significantly after the reform. Stronger performance standards do not seem to have a significant equity effect on applicants to need-based financial aid.⁴⁸

To interpret the results across periods, it is important to address the issue of the comparability of applicants across periods. Note that this is not an internal validity concern, but an exercise to understand how similar the populations of grant applicants are across periods. As performance requirements are different by period, the type of students who apply for a need-based grant might significantly differ, comparing potentially different student types across periods. In an attempt to test the comparability of the two periods for each of the allowances studied, I display a t-test of the difference in baseline means of students' observable characteristics (e.g., gender, *PAU* test score, household income, parental occupation) by period and treatment sample. The null hypothesis of equality of the observable characteristics across periods cannot be rejected for 80 and 60 percent of the variables in the

⁴⁸To see the Logit and DID analysis see Appendix, Section G.

extensive margin of grant eligibility and in additional amounts cutoffs respectively.⁴⁹ Applicants seem to be highly comparable across periods for those students applying at the vicinity of the grant extensive margin, which sub-population of applicants present slightly higher family income and fraction of movers, but lower students with blue collar parents, in comparison with those of the period with *weak* academic requirements. However, there is a more different student composition for those students entitled to additional amounts (at the lowest of the household income distribution), which makes the comparison across periods more challenging.

The equity results seem to contradict the evidence that performance incentives have negative impacts on student persistence presented by [Scott-Clayton and Schudde \(2019\)](#) in the US and [Lindo, Sanders and Oreopoulos \(2010\)](#) in Canada. However, results are consistent with [Montalvo \(2018\)](#), showing no adverse impact of a tuition increase on low-income students enrolled at universities in the region of Catalonia (Spain). The institutional context, especially the cost of college, may potentially be a relevant factor affecting the elasticity of dropout with respect to academic requirements. A setting with stronger academic requirements appears to be cost-effective, at least partially.

1.7.3 Potential confounding factors

Labeling Effect. Empirical evidence has shown that labeling may reinforce or mitigate the impact of an intervention. In the context of grants, [Barrera-Orsorio and Filmer \(2016\)](#) show that providing scholarship labeled as need-based or merit-based matter for its effects on student performance, since both types of grants increased enrollment and attendance to school, but only the merit-based displays positive impacts on student achievement. After nine years in the program, [Barrera-Orsorio, De Barros and Filmer \(2018\)](#) find that these differences remain, and those students on the merit-based program registered higher health and employment rates. In this paper, it is unlikely that the labeling effect may be a confounding factor of the results, since targeted population and label of the program remain constant over the period

⁴⁹See online appendix, section F.

of study.

Other institutional changes. In an attempt to rationalize public expenditure, the Spanish government passed Law 14/2012, which raised college tuition fees from 2012/2013 onwards.⁵⁰ The Law increased the cost per credit on the number of times registered in a particular module, such that the cost per credit that students bear raised exponentially with the number of re-registrations in a particular module.⁵¹ [Beneito, Bosca and Ferri \(2018\)](#) find that the increase in tuition fees reduces the average number of times students' register for a single module before passing it, increase the probability of passing with the first registration and raise their academic grades. A potential concern might arise on whether this change in the credit price across time might be a confounding factor of the impact of the grant. There are three reasons that potentially reject this idea. First, there are no statistically significant results of testing the isolated effect of being eligible for only fee waiver on student performance in none of the periods.⁵² Second, the significant effects on T1 eligibility threshold are mainly comparing students who got fee waiver vs. those who were granted with fee waiver plus some amount of cash allowance, holding constant the tuition fees paid by the students on both sides of the cutoffs within each period. This is due to the fact that students closer to the eligibility threshold are weighted higher using a triangular kernel. Third, this paper focuses on applicants for a need-based grant, a population of students that is potentially different to the standard student enrolled at university, which is the focus of the study previously mentioned.

In addition, a loan system was functioning in Spain from 2007 to 2011 for postgraduate studies. The timing – in the midst of a recession – was unpropitious and many students defaulted on their loan payments ([OECD, 2015](#)). The fraction of Master students covered by loans in 2010/2011 was only 2 percent. This program is unlikely to affect the results due to the tiny fraction of students covered, and by the fact that loans were offered for Master and PhD students, while this paper focuses

⁵⁰*Real Decreto-ley 14/2012, de 20 de abril, de medidas urgentes de racionalización del gasto público en el ámbito educativo.*

⁵¹Law 14/2012 established that university tuition fees should cover between 15–25%, 30–40%, 65–75%, and 90–100% of the total cost of education for the first, second, third and fourth-time and subsequent registrations respectively in a particular module.

⁵²See online appendix, section D.

on undergraduate students.

1.8 Conclusion

Using a regression discontinuity design in which I exploit the family income thresholds to being eligible for a grant, I estimate the causal effect of the Spanish large-scale need-based grant program on student performance (average GPA and fraction of credits earned) and degree completion. I find that an average provision of 825 euros cash allowance (relatively to receiving only fee waiver) increases student performance and probability of degree completion in a setting with demanding academic requirements. Students increase their final exams attendance rate and their GPA in final exams taken. Eligible students decrease the fraction of subjects that they have to retake and the total credits in which they enroll. There is no evidence of significant effects on student course selection or dropout from higher education. However, I find no effects of non-statistically different cash amounts interacted with academic requirements relatively comparable to those of the typical need-based grant programs around the world, such as the SAP of Pell Grants in the U.S.

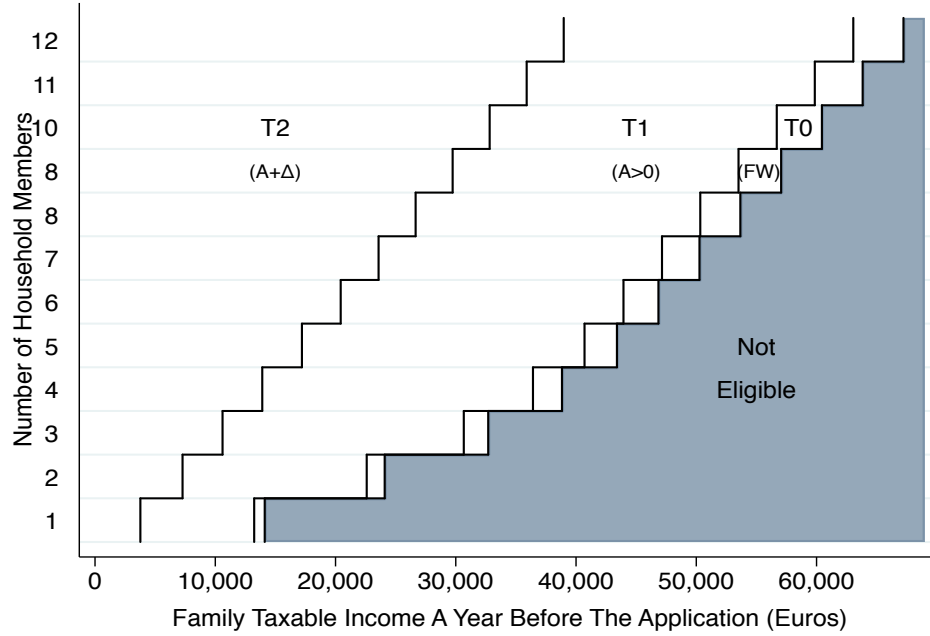
This paper points out the significant role that minimum academic requirements to retain the grant play on stimulating student performance and degree completion in the national large-scale need-based grant programs. Student performance seems to be weakly correlated with monetary incentives, but more reactive to academic incentives. It seems that grants attached to *weak* academic requirements do not affect student performance, while grants with *stronger* requirements provide large and positive effects. The intensity and design rather than the solely existence of academic requirements matter for stimulating student performance. The mechanisms indicate that a setting with *strong* academic requirements may reduce costs for taxpayers on repeated subjects failures and long attainment time rates, and improve the benefit of eligible students by improving their performance. Academic requirements in the context of higher education financial aid seem to be an effective tool to overcome moral hazard problems, though the optimal intensity of those requirements may be

institutional context-specific.

Although results may suggest that the effect on student performance is not linear to size of academic requirements, I cannot reject the possibility that the impact is in fact linear. It is still unclear which would be the distribution of the effects of a grant on performance and dropout along the different possible academic requirements (from zero strings attached to full accountability). Understanding how the impact of a grant changes when attached to all different possible academic requirements turns out to be a crucial research question, in order to improve the effectiveness of large-scale financial aid programs that cover a large fraction of students in higher education, and which spend sizable amounts of the public budget. Establishing the optimal line of academic standards and amount of financial aid that is socially optimal remains a topic for future research.

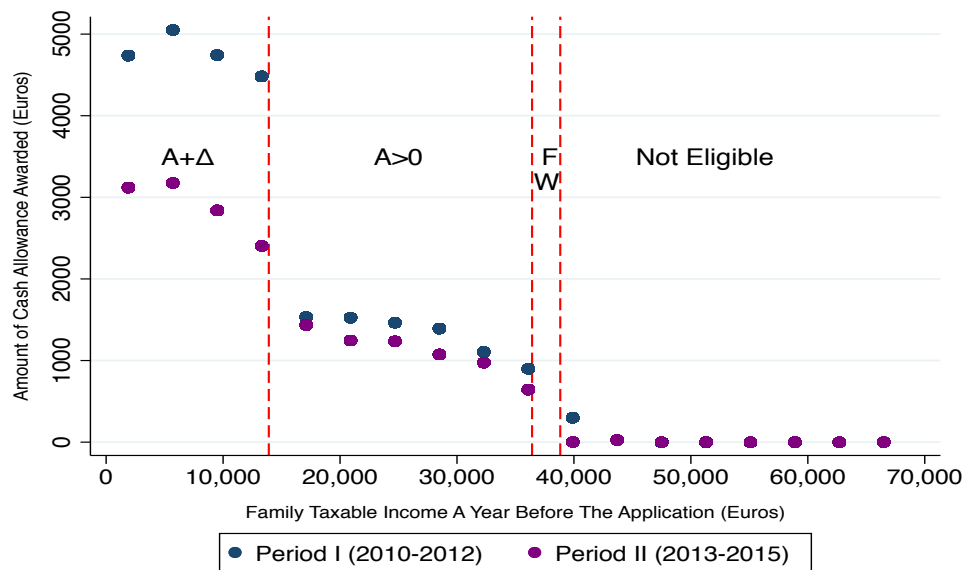
Appendix 1.A Appendix: Main Figures

Figure 1.A.1: Income eligibility thresholds for the different levels of the BCG grant.



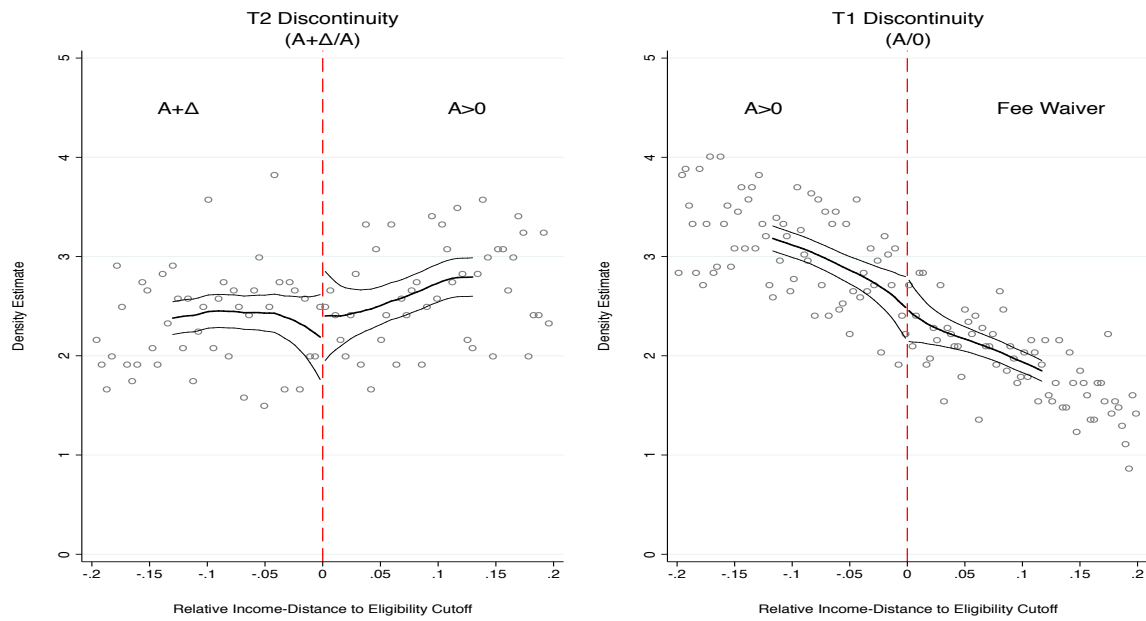
Note: The figure depicts family income thresholds for different number of household members in the period 2010–2015. Thresholds are exactly the same amounts over the six-year period. T0 refers to T0 Discontinuity in which students receive the fee waiver grant (FW), T1 to T1 Discontinuity in which students are awarded with fee waiver and cash amount (A), and T2 to T2 Discontinuity in which students are awarded with fee waiver and larger cash amount ($A + \Delta$) than (A). Thresholds are expressed in 2015 euros.

Figure 1.A.2: Amount of annual cash allowance awarded to applicants with 4 family members, as a function of their parents' taxable income by period



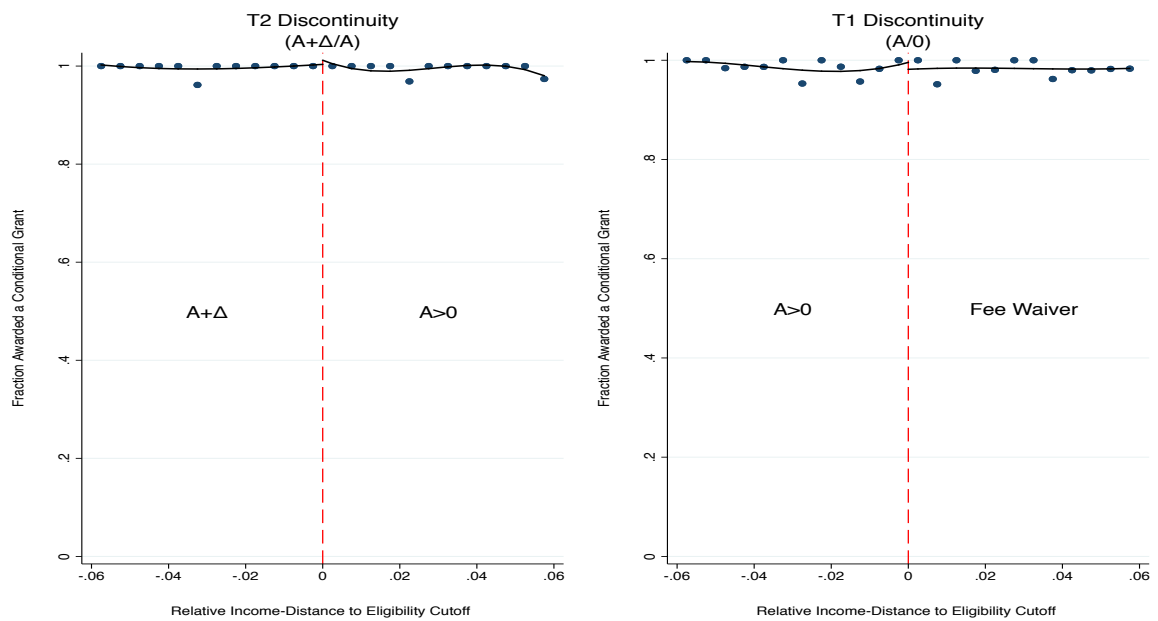
Note: Notes: The figure depicts the amount of annual cash allowance awarded to applicants with 4 family members, as a function of their parents' taxable income by period. FW refers to students receiving the fee waiver grant, A to students awarded with fee waiver and cash amount (A), and ($A + \Delta$) to students awarded with fee waiver and larger cash amount ($A + \Delta$) than (A).

Figure 1.A.3: McCrary (2008) test for 2010–2015



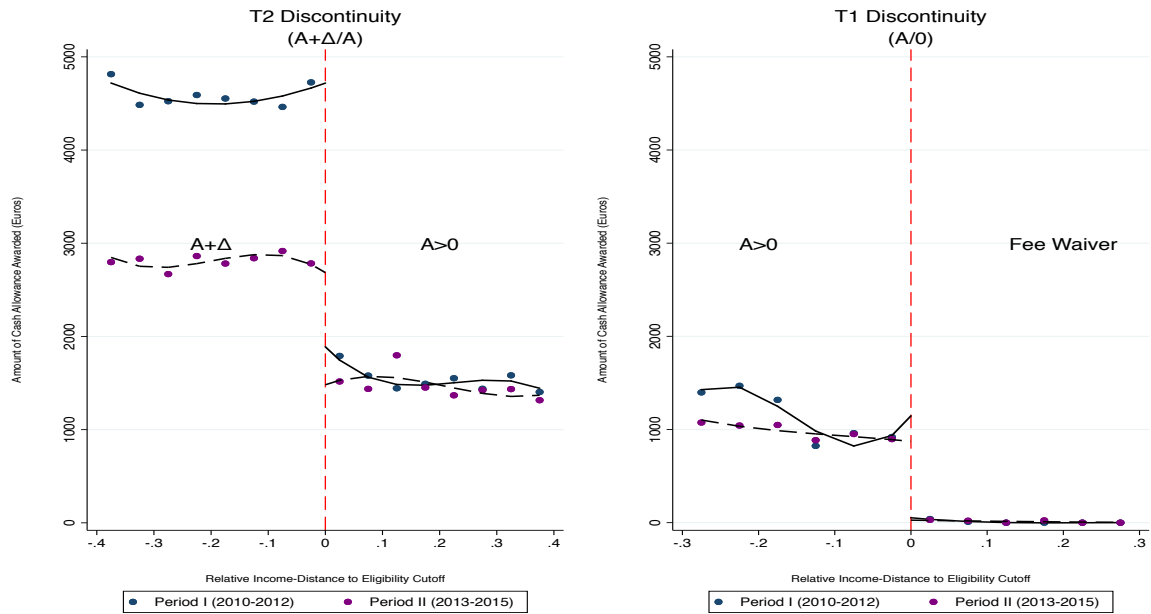
Note: The figure shows the results of the test proposed by [McCrary \(2008\)](#). The weighted kernel density estimates are plotted, computed separately for each of the sides of the income thresholds T1 and T2. Optimal bandwidth and bin size are computed by [McCrary \(2008\)](#) selection procedure. 'Relative Income-Distance to Eligibility Cutoff' refers to the relative distance of household taxable income to the income eligibility thresholds.

Figure 1.A.4: Fraction of Awarded a Conditional Grant for T1 and T2 Discontinuities (2010–2015).



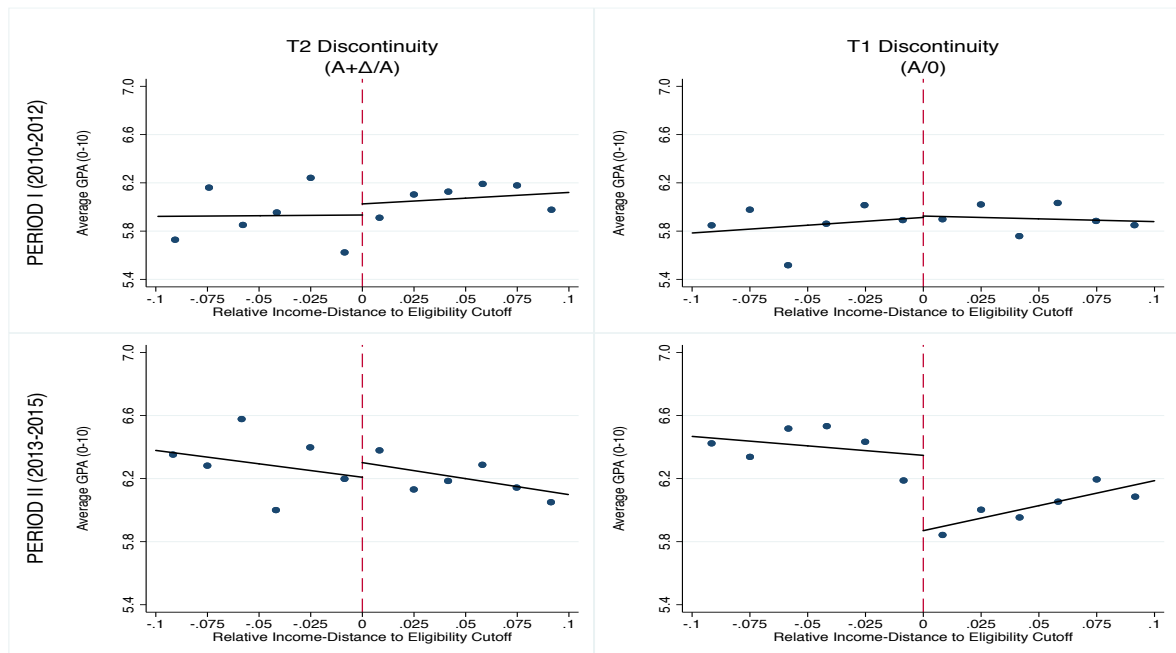
Note: The dots represent the average fraction of applicants who were awarded a conditional grant per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a third-order polynomial approximation which is estimated separately on both sides of the cutoffs. 'Relative Income-Distance to Eligibility Cutoff' refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Figure 1.A.5: Average Grant Amounts for T1 and T2 Discontinuities by Period.



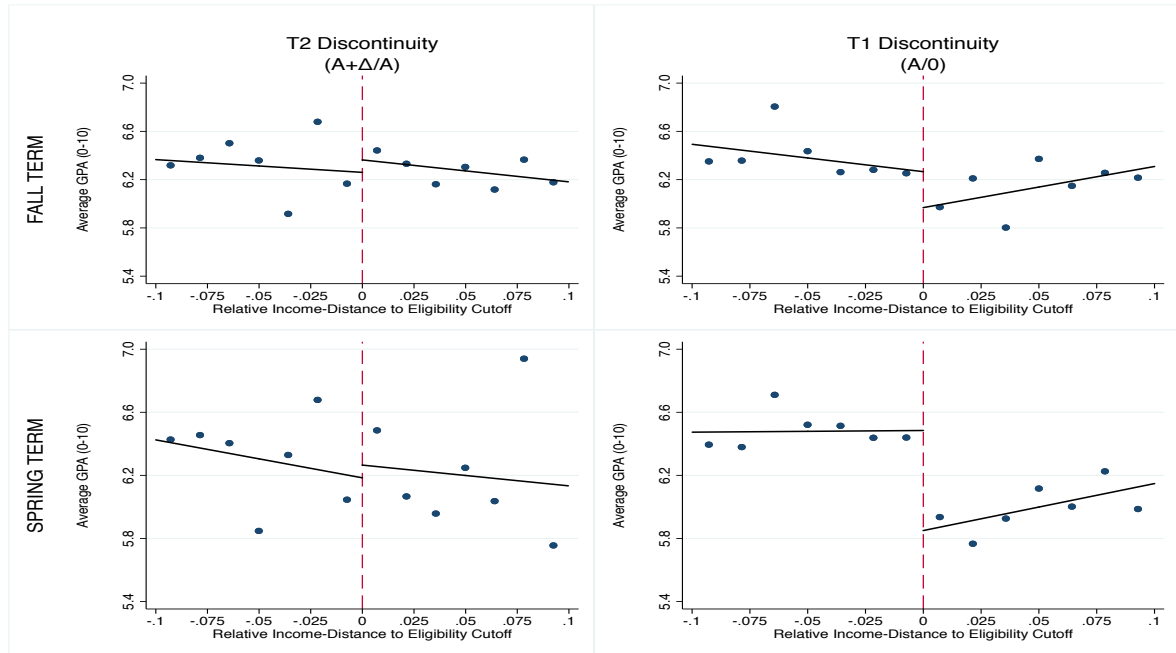
Note: The dots represent the average grant amount per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a third-order polynomial approximation which is estimated separately on both sides of the cutoffs. "Relative Income-Distance to Eligibility Cutoff" refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Figure 1.A.6: Average GPA (0-10) for T1 and T2 Discontinuities by Period.



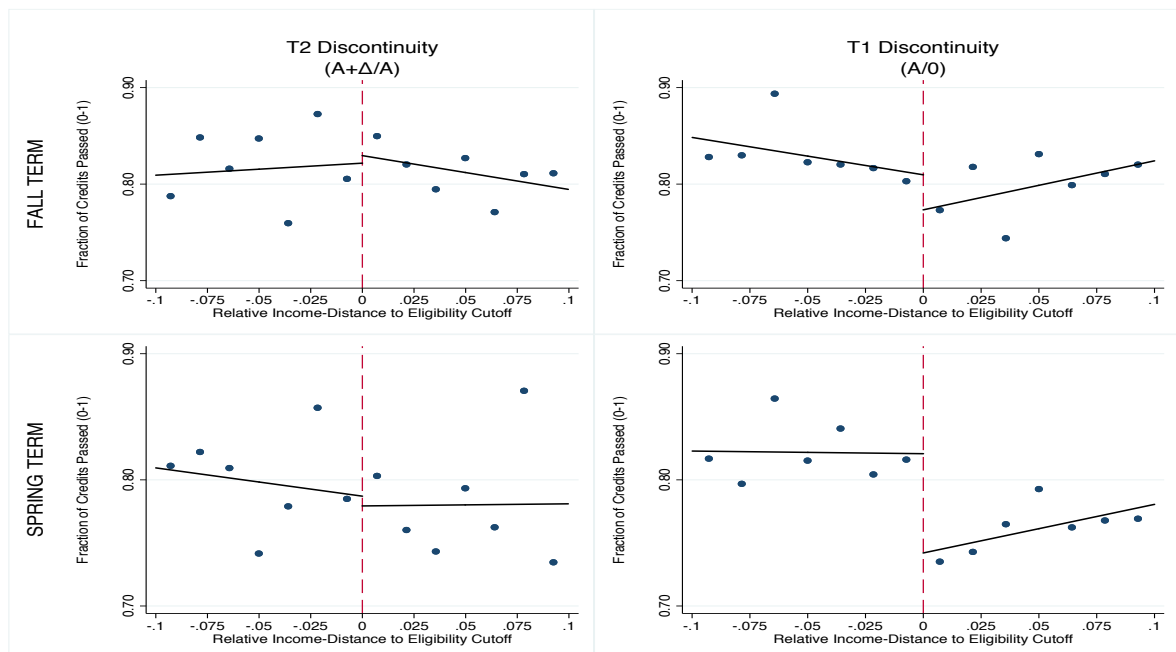
Note: The dots represent the average GPAs per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a third-order polynomial approximation which is estimated separately on both sides of the cutoffs. "Relative Income-Distance to Eligibility Cutoff" refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Figure 1.A.7: Average GPA (0-10) for T1 and T2 Discontinuities in Period II by Term



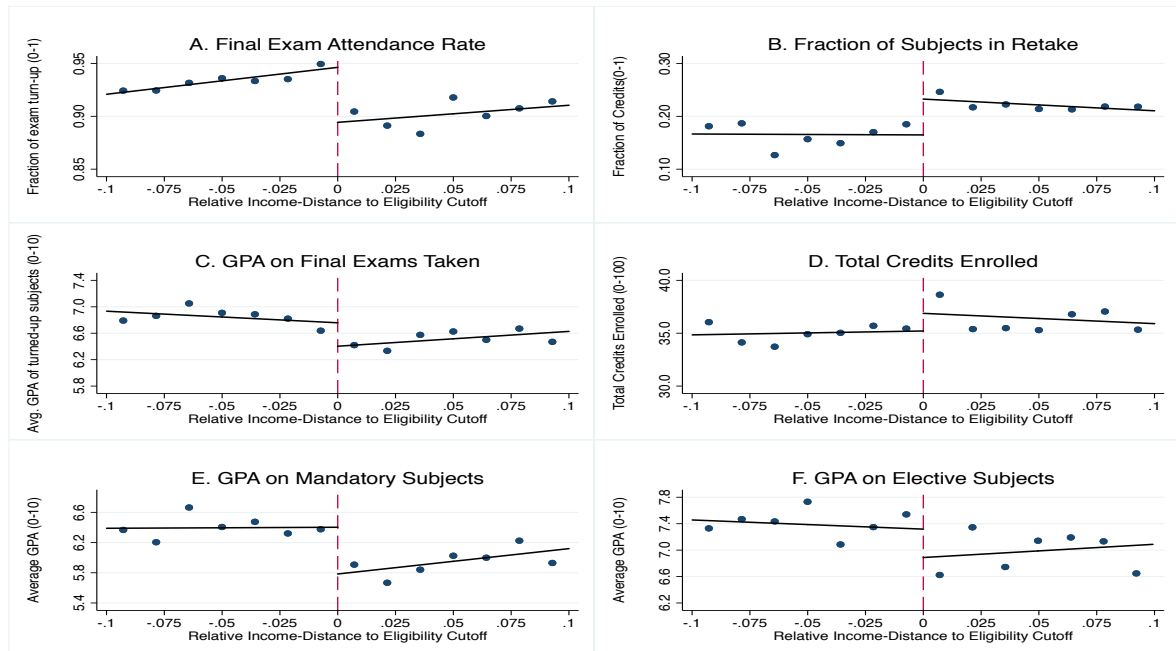
Note: The dots represent the Average GPAs per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a third-order polynomial approximation which is estimated separately on both sides of the cutoffs. "Relative Income-Distance to Eligibility Cutoff" refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Figure 1.A.8: Fraction of Credits Earned (0-1) for T1 and T2 Discontinuities over Period I and II by Term



Note: The dots represent the fraction of credits earned per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a third-order polynomial approximation which is estimated separately on both sides of the cutoffs. "Relative Income-Distance to Eligibility Cutoff" refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Figure 1.A.9: Mechanisms for Spring term at T1 Discontinuity (2013–2015)



Note: The dots represent the average of the outcome variable per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a third-order polynomial approximation which is estimated separately on both sides of the cutoffs. "Relative Income-Distance to Eligibility Cutoff" refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Appendix 1.B Appendix: Main Tables

Table 1.B.1: Number of BCG applicants (2010–2015).

		Undergraduate Old system	Undergraduate European system	Graduate Programs	Others	Total
2010	%	28.78	68.12	3.09	0	100
	N	1,555	3,680	167	0	5,402
2011	%	12.99	82.24	4.76	0	100
	N	701	4,436	257	0	5,497
2012	%	6.11	86.96	6.07	0.86	100
	N	334	4,754	332	47	5,552
2013	%	2.34	90.39	7.03	0.24	100
	N	119	4,602	358	12	5,174
2014	%	0.81	90.26	8.89	0.18	100
	N	41	4,560	449	9	5,128
2015	%	0.04	88.84	10.97	0.17	100
	N	2	4,721	582	9	5,314
Total	%	8.65	84.34	6.76	0.24	100
	N	2,745	26,755	2,145	77	31,722

Notes: Total number of BCG applicants to UC3M over the period studied 2010–2015. Undergraduate students studied are the addition of applicants in the old and new system. Undergraduate new system is typically four years degree program, harmonized with the European Union using ECTS credits.

Table 1.B.2: Descriptive Statistics on Undergraduate Applicants for Different Treatment Samples (2010–2015).

Treatment sample: (Income Eligibility Thresholds)	T2 Discontinuity ($A + \Delta/A$) (1)	T1 Discontinuity ($A/0$) (2)
Applicants		
Female	0.48	0.47
Spanish	0.92	0.98
Access to University Percentile rank	52.40 (28.60)	55.40 (28.46)
Technical degree	0.34	0.38
Applications		
Parent's taxable income (euros)	14,250 (5,689)	32,182 (10,111)
# Family members	3.7 (0.970)	3.6 (0.843)
% Disability	0.024	0.013
%Large family condition	0.17	0.11
Mover	0.32	0.28
Parental Occupation		
Entrepreneur	0.08	0.04
Blue Collar	0.44	0.3
Self-Employed	0.08	0.03
Conditional grant		
Awarded a conditional grant	0.99	0.72
Amount of Cash Allowance Awarded (Euros)	2,372 (1,858)	750.3 (1,105)
Undergraduate year		
First year	0.30	0.31
Second year	0.22	0.20
Third year	0.20	0.19
Fourth year	0.17	0.17
Years		
2010	0.167	0.164
2011	0.171	0.168
2012	0.149	0.174
2013	0.164	0.162
2014	0.171	0.165
2015	0.178	0.167
N	6,835	10,050

Notes: The sample is constructed by the administrative database of undergraduate applicants to the BCG grant in Carlos III University over 2010–2015. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student's postal code when they access higher education. The variable "Access to University Percentile Rank" is computed as the percentile rank of the students' academic year high school graduation on the *PAU* grade over the pool of BCG grant applicants from 2004-2015. Household's taxable income is expressed in constant 2015 euros. Standard errors are in parenthesis.

Table 1.B.3: Balance of Baseline Characteristics for Different Treatment Samples (2010–2015).

Treatment Sample: (Income Eligibility Thresholds)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Baseline mean (1)	Non-parametric Estimates (2)	Baseline mean (3)	Non-parametric Estimates (4)
A. Each baseline characteristic separately				
Female	0.48	0.015 (0.031)	0.46	0.033 (0.026)
Spanish	0.94	-0.014 (0.016)	0.99	0.003 (0.006)
Access to University Percentile rank	53.41	-2.044 (1.858)	56.36	3.46** (1.594)
STEM degree	0.35	0.004 (0.031)	0.41	-0.031 (0.025)
Households taxable income (euros)	17,282	-1.882 (233.841)	42,126	123.971 (308.907)
Disability	0.015	0.005 (0.009)	0.012	0.003 (0.006)
Large family condition	0.13	-0.000 (0.024)	0.13	0.007 (0.020)
#Family members	3.664	-0.004 (0.068)	3.651	-0.015 (0.047)
Live outside the family home	0.3	0.006 (0.028)	0.3	0.023 (0.023)
Entrepreneur Parent	0.073	0.007 (0.017)	0.04	-0.008 (0.009)
Blue Collar Parent	0.45	0.019 (0.032)	0.22	-0.041** (0.020)
Self-Employed Parent	0.061	0.026 (0.017)	0.021	-0.003 (0.008)
# Awarded Grants	0.061	-0.107 (0.097)	0.37	-0.020 (0.061)
First year	0.295	0.042 (0.031)	0.338	-0.012 (0.018)
Second year	0.225	-0.056* (0.029)	0.238	0.012 (0.017)
Third year	0.204	0.030 (0.023)	0.182	0.016 (0.016)
Fourth year	0.173	-0.019 (0.021)	0.148	0.005 (0.014)
B. All baseline characteristic jointly				
X2-stat		14.37		22.99
P-value		0.57		0.11

Notes: The table shows the RDD non-parametric estimates for the different applicants' observable variables. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average value of the observable variable above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. The variable "Access to University Percentile Rank" is computed as the percentile rank of the students' academic year high school graduation on the PAU grade over the poll of BCG grant applicants from 2004-2015. Household's taxable income is expressed in constant 2015 euros. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.4: Discontinuities in Allowance Amounts, GPA and Fraction of Credits Earned at Different Income Eligibility Thresholds and Periods.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. Average Allowance Amounts (euros)				
Non-parametric Estimates	2,955*** (147.657) [3,402]	1,240*** (108.270) [3,549]	675*** (98.806) [6,095]	825*** (37.662) [5,87]
Baseline mean	1,481	1,415	25.89	10.81
B. Average GPA (0-10)				
Non-parametric Estimates	-0.092 (0.190) [3,402]	0.057 (0.157) [3,549]	-0.031 (0.124) [6,093]	0.455*** (0.144) [5,868]
Baseline mean	5.97	6.29	5.91	6.15
C. Fraction of Credits Earned (0-1)				
Non-parametric Estimates	-0.017 (0.028) [3,402]	0.016 (0.026) [3,549]	0.006 (0.020) [6,093]	0.059*** (0.021) [5,868]
Baseline mean	0.78	0.80	0.77	0.79
D. Average Accumulated GPA over two years (0-10)				
Non-parametric Estimates	-0.017 (0.197) [2,801]	-0.246 (0.170) [1,851]	0.048 (0.124) [5,086]	0.511** (0.205) [3,178]
Baseline mean	6.00	6.29	5.96	6.14
E. Fraction of passed credits accumulated over two years (0-1)				
Non-parametric Estimates	-0.008 (0.024) [2,806]	0.021 (0.037) [1,860]	0.025 (0.019) [5,088]	0.063*** (0.023) [3,184]
Baseline mean	0.85	0.86	0.85	0.85

Notes: The table shows the RDD non-parametric estimates for the different applicants' average grant allowance received (Panel A), average GPA (Panel B), fraction of credits earned (Panel C) and average accumulated GPA over two years (Panel D). The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) for each separated regression. Baseline mean refers to the average grant amount above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.5: Discontinuities in Official Dropout from higher education at T1 and T2 grants by period.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. Dropout within one academic year				
Non-parametric Estimates	0.007 (0.010) [3,402]	0.017 (0.014) [3,549]	0.020 (0.015) [6,095]	-0.014 (0.014) [5,87]
Baseline mean	0.02	0.02	0.03	0.03
B. Dropout within two academic year				
Non-parametric Estimates	0.007 (0.015) [3,402]	0.027 (0.018) [3,549]	0.026 (0.019) [6,095]	-0.016 (0.019) [5,87]
Baseline mean	0.04	0.04	0.04	0.06

Notes: The table shows the RDD non-parametric estimates for the different applicants' dropout from higher education. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average dropout above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported below the standard errors. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.6: Discontinuities in Average GPA and Fraction of Credits Earned at Different Income Eligibility Thresholds and Periods by Term.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I (2010-2012)	Period II (2013-2015)	Period I (2010-2012)	Period II (2013-2015)
	(1)	(2)	(3)	(4)
A. Average GPA (0-1)				
A.1. Baseline Estimates				
Non-parametric Estimates	-0.092 (0.190) [3,402]	0.057 (0.157) [3,549]	-0.031 (0.124) [6,093]	0.455*** (0.144) [5,868]
Baseline mean	5.97	6.29	5.91	6.15
A.2. First Term (Fall)				
Non-parametric Estimates	-0.024 (0.213) [3,299]	0.020 (0.159) [3,470]	0.030 (0.127) [5,904]	0.282** (0.136) [5,703]
Baseline mean	6.01	6.33	5.91	6.20
A.3. Second Term (Spring)				
Non-parametric Estimates	-0.187 (0.222) [3,282]	0.050 (0.190) [3,400]	-0.118 (0.138) [5,885]	0.560*** (0.154) [5,600]
Baseline mean	6.02	6.33	5.99	6.19
B. Fraction of Credits Earned (0-1)				
B.1. Baseline Estimates				
Non-parametric Estimates	-0.017 (0.028) [3,402]	0.016 (0.026) [3,549]	0.006 (0.020) [6,093]	0.059*** (0.021) [5,868]
Baseline mean	0.78	0.80	0.77	0.79
B.2. First Term (Fall)				
Non-parametric Estimates	-0.010 (0.029) [3,299]	-0.003 (0.023) [3,470]	0.018 (0.023) [5,904]	0.037** (0.018) [5,703]
Baseline mean	0.79	0.82	0.77	0.80
B.3. Second Term (Spring)				
Non-parametric Estimates	-0.019 (0.031) [3,282]	0.029 (0.034) [3,400]	-0.012 (0.020) [5,885]	0.073*** (0.022) [5,600]
Baseline mean	0.78	0.80	0.78	0.79

Notes: The table shows the RDD non-parametric estimates for the different applicants' average GPA (Panel A) and fraction of credits earned (Panel B). The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) for each separated regression. Baseline mean refers to the average grant amount above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.7: Discontinuities in Average GPA (0-10) at T1 and T2 grants by period and subgroup of applicants.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)		T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
		Period I	Period II	Period I	Period II
		(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. By Gender					
A. By Gender Female	Non-parametric Estimates	-0.038 (0.264) [1,644]	0.078 (0.184) [1,689]	0.063 (0.209) [2,879]	0.250 (0.166) [2,735]
	Baseline mean	6.32	6.69	6.27	6.56
Male	Non-parametric Estimates	-0.167 (0.262) [1,758]	-0.021 (0.231) [1,859]	-0.110 (0.202) [3,209]	0.465** (0.189) [3,132]
	Baseline mean	5.64	5.94	5.58	5.83
B. PAU entrance exam percentile rank					
Above Median	Non-parametric Estimates	-0.403 (0.249) [1,838]	0.000 (0.220) [1,772]	-0.179 (0.175) [3,433]	0.523*** (0.180) [3,254]
	Baseline mean	6.41	6.75	6.32	6.61
Below Median	Non-parametric Estimates	0.344 (0.335) [1,484]	0.129 (0.213) [1,722]	0.104 (0.217) [2,504]	0.295 (0.219) [2,516]
	Baseline mean	5.44	5.76	5.35	5.53
C. By residence status					
Living with parents	Non-parametric Estimates	0.044 (0.273) [2,346]	-0.018 (0.195) [2,388]	-0.083 (0.142) [4,417]	0.450*** (0.158) [4,11]
	Baseline mean	5.90	6.19	5.81	6.03
Living outside the family home	Non-parametric Estimates	-0.385 (0.258) [1,056]	0.160 (0.257) [1,16]	-0.045 (0.272) [1,671]	0.524* (0.299) [1,757]
	Baseline mean	6.15	6.52	6.15	6.45

Notes: The table shows the RDD non-parametric estimates for the different applicants' variables. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average variable value above the eligibility threshold. The T1 treatment sample includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.8: Discontinuities in the Fraction of Credits Earned (0-1) at T1 and T2 grants by period and subgroup of applicants.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)		T2 Discontinuity (A + Δ /A)		T1 Discontinuity (A/0)	
		Period I	Period II	Period I	Period II
		(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. By Gender					
Female	Non-parametric Estimates	0.016 (0.038) [1,644]	0.008 (0.025) [1,689]	0.019 (0.026) [2,879]	0.039 (0.026) [2,735]
	Baseline mean	0.830	0.856	0.818	0.841
Male	Non-parametric Estimates	-0.063 (0.044) [1,758]	-0.006 (0.033) [1,859]	-0.004 (0.033) [3,209]	0.054** (0.027) [3,132]
	Baseline mean	0.730	0.760	0.726	0.744
B. PAU entrance exam percentile rank					
Above Median	Non-parametric Estimates	-0.058* (0.033) [1,838]	-0.019 (0.022) [1,772]	-0.004 (0.019) [3,433]	0.068*** (0.025) [3,254]
	Baseline mean	0.825	0.862	0.817	0.845
Below Median	Non-parametric Estimates	0.034 (0.051) [1,484]	0.029 (0.036) [1,722]	0.010 (0.035) [2,504]	0.032 (0.031) [2,516]
	Baseline mean	0.723	0.739	0.703	0.706
C. By residence status					
Living with parents	Non-parametric Estimates	0.004 (0.036) [2,346]	-0.012 (0.029) [2,388]	0.007 (0.025) [4,417]	0.058** (0.023) [4,110]
	Baseline mean	0.771	0.793	0.757	0.773
Living outside the family home	Non-parametric Estimates	-0.063 (0.045) [1,056]	0.044 (0.037) [1,160]	-0.007 (0.034) [1,671]	0.056 (0.039) [1,757]
	Baseline mean	0.799	0.830	0.798	0.822

Notes: The table shows the RDD non-parametric estimates for the different applicants' variables. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average variable value above the eligibility threshold. The T1 treatment sample includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.9: Discontinuities for the mechanisms variables at T1 and T2 grants by period.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012)	(2013-2015)	(2010-2012)	(2013-2015)
	(1)	(2)	(3)	(4)
A. Final exam attendance rate (0-1)				
Non-parametric Estimates	-0.001 (0.016) [3,402]	0.006 (0.012) [3,549]	0.006 (0.012) [6,093]	0.032*** (0.010) [5,868]
Baseline mean	0.904	0.932	0.912	0.929
B. Fraction of Subjects in Retake				
Non-parametric Estimates	0.004 (0.025) [3,402]	-0.002 (0.020) [3,549]	-0.013 (0.019) [6,093]	-0.046** (0.018) [5,867]
Baseline mean	0.192	0.169	0.204	0.189
C. GPA on final exams taken				
Non-parametric Estimates	-0.040 (0.156) [3,392]	0.074 (0.139) [3,541]	-0.049 (0.106) [6,077]	0.351*** (0.129) [5,859]
Baseline mean	6.512	6.699	6.396	6.552
D. GPA on Mandatory Subjects				
Non-parametric Estimates	-0.084 (0.200) [3,397]	0.107 (0.172) [3,548]	-0.050 (0.128) [6,089]	0.462*** (0.152) [5,865]
Baseline mean	5.908	6.210	5.868	6.103
E. GPA on Elective Subjects				
Non-parametric Estimates	-0.689** (0.326) [1,298]	-0.056 (0.288) [1,288]	0.136 (0.205) [2,134]	0.553 (0.385) [1,924]
Baseline mean	6.954	7.370	6.802	7.236

Notes: The table shows the RDD non-parametric estimates for the different applicants' outcome variables. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average value above the eligibility threshold. The T1 treatment sample includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.B.10: Discontinuities in Degree Completion at T1 and T2 grants by period.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012)	(2013-2015)	(2010-2012)	(2013-2015)
	(1)	(2)	(3)	(4)
A. Probability of Graduation				
Non-parametric Estimates	0.007 (0.020) [2,253]	0.116 (0.073) [1,151]	0.008 (0.023) [4,842]	0.100* (0.056) [1,947]
Baseline mean	0.93	0.92	0.91	0.8
B. Probability of Graduation in Graduation Year				
Non-parametric Estimates	-0.008 (0.022) [515]	0.000 (0.000) [428]	-0.022 (0.021) [863]	0.102* (0.056) [723]
Baseline mean	0.98	0.99	0.99	0.95

Notes: The table shows the RDD non-parametric estimates for the different applicants' degree completion. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average GPA value above the eligibility threshold. The T1 treatment sample includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 1.C Additional Figures and Tables

The section of the appendix supplements the paper “The role of performance incentives in need-based grants for higher education: Evidence from the Spanish *Becas*”. It presents details on low-income students’ performance in higher education (section 1), validity of the research design (section 2), the discontinuities in BCG grants awarded to applicants (section 3), the robustness of baseline estimates (section 4), an RDD-DID reduced-form model (section 5), the comparability between Period I and Period II (section 6), the potential equity effects (section 7), and the minimum academic requirements to being eligible for a BCG grant (section 8).

1.C.1 Low-Income Students’ Performance in Higher Education

The analysis of the BCG grant is performed on low-income high school graduates enrolled in Carlos III University who applied to a BCG grant to start or to continue undergraduate college studies. To compare Carlos III students with the rest of collegian enrolled in the Spanish public universities, and more precisely, with the population of low-income students enrolled in higher education, I use data on student attainment in higher education provided by the Ministry of Education for the academic year 2014/2015. Dropout rates were computed for the cohort of students who enrolled Spanish higher education in the academic year 2010/2011 (which expected graduation from undergraduate program was 2013/2014 or 2014/2015).

The summary statistics presented in Table [1.C.1](#) show substantial differences between Carlos III undergraduate students and their peers enrolled in the rest of the Spanish public higher education institutions. Students enrolled in Carlos III University scored higher in the standardized university access exam, reported higher graduation rates, passed a higher number of credits, and presented lower dropout rates. In contrast, BCG grant recipients in Carlos III are highly comparable to BCG grant recipients in Spain. These students reported similar GPAs, number of credits passed (over the total credits enrolled and the final exams taken), and time to graduation. The sample of grant applicants is reasonably representative of the general

population of low-income students in Spain. These students can be considered as comparable to the typical targeted population of most large-scale need-based grant programs around the world, e.g students graduated from high school and admitted to college. The results cannot be directly generalized to the population of low-income students who fail in high school graduation and might respond differently to financial aid.

Table 1.C.1: External Validity.

	All Students			BCG Grant Recipients			Non BCG Grant Recipients		
	Spain (1)	Carlos III (2)	Diff. (2)-(1)	Spain (3)	Carlos III (4)	Diff. (4)-(3)	Spain (5)	Carlos III (6)	Diff. (6)-(5)
Avg. PAU score	8,5	10,29	1,79	8,67	10,28	1,61	8,76	10,51	1,75
Avg. GPA of enrolled students	7,24	6,95	-0,29	7,3	7	-0,3	7,2	6,9	-0,3
Number of credits passed over total enrolled	92,8	90,2	-2,6	88	88,6	0,6	74,9	82,3	7,4
Number of credits passed over total final exam taken	94,9	93,3	-1,6	92,1	93	0,9	84,9	89,7	4,8
Avg. time to graduation (4-year program)	4,4	4,8	0,4	4,4	4,8	0,4	4,5	4,8	0,3
Graduation rate (graduates in 2014/ total enrolled)	15,9	20,3	4,4						
Dropout rate (cohort 2010/2011)	28,42	24,7	-3,72	25,6	28,8	3,2	29,1	23,7	-5,4
# Enrolled	1,187,976	15,394		326,693	2,879		861,283	12,515	

Notes: Self-constructed Table with data from the Spanish Ministry of Education.

1.C.2 Validity of the Research Design: McCrary (2008) Test

In order to perform a formal investigation of the validity of the research design, I test the non-random sorting of applicants at the income eligibility thresholds. I use the test proposed in McCrary (2008), which is a test based on an estimator for the discontinuity in the density function of the running variable at the cutoff, checking the non-systematic manipulation of household parental income around the thresholds.

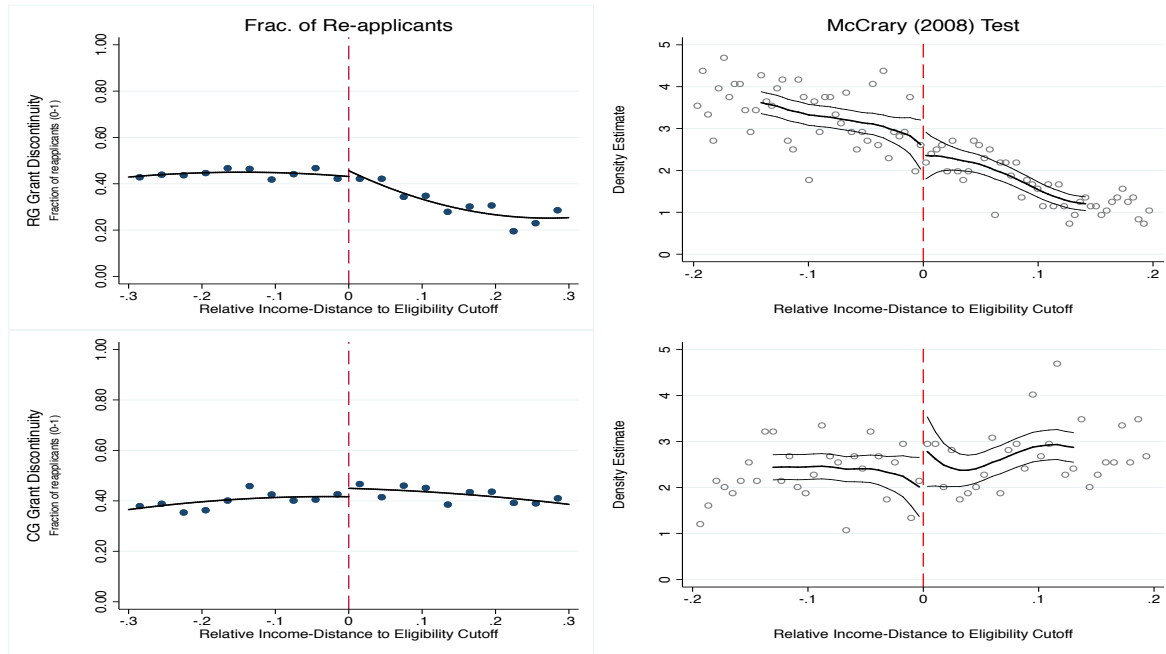
The results of the McCrary (2008) test for Fee Waiver (FW) and Displacement and Other Needs Grant (T3) are presented in Table 1.C.2. McCrary test statistics confirm that the null hypothesis of no density jump at the eligibility cutoffs cannot be rejected for these income eligibility cutoffs. Table 1.C.3 shows the results of McCrary (2008) test developed for all the different treatment samples used in this paper. Regardless of the treatment sample considered, the test statistic fails to reject the null hypothesis that the log difference in height around the discontinuity points is equal to zero. In addition, Figure 1.C.1 displays the fraction of re-applicants and McCrary (2008) test for this sub-population of students. Applicants' who were awarded a grant in a given year might be more likely to reapply the next year, especially those below the cutoff of the T1 Discontinuity. If it is the case, it may suggest that the impacts would be driven by this group of students, with no density break for applicants at the cutoffs but so for re-applicants. A robustness check testing the discontinuity in the density of re-applicants rejects this concern. Overall, these tests suggest that the probability of submitting an application does not change discontinuously at the income eligibility threshold, and thus applicants immediately above the cutoff are not able to manipulate their household parental income to being eligible for higher levels of grant.

Table 1.C.2: McCrary (2008) Test for Manipulation of the Forcing Variable for Different Treatment Samples in FW and T3 grant.

Treatment sample (Income Eligibility Thresholds)		Fee Waiver Grant (FW) (Threshold 0)			
		Log Difference in frequency bins (1)	Z-stat (2)	Bandwidth (3)	Bin size (4)
A. Total sample					
	Period I (2010–2012)	.175 (.193)	.907	.047	.004
	Period II (2013–2015)	.228 (.15)	1.47	.061	.004
Treatment sample (Income Eligibility Thresholds)		Displacement and Other Needs Grant (T3) (Threshold 3)			
		Log Difference in frequency bins (5)	Z-stat (6)	Bandwidth (7)	Bin size (8)
A. Total sample					
	Period I (2010–2012)	-.02 (.12)	.23	.09	.004
	Period II (2013–2015)	.09 (.15)	.62	.06	.004
B. By residence condition					
Living with parents	Period I (2010–2012)	-.12 (.17)	.69	.06	.005
	Period II (2013–2015)	.26 (.17)	1.49	.06	.005
Living outside the family home	Period I (2010–2012)	-.07 (.29)	.23	.05	.008
	Period II (2013–2015)	-.42 (.29)	1.43	.06	.007

Notes: The McCrary test is performed separately for each treatment sample. The FW treatment sample includes applicants whose household parental taxable income is within 30 percent of the eligibility thresholds. The T3 treatment sample includes applicants whose household parental taxable income is within 30 percent of the eligibility thresholds. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student's postal code when they access higher education. Standard deviations are in parenthesis. $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1.C.1: Fraction of re-applicants and McCrary (2008) test for re-applicants density.



Notes: The dots represent the average fraction of re-applicants and density estimates of [McCrary \(2008\)](#) test per interval of relative income-distance to the eligibility thresholds. The solid lines are fitted values from a second-order polynomial approximation which is estimated separately on both sides of the cutoffs. "Relative Income-Distance to Eligibility Cutoff" refers to the relative distance of household taxable income to the income eligibility thresholds. Red vertical lines identify the income eligibility thresholds.

Table 1.C.3: McCrary (2008) Test for Manipulation of the Forcing Variable for Different Treatment subamples.

Treatment sample (Income Eligibility Thresholds)		T2 Discontinuity (A + Δ /A)				T1 Discontinuity (A/0)			
		Log Difference in frequency bins	Z-stat	Bandwidth	Bin size	Log Difference in frequency bins	Z-stat	Bandwidth	Bin size
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. By Period									
Total Applicants	Period I (2010-2012)	.13 (.22)	.60	.06	.006	-.079 (.12)	.647	.10	.004
Total Applicants	Period II (2013-2015)	.18 (.23)	.79	.07	.006	-.107 (.13)	.80	.10	.004
B. By Gender									
Females	Period I (2010-2012)	-.14 (.43)	.33	.05	.009	-.06 (.17)	.34	.096	.006
	Period II (2013-2015)	.29 (.36)	.81	.05	.009	.004 (.23)	.02	.072	.006
Males	Period I (2010-2012)	.16 (.25)	.61	.07	.009	-.11 (.2)	.55	.074	.006
	Period II (2013-2015)	.18 (.36)	.49	.06	.008	-.03 (.18)	.16	.10	.006
C. By PAU Percentile Rank									
Above Median	Period I (2010-2012)	-.05 (.27)	.19	.07	.008	-.14 (.19)	.75	.07	.005
	Period II (2013-2015)	.05 (.28)	.17	.07	.008	-.096 (.19)	.48	.08	.005
Below Median	Period I (2010-2012)	.28 (.34)	.83	.07	.01	-.016 (.24)	.07	.06	.007
	Period II (2013-2015)	.20 (.43)	.47	.06	.009	-.016 (.22)	.07	.08	.007
D. By residence condition									
Living with parents	Period I (2010-2012)	.12 (.26)	.45	.07	.007	.075 (.14)	.52	.09	.005
	Period II (2013-2015)	-.11 (.29)	.39	.07	.007	.054 (.16)	.32	.09	.005
Living outside the family home	Period I (2010-2012)	.14 (.38)	.38	.07	.012	-.65 (.37)	1.76	.059	.008
	Period II (2013-2015)	.62 (.33)	1.87	.09	.01	-.46 (.29)	1.56	.078	.008

Notes: The McCrary test is performed separately for each treatment sample. The T1 treatment sample includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. The variable "Access to University Percentile Rank" is computed as the percentile rank of the students' academic year high school graduation on the PAU grade over the pool of BCG grant applicants from 2004-2015.. Standard deviations are in parenthesis: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.C.3 Discontinuities in Awarded Grants

This section tests the different discontinuities of average awarded grants at the different income eligibility thresholds. Table 1.C.4 presents the average allowance amounts (in constant euros of 2015) at T1 and T2 grants for the two periods studied and all the treatment samples used in the paper. This table shows that all the treatment groups present strong and statistically significant increments in average cash amount awarded at the discontinuity thresholds, except for students living with their parents when they enter university (non-movers) for T1 grant at Period I. These subgroups of students were not eligible for financial aid at this specific threshold and period. Table 1.C.5 shows the average cash amount at T1 and T2 grants for being eligible for a grant over two academic years. Conditional on applying for a grant at time t with a certain household income, it is possible to compute the average allowance amounts awarded over two years. This method would provide no sample selection concerns. However, the first stage decreases over time due to the variability of applicants' application status and household income over years. The discontinuities in the actual amount of conditional grant awarded to applicants are about 300 euros for T1 grant on both periods, and similar estimates for T2 grant but not statistically significant.

Table 1.C.4: Average Allowance Amounts (in euros) at T1 and T2 Discontinuities by period and subgroup sample.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)		T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
		Period I	Period II	Period I	Period II
		(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. Total Applicants					
	Non-parametric Estimates	2,955*** (147.657) [3,402]	1,240*** (108.270) [3,549]	675*** (98.806) [6,095]	825*** (37.662) [5,87]
	Baseline mean	1,481	1,415	25.89	10.81
B. By Gender					
Female	Non-parametric Estimates	3,231*** (222.727) [1,644]	1,339*** (158.675) [1,689]	878*** (140.786) [2,881]	946*** (57.538) [2,735]
	Baseline mean	1,604	1,561	21.59	20.70
Male	Non-parametric Estimates	2,715*** (213.360) [1,758]	1,103*** (178.747) [1,859]	472*** (131.423) [3,209]	696*** (54.368) [3,134]
	Baseline mean	1,364	1,287	29.75	2.88
C. PAU entrance exam percentile rank					
Above Median	Non-parametric Estimates	2,773*** (254.429) [1,838]	1,372*** (160.635) [1,772]	895*** (152.250) [3,435]	1,029*** (73.145) [3,256]
	Baseline mean	1,728	1,741	33.19	12.90
Below Median	Non-parametric Estimates	2,985*** (205.623) [1,484]	1,124*** (168.346) [1,722]	454*** (125.301) [2,504]	607*** (71.740) [2,516]
	Baseline mean	1,149	1,045	16.69	8.348
D. By residence status					
Living with parents	Non-parametric Estimates	2,984*** (141.017) [2,346]	1,235*** (115.897) [2,388]	-13.324 (62.261) [4,419]	445*** (19.519) [4,110]
	Baseline mean	818.2	1,019	16.70	2.652
Living outside the family home	Non-parametric Estimates	2,599*** (227.727) [1,056]	1,320*** (257.033) [1,160]	2,858*** (187.492) [1,671]	1,673*** (87.784) [1,759]
	Baseline mean	3,123	2,281	47.47	30.42

Notes: The table shows the RDD non-parametric estimates for the average allowance amount for different samples. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) for each separated regression. Baseline mean refers to the average variable value above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported below the standard errors. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.5: Discontinuities in Average Allowance Amounts at t+1 for T1 and T2 grants by period.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012)	(2013-2015)	(2010-2012)	(2013-2015)
	(1)	(2)	(3)	(4)
Non-parametric Estimates	394 (310.184) [1,039]	405** (181.193) [1,08]	377** (150.853) [1,861]	406*** (67.479) [1,87]
Baseline mean	1,926	1,784	231.6	156.2

Notes: The table shows the RDD non-parametric estimates for the different applicants' average grant allowance received. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) for each separated regression. Baseline mean refers to the average grant amount above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported below the standard errors. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.C.4 Robustness Checks

In this section, I perform a number of tests in order to check the robustness of baseline estimates. Specifically, I i) investigate the sensitivity of estimates to the choice of bandwidth; ii) perform the local polynomial regression with robust bias-corrected confidence intervals proposed by Calonico, Cattaneo and Titiunik (2014); iii) run the baseline regressions adding student individual predetermined variables and year fixed effects; iv) test for jumps at non-discontinuity points by running placebo regressions.

Sensitivity to the Choice of Bandwidth. I analyze the sensitivity of the non-parametric estimates to the choice of the bandwidth and that changing the bandwidth size to half or twice the value of the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). Panel B in Table 1.C.6 shows that results are very similar to those obtained in the baseline estimates, but larger when using half the optimal bandwidth than double.

Local Polynomial Regression with Robust Bias-Corrected Confidence Intervals. To test for the variability of the results under local polynomial regression

and a different computation for confidence intervals (robust bias-corrected) proposed by [Calonico, Cattaneo and Titiunik \(2014\)](#). Panel C in Table 1.C.6 presents non-parametric estimates very close to the baseline.

Individual Control Variables and Year Fixed Effects. I investigate the volatility of baseline results when adding individual predetermined control variables (such as PAU percentile rank, gender, or being enrolled in a STEM degree) and year fixed effects that capture time trends in the outcome variable to the main regression. Panel D in Table 1.C.6 shows that results are statistically significant at the 5 percent confidence level, but magnitudes is smaller than baseline estimates and similar to changing the bandwidth size to half of the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#).

Testing for Jumps at Non-Discontinuity Points. To test for jumps at non-discontinuity points, I run a placebo regression in which the income thresholds are artificially set at the midpoint between the actual eligibility thresholds by period. Since these midpoint do not correspond to any change in applicants' grant eligibility status, I should expect to find no significant jumps in average GPA. Panel E in Table 1.C.6 presents that the points estimates are close to zero and non-significant in all specifications.

Overall, baseline results are robust to all different specifications and vary from an effect of 0.27 to 0.5 points, which corresponds to about 4.5 to 8.3 percent with respect to the baseline mean. Although the magnitude of estimates varies across specifications due to the limited sample size, the direction of the effects hold over the different specifications, indicating a robust impact of grant eligibility on student performance when the academic standards are strong. In addition, the null effect of the grant under the other different thresholds (T2 and fee waiver grant) and periods is also robust and persistent for every sensitivity check performed.

Fee Waiver Grant (FW, Threshold 0). The fee waiver is the first type of grant that students may receive, and covers the tuition fees but does not award with amounts of cash. This eligibility threshold (FW or Threshold 0) is very close to the eligibility cutoff the the T1 grant. It makes difficult to construct two treatment

samples between T1 grant and fee waiver which do not overlap. The discontinuity induced by the tuition fee eligibility cutoff was ignored in the analysis, in order to focus on the grants types where students were awarded with cash amounts (in T1 and T2 grant).

As a robustness check, I have conducted a separate analysis of the treatment effect of being eligible for only tuition fee. Table 1.C.7 reports the discontinuities in average cash amount awarded and average GPA. This table shows no evidence of statistically significant effects on awarded cash amounts and average GPA at this threshold.

Displacement and Other Needs Grant (T3, Threshold 3). An additional robustness check is testing whether there is some statistically significant effect at T3 threshold (working only in Period I) where the amount of cash awarded was similar to the one awarded for non-movers in Period II. As a robustness check, I use the T3 grant for non-movers as comparison group for T1 non-movers at Period I, since both thresholds are very close to each other. This analysis is useful to investigate the role of grant's performance-based incentive components.

An ideal test would be a similar amount of cash awarded for non-movers at Period I and II for T1 grant, but unfortunately this is not the case. As a robustness check, I use the T3 grant for non-movers as comparison group for T1 non-movers at Period I, since both thresholds are very close to each other.

This analysis is useful to investigate the role of grant's performance-based incentive components. The key advantages of using the T3 cutoff are twofold. First, the T3 grant was located 15 percent of the relative distance below the T1 threshold, which mitigates concerns regarding the comparability of students in the vicinity of these two cutoffs. The sample of non-movers received their first cash award at T3 in Period I, which makes the comparable group similar to non-movers at T1 in Period II. Second, the discontinuities in average cash grant amounts were very similar (543 vs. 410 euros). Hence, using non-movers in Period I for T3 grant as a comparison group for T1 non-movers in Period II is convenient due to the fact that it offers an scenario where entitlement to the grant, cash allowances and sample are comparable,

but performance-incentives are different in the two periods.

The key advantages of using the T3 cutoff are twofold. First, the T3 grant is located 15 percent of the relative distance below the T1 threshold, which mitigates concerns regarding the comparability of students in the vicinity of these two cutoffs. The sample of non-movers receives their first cash award at T3 in Period I, which makes the comparable group similar to non-movers at T1 in Period II. Second, the discontinuities in average cash grant amounts are very similar (543 vs. 410 euros). Hence, using non-movers in Period I for T3 grant as a robustness check for T1 non-movers in Period II is convenient due to the fact that it offers an scenario where entitlement to the grant, cash allowances and sample are comparable, but performance-incentives are different in the two periods.

The average increase in cash allowance at T3 in Period I was 543 euros for non-movers, and cash endowments at T1 in Period II was 410 euros for non-movers. However, the null hypothesis of zero effect of being eligible for the T3 grant on non-movers student performance cannot be rejected. Results are robust to different treatment sample sizes, regarding the predetermined characteristics of applicants, year fixed effects, to set the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) as half and twice of its value, and to perform the local polynomial regression with robust bias-corrected confidence intervals proposed by [Calonico, Cattaneo and Titiunik \(2014\)](#).⁵³

Student performance was not impacted by T3 grant in Period I, and it was positively impacted in Period II by the T1, under approximately the same cash allowance amounts but different incentives. This finding suggests that performance-based incentive components seems to play a crucial role on enhancing student achievement. Nevertheless, performance standards alone do not seem to be enough to improve student outcomes, since monetary incentives appear to be also crucial (there is no grant effect on fee waiver grant). The results point out to a complementarity between certain cash allowance and strong performance-based incentives as drivers of the grant's effect on performance.

⁵³See online appendix (sections B, C and D) to see the validity of the research design, the discontinuities in grant amount and the RDD estimates.

Table 1.C.8 presents that the average discontinuity in cash allowance at T3 in Period I is 543 euros for non-movers, and cash endowments at T1 in Period II is 410 euros for non-movers. However, the null hypothesis of zero effect of being eligible for the T3 grant on non-movers student performance cannot be rejected. Results are robust to different treatment sample sizes, regarding the predetermined characteristics of applicants (PAU percentile rank, gender, STEM degree, etc), year fixed effects, to set the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) as half and twice of its value, and to perform the local polynomial regression with robust bias-corrected confidence intervals proposed by Calonico, Cattaneo and Titiunik (2014).⁵⁴

Differential effects on 2012. This section test for specific effects of the grant in 2012. The paper focus on the results on two periods, Period I and Period II. While academic incentives in Period II were homogeneous throughout the three academic years, Period II reported a change in 2012. Students had to passed 60 (80) percent of the credits attempted if the student was enrolled in STEM (non-STEM) degrees in 2010 and 2011. In 2012 the requirements rose to 65 (90) percent for students enrolled in STEM (non-STEM) degrees. Table 1.C.9 presents the results of non-parametric estimates of being eligible for a need-based grant in this academic year. The null hypothesis of a null effect of the grant on student performance (average GPA and fraction of credits earned) and dropout cannot be rejected.

The results suggest that a single increase in the fraction of credits earned does not affect student performance. This points toward the direction of this paper's conclusion concerning the design of academic incentives. A performance standard framework with clear minimum thresholds targets combining minimum course loads and certain GPAs seems to be desirable in terms of the cost-effectiveness of the policy.

⁵⁴Results available upon request.

Table 1.C.6: Discontinuities in Average GPA at T1 and T2 grants by period.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. Baseline Estimates				
Non-parametric Estimates	-0.092 (0.190) [3,402]	0.057 (0.157) [3,549]	-0.031 (0.124) [6,093]	0.455*** (0.144) [5,868]
Baseline mean	5.97	6.29	5.91	6.15
B. Sensitivity Analysis				
B. 1 Half of the optimal bandwidth				
Non-parametric Estimates	-0.131 (0.270) [3,402]	0.087 (0.223) [3,549]	-0.054 (0.172) [6,093]	0.490** (0.201) [5,868]
Baseline mean	5.971	6.291	5.908	6.155
B. 2 Twice of the optimal bandwidth				
Non-parametric Estimates	-0.158 (0.150) [3,402]	0.013 (0.132) [3,549]	-0.021 (0.114) [6,093]	0.363*** (0.112) [5,868]
Baseline mean	5.971	6.291	5.908	6.155
C. RD Robust				
Non-parametric Estimates	-0.135 (0.274) [3,402]	0.088 (0.225) [3,549]	-0.001 (0.211) [6,093]	0.501** (0.201) [5,868]
Baseline mean	5.971	6.291	5.908	6.155
D. Baseline estimates with controls				
Non-parametric Estimates	-0.065 (0.154) [3,402]	0.013 (0.140) [3,549]	-0.159 (0.104) [6,093]	0.273** (0.124) [5,868]
Baseline mean	5.971	6.291	5.908	6.155
E. Placebo test with midpoint between T1 and T2				
Non-parametric Estimates	0.0026 (0.120) [3,833]	0.1386 (0.136) [5,829]		
Baseline mean	5.990	6.327		

Notes: The table shows the RDD non-parametric estimates for applicants' average GPA. Panel A shows the baseline results estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Panel B displays the estimated treatment effect for half and twice the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#). Panel C reports the baseline results estimated performing the local polynomial regression with robust bias-corrected confidence intervals proposed by [Calonico, Cattaneo and Titiunik \(2014\)](#). Panel D exhibits the baseline estimated treatment effect controlling for year fixed effects, PAU percentile rank, STEM degree, whether the student has the Spanish nationality, and dummies equal to one for students who lived away their family home at the university entrance, female, household disability, household is considered as large family, and if the student's principal tutor is entrepreneur, blue collar or self-employed. Panel E shows a placebo test with a fictitious income eligibility threshold computed as the middle point between T1 and T2 cutoffs. Baseline mean refers to the average GPA above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported

Table 1.C.7: Discontinuities in Fraction of Credits Passed at T1 and T2 grants by period.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I	Period II	Period I	Period II
	(2010-2012) (1)	(2013-2015) (2)	(2010-2012) (3)	(2013-2015) (4)
A. Baseline Estimates				
Non-parametric Estimates	-0.017 (0.028) [3,402]	0.016 (0.026) [3,549]	0.006 (0.020) [6,093]	0.059*** (0.021) [5,868]
Baseline mean	0.78	0.80	0.77	0.79
B. Sensitivity Analysis				
B. 1 Half of the optimal bandwidth				
Non-parametric Estimates	-0.035 (0.041) [3,402]	0.013 (0.037) [3,549]	0.017 (0.028) [6,093]	0.059** (0.030) [5,868]
Baseline mean	0.779	0.805	0.769	0.787
B. 2 Twice of the optimal bandwidth				
Non-parametric Estimates	-0.021 (0.020) [3,400]	0.002 (0.018) [3,284]	0.003 (0.016) [6,087]	0.046*** (0.016) [5,855]
Baseline mean	0.779	0.805	0.769	0.787
C. RD Robust				
on-parametric Estimates	-0.035 (0.041) [3,402]	0.020 (0.029) [3,549]	0.017 (0.028) [6,093]	0.060** (0.030) [5,868]
Baseline mean	0.779	0.805	0.769	0.787
D. Baseline estimates with controls				
Non-parametric Estimates	-0.010 (0.024) [3,402]	0.012 (0.024) [3,549]	-0.014 (0.017) [6,093]	0.037** (0.019) [5,868]
Baseline mean	0.779	0.805	0.769	0.787
E. Placebo test with midpoint between T1 and T2				
Non-parametric Estimates	0.0021 (0.017) [3,833]	0.0150 (0.016) [5,829]		
Baseline mean	0.777	0.809		

Notes: The table shows the RDD non-parametric estimates for applicants' average GPA. Panel A shows the baseline results estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) for each separated regression. Panel B displays the estimated treatment effect for half and twice the optimal bandwidth proposed by Imbens and Kalyanaraman (2012). Panel C reports the baseline results estimated performing the local polynomial regression with robust bias-corrected confidence intervals proposed by Calonico, Cattaneo and Titiunik (2014). Panel D exhibits the baseline estimated treatment effect controlling for year fixed effects, PAU percentile rank, STEM degree, whether the student has the Spanish nationality, and dummies equal to one for students who lived away their family home at the university entrance, female, household disability, household is considered as large family, and if the student's principal tutor is entrepreneur, blue collar or self-employed. Panel E shows a placebo test with a fictitious income eligibility threshold computed as the middle point between T1 and T2 cutoffs. Baseline mean refers to the average GPA above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported

Table 1.C.8: Discontinuities in Average Awarded Grant and Average GPA at Fee Waiver (FW) and Displacement and other needs grant (T3) by period.

Fee Waiver Grant (FW)				
Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	Avg. Awarded Grant (euros)		Avg. GPA (0-10)	
	(Threshold 0)		(Threshold 0)	
	Period I (2010-2012) (1)	Period II (2013-2015) (2)	Period I (2010-2012) (3)	Period II (2013-2015) (4)
A. Baseline Estimates				
Non-parametric Estimates	143* (75.273) [1,787]	-75.927 (50.125) [1,714]	0.146 (0.189) [1,787]	-0.235 (0.272) [1,713]
Baseline	0.210	13.99	5.852	6.188
Displacement and Other Needs Grant (T3)				
Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	Avg. Awarded Grant (euros)		Avg. GPA (0-10)	
	(Threshold 3)		(Threshold 3)	
	Period I (2010-2012) (5)	Period II (2013-2015) (6)	Period I (2010-2012) (7)	Period II (2013-2015) (8)
A. Total Sample				
Non-parametric Estimates	730*** (137.186) [3,936]	123 (89.571) [3,909]	0.115 (0.182) [3,935]	0.149 (0.168) [3,907]
Baseline mean	662.2	673.9	5.853	6.250
B. By residence status				
Living with parents				
Non-parametric Estimates	693*** (45.053) [2,896]	56 (48.702) [2,724]	0.200 (0.238) [2,895]	0.172 (0.193) [2,724]
Baseline mean	98.90	394.1	5.748	6.154
Living outside the family home				
Non-parametric Estimates	888*** (309.610) [1,037]	-240.677 (183.962) [1,184]	0.095 (0.346) [1,037]	-0.014 (0.320) [1,182]
Baseline mean	2,303	1,342	6.138	6.480

Notes: The table shows the RDD non-parametric estimates for the average allowance amount received and average GPA for different samples. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average variable value above the eligibility threshold. The FW treatment sample includes applicants whose household parental taxable income is within 15 percent of the eligibility thresholds between fee waiver and zero. The T3 treatment sample includes applicants whose household parental taxable income is within 30 percent of the eligibility thresholds between fee waiver and Distance and Other Needs allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported below the standard errors. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.9: Discontinuities in Average Awarded Grant, Average GPA, Fraction of Credits Earned and Dropout at Residence Grant and Compensate Grant in 2012.

Treatment Sample: (Income Eligibility Threshold)	T2 Discontinuity	T1 Discontinuity
	(A/0)	(A + Δ/A)
	(1)	(2)
A. Average Allowance Amounts (euros)		
Non-parametric Estimates	2,792*** (213.989) [1,035]	657*** (141.691) [2,038]
Baseline mean	1,461	14.20
B. Average GPA (0-10)		
Non-parametric Estimates	-0.660 (0.401) [1,035]	-0.392 (0.298) [2,037]
Baseline mean	6.28	5.98
C. Fraction of Credits Earned (0-1)		
Non-parametric Estimates	-0.078 (0.054) [1,035]	0.017 (0.029) [2,037]
Baseline mean	0.81	0.77
E. Dropout from higher education		
Non-parametric Estimates	0.013 (0.011) [1,035]	-0.000 (0.010) [2,038]
Baseline mean	0.02	0.02

Notes: The table shows the RDD non-parametric estimates for the average allowance amount received and average GPA for different samples. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by [Imbens and Kalyanaraman \(2012\)](#) for each separated regression. Baseline mean refers to the average variable value above the eligibility threshold. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported below the standard errors. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.10: RDD-DID Estimates for T1 Grant.

Treatment Sample (Income Eligibility Thresholds)	T2 Discontinuity ($A + \Delta/A$)	T1 Discontinuity ($A/0$)
A. Average Allowance Amounts (euros)		
Non-parametric Estimates	-1.693*** (173.432)	121.9168 (85.365)
Baseline mean	1,481	25.89
B. Average GPA (0-10)		
Non-parametric Estimates	0.1389 (0.232)	0.3162** (0.148)
Baseline mean	5.971	5.908
C. Fraction of Credits Earned (0-1)		
Non-parametric Estimates	0.0307 (0.028)	0.0505** (0.023)
Baseline mean	0.779	0.769
D. Official Dropout (0-1)		
Non-parametric Estimates	0.0149 (0.015)	-0.0190 (0.018)
Baseline mean	0.02	0.03
E. Degree Completion (0-1)		
Non-parametric Estimates	0.0075 (0.047)	0.0964* (0.056)
Baseline mean	0.93	0.91
N	6,951	11,965

Notes: Standard errors are clustered at the student level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.11: Discontinuities for the mechanisms variables at T1 and T2 grants by period and term.

Treatment Sample: (Income Eligibility Thresholds) Period: (academic years)	First Term (Fall)				Second Term (Spring)			
	T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)		T2 Discontinuity ($A + \Delta/A$)		T1 Discontinuity ($A/0$)	
	Period I (2010-2012)	Period II (2013-2015)	Period I (2010-2012)	Period II (2013-2015)	Period I (2010-2012)	Period II (2013-2015)	Period I (2010-2012)	Period II (2013-2015)
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
A. Final exam attendance rate								
Non-parametric Estimates	-0.001 (0.015) [3,299]	-0.004 (0.010) [3,47]	0.014 (0.012) [5,904]	0.022*** (0.008) [5,703]	-0.014 (0.016) [3,282]	0.018 (0.019) [3,4]	-0.004 (0.013) [5,885]	0.046*** (0.015) [5,6]
Baseline mean	0.916	0.947	0.920	0.941	0.901	0.921	0.911	0.922
B. GPA on final exams taken								
Non-parametric Estimates	-0.005 (0.151) [3,288]	0.081 (0.146) [3,461]	-0.033 (0.105) [5,881]	0.185 (0.116) [5,693]	-0.070 (0.168) [3,253]	-0.007 (0.136) [3,343]	-0.100 (0.107) [5,815]	0.345*** (0.117) [5,545]
Baseline mean	6.482	6.631	6.349	6.541	6.564	6.811	6.485	6.610
C. Selection on courses								
Non-parametric Estimates	0.333 (1.363) [3,299]	0.623 (1.293) [3,47]	2.288** (1.136) [5,904]	1.427 (0.992) [5,703]	1.504 (1.663) [3,282]	0.541 (1.504) [3,4]	2.499** (1.184) [5,885]	1.683* (1.006) [5,6]
Baseline mean	51.57	51.06	52.56	53.51	51.37	50.90	53.13	53.40
D. GPA on Mandatory Subjects								
Non-parametric Estimates	-0.058 (0.208) [3,274]	0.041 (0.160) [3,426]	0.086 (0.158) [5,865]	0.293** (0.140) [5,628]	-0.149 (0.223) [3,261]	0.122 (0.205) [3,374]	-0.183 (0.146) [5,85]	0.491*** (0.155) [5,57]
Baseline mean	5.961	6.281	5.868	6.160	5.942	6.227	5.954	6.122
E. GPA on Elective Subjects								
Non-parametric Estimates	0.089 (0.381) [1,065]	-0.650* (0.391) [1,038]	0.055 (0.306) [1,693]	-0.210 (0.303) [1,48]	-1.136*** (0.396) [1,101]	0.312 (0.355) [1,021]	0.089 (0.262) [1,777]	0.516 (0.333) [1,509]
Baseline mean	7.077	7.451	6.992	7.394	7.029	7.381	6.875	7.230

Notes: The table shows the RDD non-parametric estimates for the average allowance amount received and average GPA for different samples. The treatment effect is estimated using a triangular kernel and the bandwidth is computed as the optimal bandwidth proposed by Imbens and Kalyanaraman (2012) for each separated regression. Baseline mean refers to the average variable value above the eligibility threshold. Standard errors are clustered at the student level and displayed in parenthesis. The number of observations used in the non-parametric estimations are reported below the standard errors. Total number of observations are in squared brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.C.5 RDD-DID

I perform a reduced-form analysis of RDD-Differences-in-Differences (RDD-DID). Let $E_{i,k,t}$ be a dummy variable that takes value one if applicant i is eligible for a grant of level k ($k = 1, 2$) at year t , and zero otherwise. Eligibility for a level k grant is a deterministic function of the applicant's net household taxable income c_{it} and the number of family members m_{it} :

$$E_{i,k,t} = \mathbb{1}\{c_{it} \leq \bar{c}_k(m_{it})\} \quad (1.17)$$

$$Y_{it} = \alpha + \sum_{k=1}^2 \beta_k E_{i,k,t} + c_{it} + c_{it} * E_{i,k,t} + \epsilon_{it} \quad (1.18)$$

Let Period II be a dummy variable that takes value one if years 2013-2015 (Period II). The coefficient of interest is β :

$$Y_{it} = \alpha + c_{it} + c_{it} * E_{i,k,t} + c_{it} * \text{Period II} + c_{it} * E_{i,k,t} * \text{Period II} + E_{i,k,t} + \text{Period II} + \sum_{k=1}^2 \beta_k E_{i,k,t} * \text{Period II} + \epsilon_{it} \quad (1.19)$$

Note that in this model, the coefficient estimates do not change. It is the same that perform a RDD separately by each period than RDD-DID in terms of the coefficients (standard errors can slightly vary because of the degrees of freedom). The advantage of this model is the fact that this method allow us to compute the standard errors of the difference in changes directly, since it is not straightforward to compute the standard error of the difference between the two coefficients estimated separately by period (it would require bootstrapping).

Results are reassuring on non-statistically different average grant amounts across periods, an increase in student performance and degree completion, and no statistically significant effect on dropout from higher education.

Table 1.C.12: RDD-DID Estimates for T1 Grant.

Treatment Sample (Income Eligibility Thresholds)	T2 Discontinuity ($A + \Delta/A$)	T1 Discontinuity ($A/0$)
A. Average Allowance Amounts (euros)		
Non-parametric Estimates	-1.693*** (173.432)	121.9168 (85.365)
Baseline mean	1,481	25.89
B. Average GPA (0-10)		
Non-parametric Estimates	0.1389 (0.232)	0.3162** (0.148)
Baseline mean	5.971	5.908
C. Fraction of Credits Earned (0-1)		
Non-parametric Estimates	0.0307 (0.028)	0.0505** (0.023)
Baseline mean	0.779	0.769
D. Official Dropout (0-1)		
Non-parametric Estimates	0.0149 (0.015)	-0.0190 (0.018)
Baseline mean	0.02	0.03
E. Degree Completion (0-1)		
Non-parametric Estimates	0.0075 (0.047)	0.0964* (0.056)
Baseline mean	0.93	0.91
N	6,951	11,965

Notes: Standard errors are clustered at the student level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.C.6 Comparability between Period I and Period II

. An important concern corresponds to the degree of comparability between applicants for a need-based grant in Period I and Period II. I test whether the students' observable characteristics of the comparison group at T1 and T2 grant in Period I are similar to applicants in Period II. I test the comparability of these students performing a t-test of the difference in observable characteristics between period. **Table E1** presents the results of this analysis. The null hypothesis of equality of the observable characteristics between periods cannot be rejected for three quarters of the variables at T1 grant, and for more than half in T2 grant.

Table 1.C.13: Difference in Baseline Means by period and treatment sample.

Treatment Sample: (Income Eligibility Thresholds)	T2 Discontinuity ($A + \Delta/A$)			T1 Discontinuity ($A/0$)		
	Baseline Mean	Difference	P-Value	Baseline Mean	Difference	P-Value
	PI (1)	PI vs. PII (2)		PI (3)	PI vs. PII (4)	
Female	.49	.017	.56	.47	.025	.44
Spanish	.95	.014	.02	.99	-.0023	.28
Access to University Percentile rank	54.57	2.27	.00	56.54	.38	.25
STEM degree	.34	-.02	.21	.40	-.025	.36
Households taxable income (euros)	17,959	1,336	.00	43,703	3,306	.00
Number of family members	3.7	.0074	.46	3.7	.042	.28
Disability	.011	-.007	.13	.014	.005	.66
Large family condition	.12	-.016	.08	.13	.012	.84
Live outside the family home	.29	-.027	.14	.3	.005	.00
Entrepreneur Parent	.07	-.005	.00	.04	.003	.29
Blue Collar Parent	.42	-.05	.00	.2	-.034	0.01
Self-Employed Parent	.064	.004	.4	.023	.0044	.16
Awarded Grants	1	-.013	.44	.41	.088	.36

Notes: The table shows a t-test for the differences in baseline means on different applicants' observable variables. Baseline mean refers to the average value of the observable variable above the eligibility threshold. The T1 treatment sample (column 1) includes applicants whose household parental taxable income is within 50 percent of the eligibility thresholds between fee waiver and Residence Grant allowances. The T2 treatment sample (column 2) includes applicants whose household parental taxable income is within 60 percent of the eligibility thresholds between Residence Grant and Compensate Grant allowances. The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. The variable "Access to University Percentile Rank" is computed as the percentile rank of the students' academic year high school graduation on the *PAU* grade over the poll of BCG grant applicants from 2004-2015. Household's taxable income is expressed in constant 2015 euros. Standard errors are clustered at the student level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.C.7 Equity Effects

In order to analyze whether the increase of academic requirements have significant equity effects to the average population of applicants to the grant, I performed a Logit analysis and a Difference-in-Difference (DID) model, in addition to the RDD analysis. I count as dropout from higher education if the student's file label as close (i.e., student has not graduated and file has not been moved to another higher education university). The analysis explores two different dropout variables: (i) Yearly dropout, signaling the flow of students who dropout at the end of certain academic year (open files are included in this computation); (ii) Net dropout, which is the total cohort dropout (in this variable, the student can only be graduated or dropout). This method uses two approaches: (i) Compares two cohorts, first (non-first) year students applicants the previous year of the reform versus the year of the reform; (ii) Uses the full span of six-year period to see how dropout changed before and after the reform.⁵⁵ The regressions control for all available student predetermined observable characteristics.

The Logit analysis shows that, in general, female and students with higher rank in the access to university exam presents lower probability of dropping out, but movers and non-first year enrolled in STEM degree show higher probability of drop out. However, both dropout measures do not significantly change after the reform, neither when controlling for student individual observable characteristics nor when do not. This results are robust for non-first year and first year students, which is the sub-population of students with the highest probability of dropping out.

The DID model compares students who performed higher than the academic requirements of Period I but not overcame those of Period II or 2012's (Treatment Group), with students who do not meet the academic requirements of Period I

⁵⁵Note that for the first approach I can only use the first change in the reform at 2012/2013, since I have data from 2010–2015. Net dropout rate for cohort 2013/2014 would be bias, due to the fact that undergraduate program covers four years of education, and this cohort has only attend three years of bachelor in 2015/2016. Given the definition of net dropout rate (they can either be graduated or dropout), it would increase their fraction of dropout by construction. Thus, evaluating this variable for 2012/2013 cohort is interesting due to the fact that it was the first increase in academic requirements.

(Control Group), before and after the reform:

$$\text{Dropout}_{it} = \alpha + \beta \text{Treatment}_{it} * \text{Period II}_t + \gamma \text{Treatment}_{it} + \delta \text{Period II}_t + \nu X_{it} + \epsilon_{it} \quad (1.20)$$

where Treatment_{it} is a dummy variable that takes value one if the students belongs to the treatment group and zero otherwise, Period II_t takes value one if the academic year is 2013 or higher, and X_{it} is a vector of student observable characteristics. [Bénabou and Tirole \(2000\)](#) model predicts that when the academic requirements increase, some of the weaker agents types would *give up*, exert zero effort and drop out from higher education. Results show that an the increase in the academic requirements of Period II and 2012 did not have any statistically significant impact on students drop out. Students who theoretically would not meet the academic requirements for the next academic year but they do with the *weaker* standards, do not significantly increase their level of drop out, neither these who were surprised by the reform (in 2012 and 2013 respectively) nor these of the whole Period II. This result seem to be true for first and non-first year students. The increase in the academic requirements does not seem to be high enough to induce these students to drop out from higher education. In addition, the RDD-DID analysis shows no statistically significant change in dropout from higher education across periods.

Table 1.C.14: Logistic models of net dropout rate: Academic year 2011-12 vs. 2012-13.

Variables	First Year Students			Non-first Year Students		
	(1)	(2)	(3)	(4)	(5)	(6)
Academic year 2012–2013	0.235 (0.155)	0.235 (0.157)	0.232 (0.158)	-0.195* (0.114)	-0.168 (0.116)	-0.143 (0.117)
Households taxable income (euros)		-0.000 (0.000)	-0.000 (0.000)		-0.000*** (0.000)	-0.000*** (0.000)
Female		-0.659*** (0.176)	-0.698*** (0.177)		-0.194 (0.131)	-0.184 (0.132)
Access to University Percentile rank		-0.008*** (0.003)	-0.008*** (0.003)		-0.007*** (0.002)	-0.007*** (0.002)
STEM degree		-0.328* (0.178)	-0.284 (0.180)		1.011*** (0.127)	1.038*** (0.130)
Spanish		0.625 (0.412)	0.611 (0.415)		0.987** (0.459)	0.939** (0.461)
Number of family members			-0.255*** (0.098)			0.219*** (0.071)
Live outside the family home			0.318* (0.171)			0.215 (0.137)
Disability			0.043 (0.586)			0.669* (0.375)
Large family condition			0.163 (0.274)			-0.512** (0.225)
Entrepreneur Parent			0.101 (0.314)			-0.145 (0.240)
Blue Collar Parent			-0.131 (0.189)			-0.301** (0.135)
Self-Employed Parent			0.102 (0.338)			-0.597* (0.319)
Observations	1,361	1,361	1,361	6,390	6,390	6,390

Notes: The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. The variable "Access to University Percentile Rank" is computed as the percentile rank of the students' academic year high school graduation on the PAU grade over the poll of BCG grant applicants from 2004-2015. Household's taxable income is expressed in constant 2015 euros. Standard errors are clustered at the student level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.15: Logistic models of yearly dropout rate (2010–2015).

Variables	First Year Students			Non-first Year Students		
	(1)	(2)	(3)	(4)	(5)	(6)
Period II (2013–2015)	-0.226 (0.151)	-0.226 (0.152)	-0.207 (0.152)	-0.731*** (0.157)	-0.715*** (0.157)	-0.704*** (0.158)
Households taxable income (euros)		0.000* (0.000)	0.000** (0.000)		0.000 (0.000)	0.000 (0.000)
Female		-0.369** (0.169)	-0.386** (0.170)		-0.087 (0.161)	-0.088 (0.161)
Access to University Percentile rank		-0.019*** (0.003)	-0.019*** (0.003)		-0.013*** (0.003)	-0.015*** (0.003)
STEM degree		-0.024 (0.164)	0.005 (0.166)		0.372** (0.159)	0.421*** (0.161)
Spanish		0.062 (0.350)	0.042 (0.352)		0.098 (0.391)	0.012 (0.393)
Number of family members			-0.233*** (0.090)			-0.049 (0.089)
Live outside the family home			0.111 (0.172)			0.357** (0.166)
Disability			-1.346 (1.020)			-0.322 (0.599)
Large family condition			-0.124 (0.288)			0.347 (0.241)
Entrepreneur Parent			-0.078 (0.300)			-0.143 (0.295)
Blue Collar Parent			-0.370** (0.182)			-0.542*** (0.185)
Self-Employed Parent			-0.329 (0.372)			-0.207 (0.349)
Observations	8,440	8,440	8,440	21,106	21,106	21,106

Notes: The variable "Live away their family home" refers to the fraction of applicants who live away their family home, measured by the student' postal code when they access higher education. The variable "Access to University Percentile Rank" is computed as the percentile rank of the students' academic year high school graduation on the *PAU* grade over the poll of BCG grant applicants from 2004-2015. Household's taxable income is expressed in constant 2015 euros. Standard errors are clustered at the student level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.C.16: DID model of yearly dropout rate (2010–2015).

A. Period I vs. Period II						
Variables	First Year Students (1)	Students (2)	First Year Students (2010–2013) (3)	Students (2010–2013) (4)	Non-first Year Students (5)	Students (6)
Treatment*Period II	0.011 (0.009)	0.012 (0.009)	0.016 (0.013)	0.017 (0.013)	-0.001 (0.004)	-0.001 (0.004)
Treatment	-0.043*** (0.004)	-0.042*** (0.004)	-0.043*** (0.004)	-0.042*** (0.004)	-0.018*** (0.002)	-0.018*** (0.002)
Period II	-0.020** (0.008)	-0.021*** (0.008)	-0.022* (0.012)	-0.024** (0.012)	-0.002 (0.004)	-0.005 (0.004)
Student Controls	No	Yes	No	Yes	No	Yes
Observations	10,424	10,424	8,999	8,999	15,921	15,921

B. Period I vs. Increase in 2012's Requirements						
Variables	First Year Students (7)	Students (8)	First Year Students (2010–2012) (9)	Students (2010–2012) (10)	Non-first Year Students (11)	Students (12)
Treatment*Year2012	0.004 (0.008)	0.005 (0.008)	0.006 (0.010)	0.007 (0.010)	-0.005 (0.004)	-0.006 (0.004)
Treatment	-0.034*** (0.006)	-0.030*** (0.006)	-0.034*** (0.006)	-0.030*** (0.006)	-0.015*** (0.003)	-0.017*** (0.003)
Year 2012	-0.029*** (0.006)	-0.036*** (0.007)	-0.035*** (0.008)	-0.045*** (0.010)	0.001 (0.003)	-0.002 (0.004)
Observations	8,992	8,992	7,171	7,171	12,617	12,617
Student Controls	No	Yes	No	Yes	No	Yes

Notes: Treatment refers to the group of students whose performance is in between academic requirements of Period II (2012) and Period I. These students would meet the requirements in Period I, but not in Period II (2012). Student controls include all the student predetermined observable characteristics of the students available in the data. Standard errors are clustered at the student level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

1.C.8 Academic requirements for BCG grant

This section summarizes the academic standards for being eligible for a BCG grant over the six-year period studied (2010–2015). In order to be eligible for a need-based grant, students must have complied with a minimum fraction of credits earned and average GPA the year before application. **Table F1** shows a summary of the different performance standards required by year, degree and cohort. It is remarkable the increase in the fraction of credits earned required in 2012, and the posterior change of the entire framework of academic incentives in 2013, which incorporates the average GPA of the year before application plus a variable component which depends on performance and family income the year of grant application (the variable component formula is described in **equation 11**). Figure displays a graphical summary of academic requirements for non-freshmen students. Academic incentives varied between the two periods:

- *Period I (2010–2012)*: incentives were based on the fraction of credits earned the year before application.
- *Period II (2013–2015)*: academic standards were based on the fraction of credits earned the year before application, the average GPA the year before application and in the application year (through the grant’s individual variable component).

Table 1.C.17: BCG grant academic requirements.

A. Non-First year students					
	Fraction of pass credits in the last academic year over 60 ECTS		Average GPA		Grant rights
	STEM	Humanities and Social Sciences	STEM	Humanities and Social Sciences	
Before 2012	60%	80%	None	None	All
2012	65%	90%	None	None	All
2013 onward	85%	100%	None	None	All
	65%	90%	>=6	>=6.5	All
	65%	90%	<6	<=6.5	Only Fee Waiver
B. First year students					
	Average Grade in PAU				Grant rights
Before 2013	5/10				All
2013 onward	6.5/10				All
	5.5/10				Only Fee Waiver

Notes:

Variable component formula.

$$C_j = C_{min} + \left[(C_{total} - S * C_{min}) * \frac{(N_j/N_{max} * (1 - (\frac{R_j}{R_{max}})))}{\sum_{i=1}^S (N_i/N_{max}) * (1 - (\frac{R_i}{R_{max}}))} \right] \quad (1.21)$$

where C_j = variable component amount that student j receives; C_{min} = minimum variable component; C_{total} = total amount of variable component to distribute among grant's recipients (depend on the year); S = number of applicants who receive variable component; N_j = applicant's average GPA; N_i = average GPA of each applicant to which S refers; N_{max} : average GPA obtained by the best decile of the same degree; R_j = applicant's income per capita; R_i = income per capita of each applicant to which S refers; R_{max} = maximum income per capita to be awarded with variable component.

Chapter 2

The Gender Gap in Student Performance: The Role of the Testing Environment*

*We would like to thank Ghazala Azmat, Antonio Cabrales, Caterina Calsamiglia, Esther Dufo, Gabrielle Fack, Martin Fernandez-Sanchez, Marc Gurgand, Julien Grenet, Guy Michaels, Rosa Sanchis-Guarner and Olmo Silva. We would like to especially thank Ismael Sanz, Luis Pires and Gerardo Azor working for their invaluable help in collecting the data. We thank seminar participants at Paris School of Economics, London School of Economics, *College de France*, Lund Economic Demography Seminar, Bocconi Dondena Seminara and Sevilla Economics of Education Seminar. In addition, José Montalbán thanks to the Bank of Spain for its financial support. All errors are our own.

Abstract

There is a substantial body of literature that focuses on measuring how gender differences in cognitive abilities and gender-stereotyping norms impact the gender gap in student performance. However, little attention has been devoted to investigating how the organization of student testing may influence the relative performance of male and female students. This paper analyzes the gender gap in test scores that arises as a result of differential responses by boys and girls to the testing environment. To that end, we exploit a unique randomized intervention on the entire population of students in the 6th and 10th grades in the Region of Madrid (Spain). The intervention assigned schools to either internally or externally administered testing. We find that girls do worse than boys in exams that are externally administered, especially in subjects that are male-dominated. Additional survey evidence on stress, self-confidence, and effort suggests that lower relative female performance in externally administered tests results from a lower ability to cope with stressful situations as a result of less familiarity with the testing environment.

JEL Codes: D03, J16, I21, C30

Keywords: stress, gender gaps, performance

2.1 Introduction

Gender differences in academic attainment and achievement have dramatically reversed over the past 60 years. Whereas several decades ago men graduated from college at much higher rates than women, the situation is now reversed [Goldin, Katz and Kuziemko \(2006\)](#). In most industrialized countries, men continue to outperform girls in Science, Technology, Engineering, and Mathematics (STEM) ([Fryer Jr and Levitt, 2010](#); [Ellison and Swanson, 2018](#)). Women participation rates in STEM fields is lower than those of their male peers ([Card and Payne, 2017](#)), and the increasing female participation trend has leveled off in most OECD countries ([OECD, 2016](#)).¹ This gender difference has implications for future labor market outcomes, since STEM occupations offer higher wages ([Brown and Corcoran, 1997](#); [Black et al., 2008](#); [Blau and Kahn, 2017](#)).

Most industrialized countries regularly undertake standardized testing of school children at different stages of the education path. As well as being used as an essential educational policy tool for improving student's performance and to shape education systems, standardized test results also determine an individual's educational choice set later on in life. However, measuring student ability through standardized testing may be problematic. Test scores depend not only on cognitive ability, but also in other factors such as non-cognitive skills ([Cunha and Heckman, 2007](#)), the ability to cope with competition ([Niederle and Vesterlund, 2010](#)), and the student's stereotypical beliefs ([Ehrlinger and Dunning, 2003](#)), among others. All these factors may artificially distort the underlying distribution of cognitive skills, making it difficult to infer a student's true ability. Despite the relatively extensive literature on the gender differences in student performance, it remains unclear to what extent the organization of student testing influences the relative performance of male and female students in standardized tests. This paper investigates the causal effect of the testing environment on students' standardized test scores.

We use a randomized control trial which introduced changes to the test-taking

¹See [Kahn and Ginther \(2017\)](#) for a review of the literature on the gender gap in STEM fields.

environment implemented on the full population of students in 6th and 10th grades in the Region of Madrid (Spain). Some schools were randomly allocated to either an internal or an externally administered standardized testing procedure. The goal of the intervention was not aimed at reducing gender gaps in performance in any way. Instead, the objective was to reduce the costs associated to running the tests, which had been run by the Regional Government until then.

In schools where the testing was internally administered teachers had a lot more control on how they presented the test to students, and were less likely to emphasize the importance of the test. In schools where the testing was externally administered, teachers had less control over the way the test was presented to the students. The increased formality in externally administered tests suggests that the importance assigned to the testing process was higher in externally administered test-taking environments. In the internally administered test-taking environment invigilators, and markers were already working at the school the student was currently attending, and thus likely to be known to the student. In the externally administered test-taking environment, students did not know the invigilator and markers, who were not employed by the student's school.

We estimate the causal effect on student performance under familiar versus non-familiar test-taking environments using unique linked administrative student-level information on test performance for the population of students in 6th and 10th grades in 2016/2017 and 2017/2018. We find that girls underperform in externally administered testing environments relative to boys. In particular, whereas under an internally administered testing environment boys in 6th Grade outperform girls by 0.1 standard deviations in Mathematics, and girls outperform boys in Science, Spanish and English by 6.8, 38 and 24 percent of a standard deviation, the gender gap widens in about 0.05 of a standard deviation for Mathematics (an effect size of 50 percent), and narrows in Science, Spanish and English (an effect size of 58, 13 and 18 percent respectively) under an externally administered testing environment. Similarly, whereas under an internally administered testing environment boys in Grade 10th outperform girls in Mathematics and Social Sciences by about 16 and 7

percent of a standard deviations, and girls outperform boys in Spanish and English by 15 and 11 percent of a standard deviations, we find that the gender gap in student performance widens by about 0.08 and 0.05 of a standard deviation in Mathematics and Social Sciences (an effect size of 52 and 62 percent respectively) under an externally administered test.

Girls' lower relative performance in externally administered testing environment is more acute for male-dominated subjects such as Mathematics and Sciences relative to other female-dominated subjects such as languages. This result is in line with the evidence from competitive test-taking environments that report girls performing worse than boys, particularly when the task has a strong gender stereotype and when a large fraction of competitors on tests are male ([Shurchkov, 2012](#); [Niederle, 2017](#); [Iriberri and Rey-Biel, 2017](#)). The negative effect of an externally administered testing environment on student's performance is robust to several sensitivity checks, such as controlling for school fixed effects and school characteristics. Results are also robust to different school samples, and to placebo tests that randomly allocate the intervention across schools. Heterogeneity analysis shows that our results hold across different samples of schools with different socio-economic characteristics, as well as the school's ownership status (public or private).

We also rule out alternative hypothesis such as girls' marks being artificially bumped up in schools where the test was internally administered ([Lavy, 2008](#); [Cornwell, Mustard and Van Parys, 2013](#); [Breda, 2015](#); [Terrier, 2016](#)). Marking was blind, and the tests were multiple-choice tests for most subjects so that markers could not figure out whether the student was a boy or a girl from the hand-writing. We find that the negative effect of an externally administered testing environment on student scores is stronger in the multiple-choice part of the test than in the writing part. We also rule out that our results are driven by gender differences in motivation driven by the gender and other characteristics of the student's class teacher ([Ouazad, 2013](#)).

Our results are consistent with women performing relatively worse than men in unfamiliar situations with heightened pressure, particularly in traditionally

male-dominated in which the stereotype threat is higher than in female-dominated subjects. There is evidence that dealing with high pressure environments such as an exam competes with the resources in the working memory that would otherwise be used to solve the test instead, leading to lower performance than expected given an individual's cognitive ability (Beilock, 2010).² In fact, experimental evidence from the psychology literature shows that pressure resulting from an emphasis on the importance of the positive consequences of good performance and/or the negative consequences of bad performance (the so-called importance of the process) can lead individuals to perform more poorly than expected given their skill level (Beilock, 2008). This is particularly so for women in male-dominated subjects, because of the “stereotype threat” about how one's social group should perform (e.g., “girls can't do math”) produces less-than-optimal execution (Steele, 1997; Carr and Steele, 2009).

To directly test whether this mechanism is at play, the last part of the paper uses responses to a student questionnaire. This questionnaire was exam-specific, and was conducted right after students had finished each exam in the second year of the intervention. In particular, students were asked about how stressed they were during the test, whether they felt were good at the particular subject being tested, the amount of effort put into preparing for the test, and the amount of effort they put during the test. Across all subjects, we find that girls report more stress levels during the test and lower levels of confidence. However, girls studied more for all tests, and put more effort during all tests. We find that gender differences along these dimensions are exacerbated during externally administered testing environments, although the coefficients are less precisely estimated.

Our paper contributes to a growing literature analyzing the effects of test-taking environments on students' performance, and how it varies by gender, in three important ways. Following the experimental literature showing that women underperform

²The psychology literature has argued that the attention, memory, and computational processes involved in solving math problems that require the use of working memory may be particularly susceptible to pressure-laden environments, such as tests (Beilock, 2008). High-pressure environments can also cause individuals, involuntarily, to consciously think about the task, shifting control from ‘automatic’ to ‘controlled’ processes that are less effective (see Camerer, Loewenstein and Prelec (2005) for a detailed account of automatic and controlled processes; Ariely et al. (2009)). This phenomenon is commonly known in the literature as “choking under pressure”, see Baumeister (1984).

relative to men in competitive environments ([Niederle and Vesterlund, 2007](#); [Ors, Palomino and Peyrache, 2013](#)), several studies focus on the competitive nature of exams. The number of competitors usually measures competitiveness and the final score being dependent on relative performance. These studies usually find that increases in competitive pressure favor men over women ([Iriberry and Rey-Biel, 2011](#); [Ors, Palomino and Peyrache, 2013](#)). A recent study looks at girls underperformance in a non-competitive environment where scores depend on absolute, rather than relative performance. They show that women perform worse when under pressure, defined as higher stakes associated with the exam ([Azmat, Calsamiglia and Iriberry, 2016](#)). Here we focus on the familiarity of the testing environment as an indicator of pressure. The randomized nature of the intervention allows us to focus on the effect of test-taking pressure on student's scores as a result of an exogenous variation in the familiarity with the testing environment, while other factors such as the gender of markers and teachers, the competitiveness of the environment, and the skills being tested are held constant. Second, our results are less likely to suffer from external validity and generalization. We exploit a randomized intervention rather than a quasi-natural experiment or a lab experiment as in previous studies. We also use population-level data for the entire region of Madrid, rather than survey-based data or small populations based on a single school or a selected sample of high-ability individuals. Third, additional evidence from students' questionnaires allows us to test the mechanism directly at play. We rule out effort as the driver behind the increased gender differences in scores under externally administered testing environments. Instead, increased stress and lower confidence under externally administered testing environments seems to be the major factor at play.

The remainder of this paper is structured as follows. Section 2 describes the institutional background of the Spanish education system, standardized testing, and on the randomized intervention. Section 3 describes the data used in the paper and presents descriptive statistics. Section 4 discusses the empirical strategy and the internal validity of the research design. Section 5 presents the main results, and Section 6 explores possible mechanisms behind the increased gender differences in

scores under externally administered testing environments. Section 7 concludes.

2.2 Institutional Background

2.2.1 Standardized Testing in Spain

The Spanish education system has six years of primary school (from age 6 to 12) and four years of compulsory secondary education (from age 13 to 16). At the age of 16, students can either leave the education system or continue with two additional years choosing between an academic (leading to a university degree) or a vocational track. About 85% of schools in the Region of Madrid are publicly funded. Public schools can either be fully funded by the Government and managed by civil servants and local school boards, or can be mostly financed through public funds but be privately run (*centros concertados*, or semi-public schools).

The education policies are highly decentralized and typically are run at the regional level. The Spanish Government regulates by law the general principles guiding the provision of education in Spain.³ However, regional governments have a high degree of autonomy on the implementation of education policy (e.g., regional governments set their curriculum). These governments have the authority to implement various education policies as long as these do not contradict the Spanish national law. One such education policy is standardized testing. The Spanish Government determines the general characteristics, design and requirements of the test to guarantee common standards at the national level. The law establishes that every student in Spain must be tested at the end of primary (6th Grade) and secondary education (10th Grade).⁴ However, regional governments are in charge of the final design, implementation, and marking of the test.⁵

The Regional Government of Madrid has been performing standardized testing in several subjects since the 2015/16 academic year.⁶ The exam is taken by the entire

³The [Organic Law 8/2013](#) (*Ley Orgánica para la mejora de la calidad educativa, LOMCE*).

⁴Article 21.1 and 29 of the [Organic Law 8/2013](#).

⁵Article 4.1 of the [Royal Decree 1058/2015 of November 20](#).

⁶Before 2015/16, the Region of Madrid was performing a standardized exam administered for all 6th Grade students in the region since 2004/2005 school years. This test was called Essential

population of students in Primary School Grades 3 and 6, and Secondary School Grade 10. The tests are intended to provide additional information on the students' relative performance to teachers, parents, and students, and the results from the tests have no academic consequences. Parents are sent a letter with the child's test scores, the average score of the school they are attending, and the average score of the entire region. There are no statistics made publicly available from the test. Unlike in other countries, the results from the test are not used to provide a public ranking of schools in the Region of Madrid.

During 6th Grade, students are tested in Mathematics, Science, Spanish and English. During 10th Grade students are tested in Mathematics, Spanish, English and "Social and Civic Competence" (Social henceforth); the latter corresponds to the material taught in History. Students diagnosed with some learning difficulties such as dyslexia were allowed extra time (an average of twenty additional minutes). The exam was conducted in Spanish for all students (except for the English test), and tests from previous academic years can be downloaded in the website of the Region of Madrid.⁷

The exams in Mathematics and Science are multiple-choice tests, while exams have two parts for English and Spanish: linguistic competence (multiple-choice test) and written communication (written test). Although the Social test does not include an explicit written part, there are several questions in which the student must write specific words or sentences. This part is considered as non-multiple-choice in the analysis further on. The length of the 6th Grade tests was one hour in Mathematics, 40 minutes in Science, one hour and fifteen minutes in English (25 minutes of written communication and 50 minutes of linguistic competence), and one hour and twenty minutes for Spanish (25 minutes of written communication and 55 minutes of linguistic competence). The length of the 10th Grade tests was one hour for each subject.

Knowledge and Skills test (CDI-*Conocimientos y Destrezas Indispensables*).

⁷An example of 6th Grade exams can be found in the following link: <http://www.comunidad.madrid/servicios/educacion/evaluacion-6o-primaria>.

2.2.2 A Randomized Intervention in the Administration of Standardized Testing

In the 2016/17 and the 2017/18 academic years, the Regional Government of Madrid implemented a randomized intervention in the way the standardized testing of 6th and 10th Grades was administered.⁸ Some schools were randomly assigned to internally administered standardized testing, whereas other schools were assigned to externally administered standardized testing. The intervention was designed so that by the second academic year 2018/19 all schools had undergone internally administered testing implemented at least once. The justification for such an intervention was purely financial, as the previous practice of externally administering testing turned out to be extremely costly for the Regional Government.

The randomization was performed at the school level and selected a representative school sample within each of the five administrative districts in the region of Madrid. The selection process was stratified at the level of school type (public, semi-public, and private), and language of instruction (schools with English and Spanish instruction and schools with Spanish instruction only). The Region of Madrid is divided into five administrative districts (called DATs):⁹ The Capital (municipality of Madrid), the North, the East, the West, and the South. The randomization was performed in administrative districts North, South, East and West in 2016/2017, and districts Capital, East and West in 2017/2018. Figure 2.A.1 and Figure 2.A.2 present the distribution of the randomization across years for 6th and 10th Grade, respectively.

In both internally and externally administered testing, the teacher who invigilated the test was not the same one who marked the test. Each student had an encrypted student identifier. This identifier was used in the marking of the test, so that the teacher who marked the test did not have access to the student's personal information, such as name or gender. The exams, the schedule, and the time when the test was performed were exactly the same in schools that administered the test internally or

⁸These interventions were regulated by the regional law [B.O.C.M. number 110, May 10th of 2017, page 248](#).

⁹DAT means *Direcciones de Área Territorial*.

externally. The Regional Government of Madrid Education Unit sent some staff to several internally administered schools to monitor whether the exam was implemented according to the rules.

In the schools where testing was internally administered, the school appointed invigilators (proctors) and markers who were teachers already working at the school. None of the internal invigilators and markers were assigned to students whom they taught. In these schools, the tests were sent by the Regional Government two hours before the testing started, and the school printed out the questions before distributing the test to students. When the test was completed, the internal invigilators brought the exams to the school principal's office, who assigned several school teachers to mark them.

In schools where testing was externally administered, the Regional Government of Madrid directly appointed invigilators and markers, who were teachers working in public schools in the Region of Madrid but were not employed by the school where the test was taking place. In the case of externally administered tests, invigilators picked up a sealed box from the regional office with the school label on it. He or she brought this box to the school in the morning of the test and opened the boxes in the principal's office two hours before the test took place. When the test was completed, invigilators put the tests in an official box marked with an encrypted school identifier and brought the sealed box to the Regional Government of Madrid Education Unit, where assigned teachers marked them in an office of the educational unit office.

2.3 Data and Descriptive Statistics

The data for our main analysis are based on administrative student-level data of 6th and 10th graders who took the standardized test in the Region of Madrid in the academic years 2016/2017 and 2017/2018, together with administrative information on school characteristics. These data sets are matched on the basis of an encrypted student and school identifier. Over 95% of students enrolled in the Region of Madrid

took the test. Alongside with test scores, the Regional Government of Madrid provided us with additional administrative information on school characteristics, such as the average student socio-economic status and the percentage of students who do not pass a grade. There is no attrition problem across treatment and control groups since all schools take part in the standardized testing.

The students' test scores database includes information on the overall student score, as well as the score for every single question in the test for every subject: Science in 6th Grade, Social Sciences in 10th Grade, and Mathematics, Spanish and English in both grades. School characteristics include information on whether the school administered the standardized testing internally or externally, as well as other descriptive information such as the socio-economic status of students and whether the school is public or private. For a full description of these variables see Table 2.B.2.

Panel A of Table 2.A.1 presents the number of students per academic year and grade by administrative district (DAT) to illustrate the randomization procedure for each grade. The first columns of each grade show that students in the 6th and 10th grades attending schools in all but the district of Madrid were allocated to either an internal or external invigilator in the examination of the 2016/17 academic year. The second columns of each grade show all students except those attending schools in the North and South districts were randomly allocated to either an internal or external invigilator for the exam in the 2017/18 academic year. Panel B presents the number of schools and students per grade and year in our baseline sample. We limit the sample to students attending schools in administrative districts where randomization took place in a given year, and students who took exams in all the subjects ¹⁰ Overall, there are 67,841 students (1,144 schools) in Grade 6th and 56,880 students (788 high schools) in Grade 10th. In 6th Grade, about 80 percent of students performed the test in a school internally administered, whereas 20 percent took it in a school externally administered school. In 10th Grade, about 65 percent of students performed the test

¹⁰In 6th Grade, we drop 3,656 and 4,425 students who did not attend the four tests in 2016/17 and 2017/18 academic years. In 10th Grade, we discard 6,860 and 8,907 students who did not attend the four tests in 2016/17 and 2017/18 academic years. Results are robust to the inclusion of the full population of students and available under request.

in a high school internally administered, whereas 35 percent took it in a high school externally administered school. The proportion of girls is the same as the proportion of boys. Test scores are standardized to have a distribution with zero mean and a unit standard deviation. The standardization is performed by academic year and subject.

2.4 Empirical Strategy

To estimate the causal effect of having externally versus internally administered testing on gender relative student performance, we use the following specification:

$$Y_{isdt} = \alpha + \delta External_s * Female_i + \beta Female_i + \gamma External_s + \theta_{dt} + \epsilon_{isdt} \quad (2.1)$$

where Y_{isdt} is the standardized score in a given subject for student i , in school s , located in administrative district d , in the academic year t ; $Female_i$ takes the value 1 if the student is female and 0 otherwise; $External_s$ takes value 1 if the student attended a school that was assigned to administer the test externally, and 0 otherwise; θ_{td} are year-administrative district fixed effects; and ϵ_{isdt} is the error term. The standard errors are clustered at the school level (the treatment unit).

The OLS estimate of parameter δ captures the differences between the gender gap in student performance for students attending a school where the test was externally administered (treatment group) compared to students attending a school where the test was internally administered (control group). δ takes a negative (positive) value if girls perform worse (better) than boys as a result of being exposed to an externally administered test. Due to the sampling frame and the nature of the intervention, we estimate average treatment effects, rather than intention to treat.

Internal Validity of the Empirical Design. Our identification strategy relies on the random assignment of schools to either externally or internally administered tests. Since randomization was done at the administrative unit level, we present the balance of these characteristics. To test whether the randomization was well

executed, we compare whether school baseline characteristics are balanced between treatment (schools with external testing) and control (schools with internal testing) groups. Finding that some groups of schools are more likely to sort on the externally administered test group would indicate endogenous sample selection. Linear regressions are performed using each of the school observable characteristics (i.e., school type, number of students, ESEC socioeconomic index, the fraction of non-Spanish parents, etc.) as the dependent variable. Table 2.A.2 shows the balancing tests by grade. Four out of eighteen contrasts are significant at the 10% level in 6th and 10th Grade. In 6th Grade, the number of public schools is lower and the socioeconomic status of their students (measured as the ESEC index) is higher in schools with external testing. However, other variables that are correlated with the student socioeconomic status, such as the fraction of non-Spanish parents, students that started school after three years old, and students that have less than 50 books at home, are not statistically significantly different between the treatment and control group. In 10th Grade, the number of semi-public schools is lower and the number of students is higher in schools with external testing. We also test for the joint significance of the observable characteristics for treatment and control groups. We find that none of the F-statistics is statistically significant at the 10% level. In addition, Table 2.B.1, Table ?? and Table ?? present the balancing tests by grade and academic year. Table 2.B.1 shows that several baseline school characteristics are statistically significantly different for West DAT for 6th Grade in 2016/2017, which may indicate that randomization was not particularly well executed in this unit and year. We thus exclude DAT West for 6th Grade in 2016/2017, since including it may bias the estimates due to significant differences between treatment and control group.

2.5 Results

2.5.1 Baseline Results

Table 2.A.3 presents the results from estimating Equation (1) separately for each subject. The coefficient on the interaction of the female dummy and externally

administered testing is negative, indicating that females perform worse than boys under externally administered testing environments. In terms of magnitude, Panel A shows that in Grade 6 the size of the coefficient on the interaction between the female dummy and the external testing dummy is very similar across subjects, at about 0.04 of a standard deviation. Panel B shows that in Grade 10th the size of the coefficient on the interaction between the female dummy and the external dummy varies across subjects, being greatest for Mathematics and Social Sciences, and losing statistical significant (although keeping the negative sign) in Spanish and English.

An interesting observation from the results in Table 2.A.3 is that whereas the size of the coefficient on the interaction between the female and the externally administered school dummies is the same across subjects, the relative size with respect to the coefficient on the female dummy is much larger in Mathematics, Sciences, and Social Sciences than in languages subjects, which are traditionally male-dominated. This is consistent with the literature showing that women under-perform when the task used is perceived as favoring men (Iriberry and Rey-Biel, 2017). A closer look at the estimates in Table 2.A.3 reveals that the attainment gap between male and female students widens in subjects where males outperform females such as Mathematics in both grades, and Social Sciences in 10th Grade. In particular, boys in 6th Grade outperform girls by 0.10 of a standard deviation under examinations with internal invigilation, so that the gender difference increases by 50 percent in schools where exams took place under external invigilation. Conversely, girls outperform boys in the rest of the subjects, particularly in Spanish and English, by 0.38 and 0.23 standard deviations respectively. This gender difference is narrowed by 13 percent in the case of Spanish and 18 percent in the case of English when tests are externally administered. In Science, where the difference between boys and girls is 0.04 standard deviations in schools under internal invigilation regimes, the gender gap in attainment is reduced by 53 percent as a result of having an external invigilator in the exam. Similarly, the difference between boys' and girls' attainment scores in Mathematics and Social Sciences increases by more than 50 percent for students in 10th Grade.

Table 2.A.4 shows that these results are mostly driven by girls performing worse

in exams with external inspection, rather than by boys performing better in exams with internally administered tests in Mathematics. It seems that both females and males decrease their relative performance in comparison to students of the same gender students in internally administered tests in the rest of subjects, but the relative reduction in performance is higher for females than for males.

To provide further evidence on the causal effect of having an external testing on girls' test scores, we perform a test for randomization inference. We perform a permutation test by generating a random allocation of internal or external testing to schools, and we effectively construct a “false” external dummy variable based on this randomization. We then re-estimate our benchmark model, Equation (1), using the “false” external dummy variable, and store the estimates. We repeat the exercise 1,000 times. The empirical density function and density of the estimated coefficients on external invigilators are shown in Figure 2.A.3 and Figure 2.A.4. The distribution of the estimated coefficients on the placebo external variable is centered around zero, as expected, and our benchmark estimates indicated by the red vertical line lies outside the range of coefficients estimated in our simulation exercise.

2.5.2 Robustness Checks and Heterogeneity Analysis

We next check whether our results are robust to using different samples. Table 2.A.5 shows that baseline results remain virtually the same when including the sample of students in schools where no randomization took place, in which all performed the test under external testing. Adding these observations increases the number of students to 112,299 and the sample of schools to 1,855. Table 2.A.6 shows that the main results remain unchanged when adding school fixed effects in addition to administrative district dummies interacted with year dummies. Table 2.A.7 estimates a version of Equation 1 that adds school controls. In particular, we control for several school observable characteristics, such as whether the school is public, the fraction of students that repeat a year, and the fraction of students whose parents are not Spanish.¹¹ The results remain qualitatively the same, although they are less efficiently

¹¹The construction and summary statistics of these variables are shown in Table 2.B.2

estimated.¹²

Table 2.A.8 and Table 2.A.9 analyze heterogeneity by school's ownership (public versus semi-public and private schools) and socio-economic characteristics respectively, to see which girls are more affected across the socio-economic spectrum. Results from Table 2.A.8 shows that both the gender gap under internal testing and its reduction under external testing remains qualitatively the same as in the baseline estimates. Girls' lower relative performance in externally administered testing is not statistically significantly different by public and semi-public and private schools in most specifications. We proxy socio-economic status using school characteristics on parental education. In particular, we estimate Equation (1) for a sample of students attending schools where the fraction of students whose father have a college degree is below the median, and compare these estimates to those where the fraction of students whose fathers have a college degree is above the median.

Results from Table 2.A.9 show that the female dummy remains qualitatively the same as in the benchmark estimates of Table 2.A.3, with girls over-performing boys in all subjects but Mathematics in 6th Grade, and in Spanish and English in 10th Grade. The point estimates on the interaction between the female dummy and the external dummy remains negative in most specifications as before. The effects seem to be mostly driven by students in schools with lower socio-economic status for 6th-graders, although the difference across socio-economic school status is not statistically different in most specifications. In 10th Grade, the reduction in the gender gap on Mathematics seems to be driven by students in schools with higher socioeconomic status.

2.5.3 Ruling out alternative explanations

Results so far point to girls underperforming relative to boys in externally administered testing environments. Here we explore whether other explanations may be driven the gender differences in scores previously documented.

¹²The sample size drops in 3,858 (1,707) observations due to the lack of information on several school characteristics in 16 (10) schools (high schools) out of 1,3000 (788).

Grade inflation. There is indeed evidence on grade inflation when internal testing was administered to students. [García-Brazales \(2019\)](#) finds that there is an average grade inflation in schools which undertook internal testing, but this is entirely driven by semi-public schools, finding no effects for public schools. Table [2.A.8](#) shows heterogeneity effects by school ownership. This Table presents that in both types of schools, we observe a negative coefficient on the interaction between female and external, indication that girls perform worse than boys under external testing regardless of the type of school. This result is even more significant in public schools, in which there is no evidence on grade inflation. Our identification relies on the double difference between internal vs. external testing and girls vs. boys. If there is grade inflation in internal schools and all grades are bumped up, it would not affect our identification since the effect is identified by the difference between girls and boys. Our identification relies on the assumption that there is no differential grade inflation across student gender.

Surprise Effects on the first year of implementation. It may be the case that the first year in which the policy intervention took place (2016/2017), a surprise effect may arise. This may be the case if in the first year of the policy schools that experienced internal examination for the first time also experienced other factors in the organization of the testing that could have affected the relative performance of girls. In this case, we might observe an effect only on the first year of the policy implementation and a zero effect thereafter, when students adjusted to the new situation. Results from Table [2.A.10](#) suggests that this was not the case. When testing for differential effects by year, we observe a negative coefficient on the female times' external dummy in both years (albeit the coefficients are less precisely estimated since the sample is split into two parts).

Gender-biased marking. A possible explanation to our findings is that the grading in school where the test was internally administered might be gender-biased in favor of girls. Several studies have found that girls are treated more favourably by teachers when marking (see e.g., [Lavy \(2008\)](#); [Cornwell, Mustard and Van Parys \(2013\)](#)), particularly in Mathematics exams [Terrier \(2016\)](#)). Gender-biased marking

can then be a confounding factor if finding that teachers grade girls more favorably than boys in internal relative to external testing.

This alternative explanation seems unlikely for two reasons. First, tests were blind. Markers had only access to the student's encrypted identifier but not to any other personal information, such as student's name or gender. Second, the majority of the tests were multiple-choice questions. However, in order to test whether markers may use the students' handwriting to guess whether an exam is filled out by a boy or a girl could explain our results, we look at the differential effect from having a test externally administered by comparing the scores from the part involving multiple-choice test and the part involving writing.

To that end, we look at the score in each of the parts of the test, i.e., the multiple-choice question part and the non-multiple-choice question part involving some writing. Table 2.A.11 shows that the main results are robust to considering only the multiple-choice part of the test. Interestingly, there is no statistically significant impact on external testing when we restrict to the non-multiple-choice part for English and Spanish, although there is one in Social Sciences. These results may suggest that either the widening of the gender gap in scores as a result from having an externally administered test may be different when performing a multiple-choice versus a non-multiple-choice question, or, alternatively, that there is gender bias in non-multiple-choice marking that offsets the gender gap increase in an external testing environment as opposed to an internal testing, at least for languages.

Student's perception of the teacher's biases. Another potential explanation of our findings is that students may exert different levels of effort depending on who is their main teacher. If students have the perception that their teacher may be able to access the information of their final grade in the standardized test, they may change their behavior according to it. It is more likely that students under internally administered testing may think this way, as the exams are marked by teachers from the same school. There is also evidence that there are gender differences with respect to perceptions regarding marking discrimination by gender. [Ouazad \(2013\)](#) find evidence Using a large-scale experiment that students' choices reflect the perception

of biases in teachers' grading practices. In particular, boys invest less with female teachers, and girls invest more with male teachers. We perform a heterogeneity analysis to examine whether the effects are different depending on the observable characteristics of the main teacher of the student. Identifying the characteristics of each student's teacher, we can match each student with her teacher. Table 2.A.12, Table 2.B.4 and Table 2.B.5 show that the results are not significantly different across teacher gender, experience and age respectively.

Potential Unobserved Factors. There are two additional factors that may differentially affect students' performance under external and internal testing. First, schools under internal testing may have provided a higher flexibility in the duration of the test relative to those of external testing. An alternative hypothesis is that the schools where the exam was internally administered provided more time to perform the test in comparison to those externally administered. Shurchkov (2012) finds that additional time is better exploited by female students relative to males, which provides evidence of a potential differential effect by gender. However, the exam rules state that the test duration cannot be modified, and these rules were monitored in schools undergoing internal testing by independent inspectors from the Madrid Educational Authority. Another possible explanation to our findings could be that girls in schools where the tests were internally administered got more help during the invigilation period than girls in externally administered ones. Notice that for this explanation to explain the heterogeneous increase in the gender gap in scores across subjects as a result of external testing girls would have needed to get help in some subjects and not in others, which seems unlikely.

2.6 Identifying the mechanisms through which the testing environment affects the gender gap in math

We have argued that having an externally administered test results in being exposed to a less familiar testing environment in comparison to having an internally administered test. The exposure to a non-familiar environment may lead to higher levels of pressure relative to a more familiar environment. This potential increase in the level of pressure may lead to differential responses depending on the student gender. In this section, we provide empirical evidence on how pressure may be driving the gender differences in performance documented so far.

This section exploits information from a student survey on student's attitudes and beliefs that were collected right after each of the exams in the second year of the evaluation (2017/2018). In particular, students were asked, for each subject and after each of the exams was finished, questions on whether they had felt stressed, they considered themselves good in the subject, how much effort they had put in preparing for the test, and how much effort they had put in the actual execution of the test were asked. We construct dummy variables with value 1 if the student answered the highest degree level of stress, the confidence level in the subject, and preparation of the test, on the first three questions, and zero otherwise. For a full description of the questions, see Table 2.B.3.

Figure 2.A.5 and Figure 2.A.6 show the average value of the responses to the four survey questions by gender, subject and grade. Panel A shows that girls in both grades overwhelmingly felt more stressed than boys. The subject that induces the highest level of stress is Mathematics for both boys and girls. However, the largest differences in stress between boys and girls are found in Mathematics, Science, and Social Sciences. Panel B shows that boys' self-assessed performance is higher than that of girls. The fraction of boys who consider themselves as very good in each subject is higher than girls, except for English in 6th Grade. Interestingly, the

highest value for boys is reported in Spanish for 6th graders, a subject in which girls overcome boys in about 0.38 standard deviation. However, the larger differences in the level of confidence between boys and girls are in Mathematics, Science, and Social Sciences. Panel C presents that girls prepared the test more intensively than boys on average. The fraction of girls who self-report a high preparation of the test in each subject is larger than for boys. However, the differences between boys and girls are not statistically different for Mathematics, Social and Spanish in 10th Grade. Panel D shows that girls are reporting higher levels of effort than boys in the exams of every subject. The lowest difference between boys' and girls' effort is in Mathematics.

It seems that there are substantial differences in the levels of stress, self-confidence, and effort exerted during the exam between boys and girls. Interestingly, these differences are more salient for those subjects in which the size of the effect of externally administered test is the highest with respect to its baseline mean. This may indicate a potential association between these variables and the gender differences under different testing environments. In order to further test this hypothesis, we estimate our main specification with each of the survey variables as dependent variables. Table 2.A.13 and Table 2.A.14 show the results for 6th and 10th Grade respectively. We find that the gender gap for students attending a school where the test was externally administered is higher for the levels of stress, lower for self-confidence, higher test preparation and larger effort exerted during the exam, in comparison to students attending a school where the test was internally administered. However, point estimates are small and not statistically significant for every variable except the self-reported effort during the exam. The non-significance of the estimates may potentially be due to the reduction in statistical power due to the smaller sample size relative to the baseline estimates, since the survey was only conducted in 2017/2018. Alternatively, the variables of stress, self-confidence and preparation, are categorical variables, and thus may potentially be not flexible enough to be able to detect an effect. We will have the opportunity to increase the number of observations when we receive the data of the next randomization wave performed in 2018/2019,

in which the survey questions were incorporated as well.

2.7 Conclusion

Using a randomized control trial implemented on the entire population of students in 6th and 10th grades in the Region of Madrid (Spain), we estimate the causal effect on student performance from taking a test under familiar versus non-familiar test-taking environments. Students are exposed to a non-familiar test environment when the Regional Government of Madrid externally administered the test, and they are exposed to a familiar environment when the test is internally administered by the school they attend.

We find that female students' performance declines when performing the test in a school with external testing in comparison to internal testing. This result is particularly strong in subjects that are traditionally male-dominated. We find suggestive evidence that gender differences in the levels of stress while taking the test, self-confidence, and effort exerted during the test may explain this difference in performance. We rule out alternative hypotheses such as gender-biased marking.

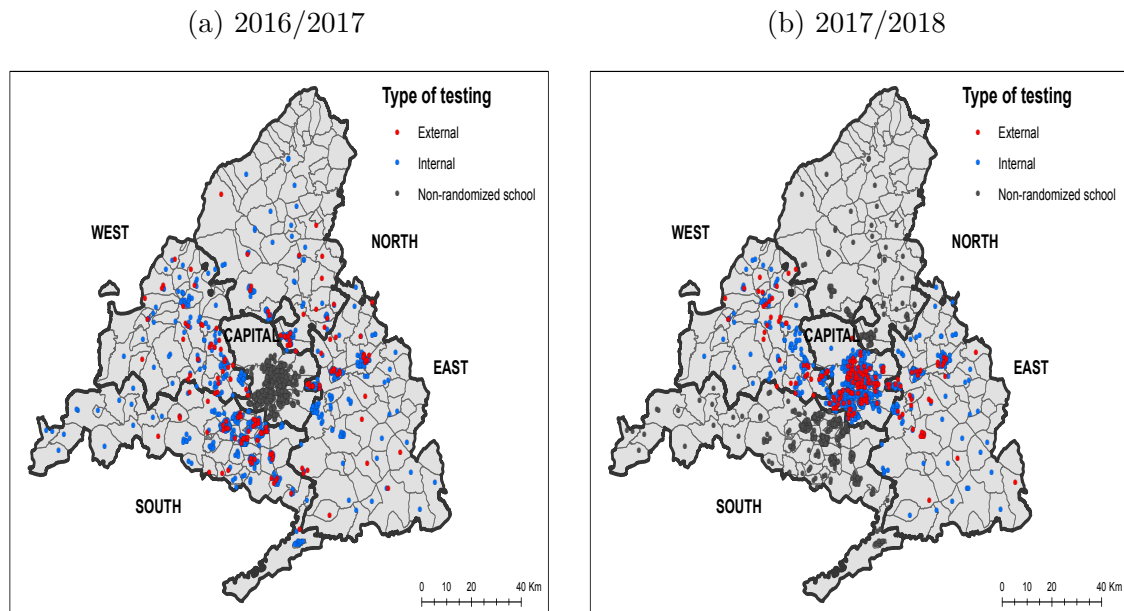
Empirical evidence has shown that the higher the stakes, the larger the increase in the level of pressure ([Azmat, Calsamiglia and Iriberry, 2016](#)). The estimates of this paper are based on a low stakes test. Therefore, it seems likely that the estimates from our paper are a lower bound of the potential effects in high stakes testing. Most of the education systems rely on high stakes external standardized testing to access different education tracks. We show that girls are more sensitive than boys to the testing environment (independently of their true level of ability), decreasing their performance when the latest is less familiar to students. Result imply that a non-familiar testing-environment in standardized test exacerbates the gender differences in performance, which may translate into differential career choice, and thus persistent gender inequalities over the earnings life-cycle.

Further research needs to explore potential heterogeneity in the invigilation of the test that may contribute to the observed gender differences in student performance.

It would be useful to understand whether the characteristics of the invigilator (e.g., gender, age, and degree of familiarity with the student) and the behavior of the invigilator influences student performance differently across genders. In addition, the characteristics of the testing-environment, such as the disposition of tables or size of classes, may also affect student performance differently for boys and girls. It is also worth investigating whether there are some subgroups of the population that may be potentially affected by the testing environment more than others. Answering these types of questions can help to improve the design of testing environments to elicit students' true level of ability better.

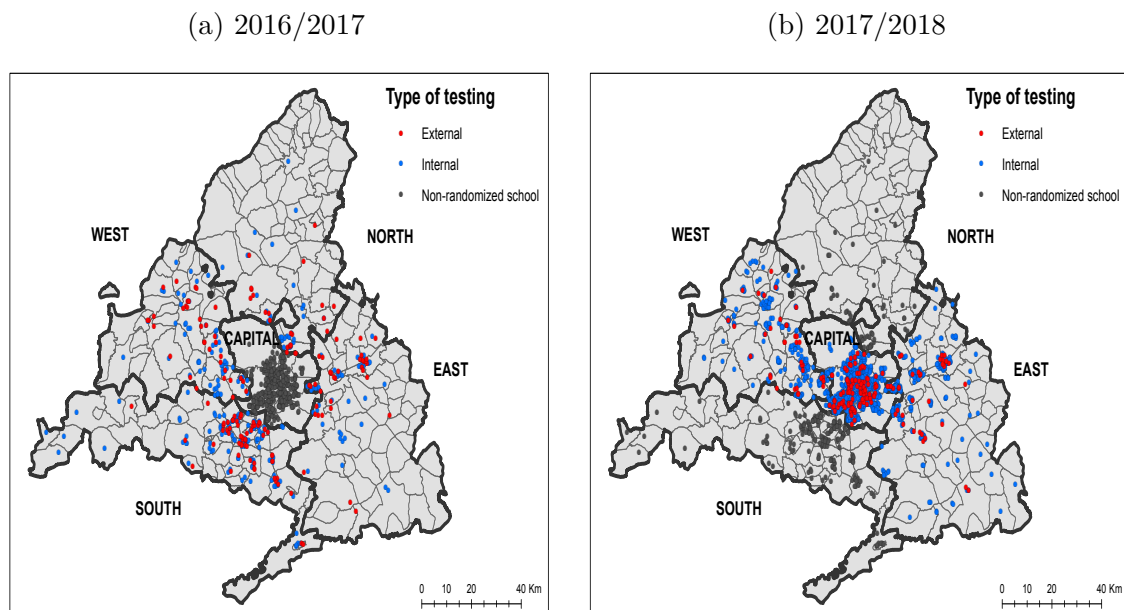
Appendix 2.A Main Figures and Tables

Figure 2.A.1: Distribution of internally vs. externally schools in 6th Grade.



Note: Red dots signal externally administered schools. Blue dots signal internally administered schools.

Figure 2.A.2: Distribution of internally vs. externally schools in 10th Grade.



Note: Red dots signal externally administered schools. Blue dots signal internally administered schools.

Figure 2.A.3: Permutation Test for 6th Grade Sudents.

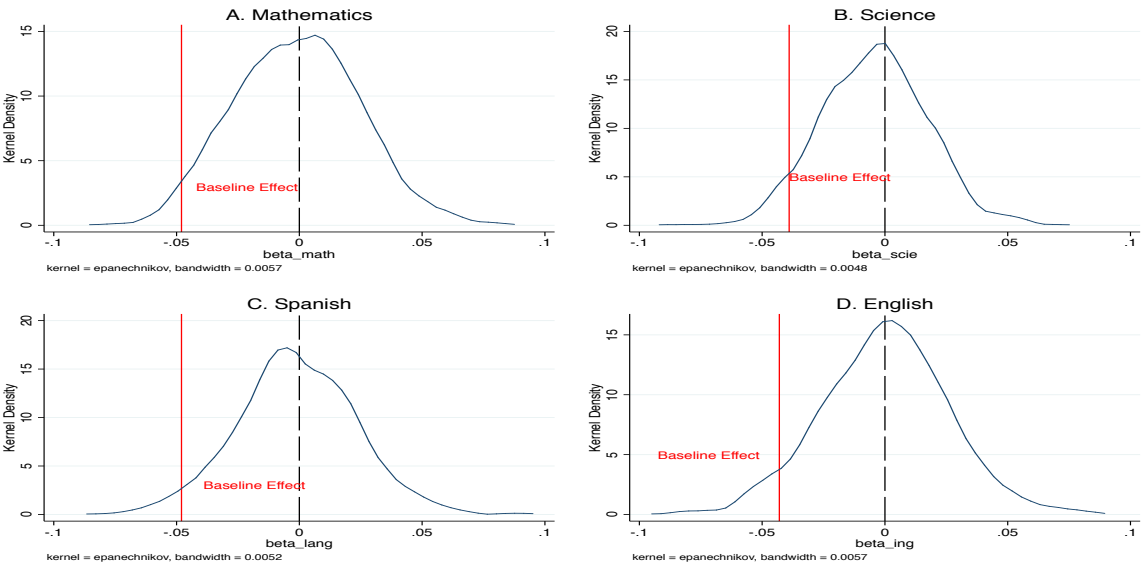


Figure 2.A.4: Permutation Test for 10th Grade Students.

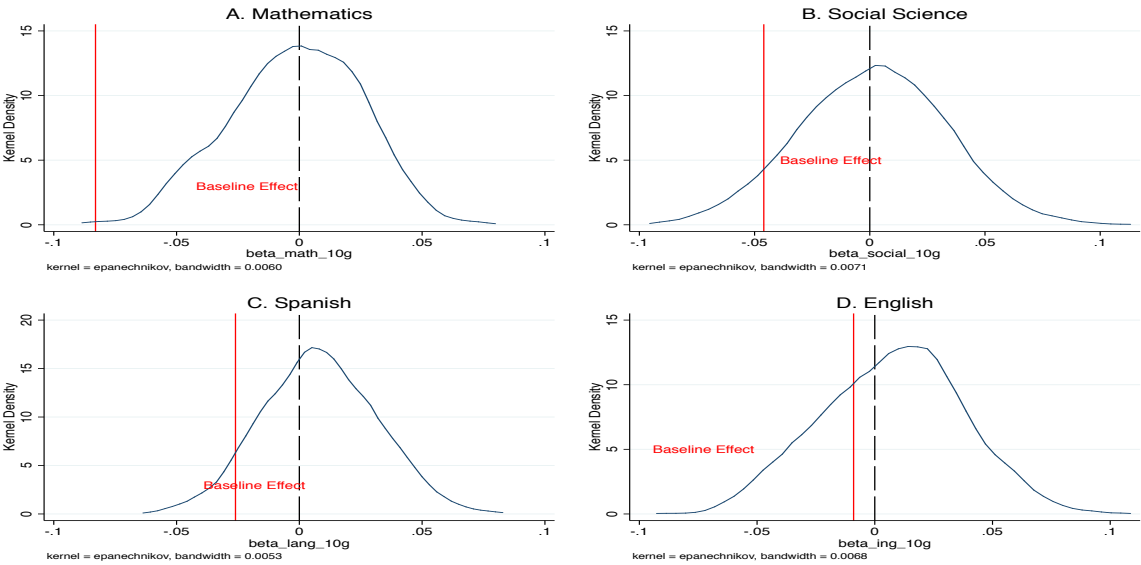
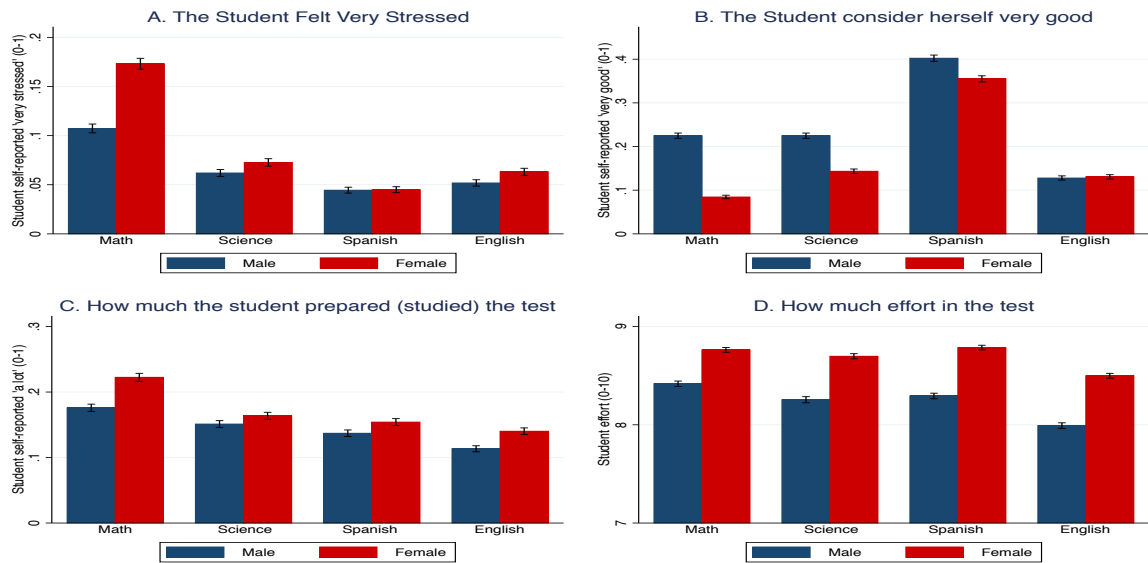
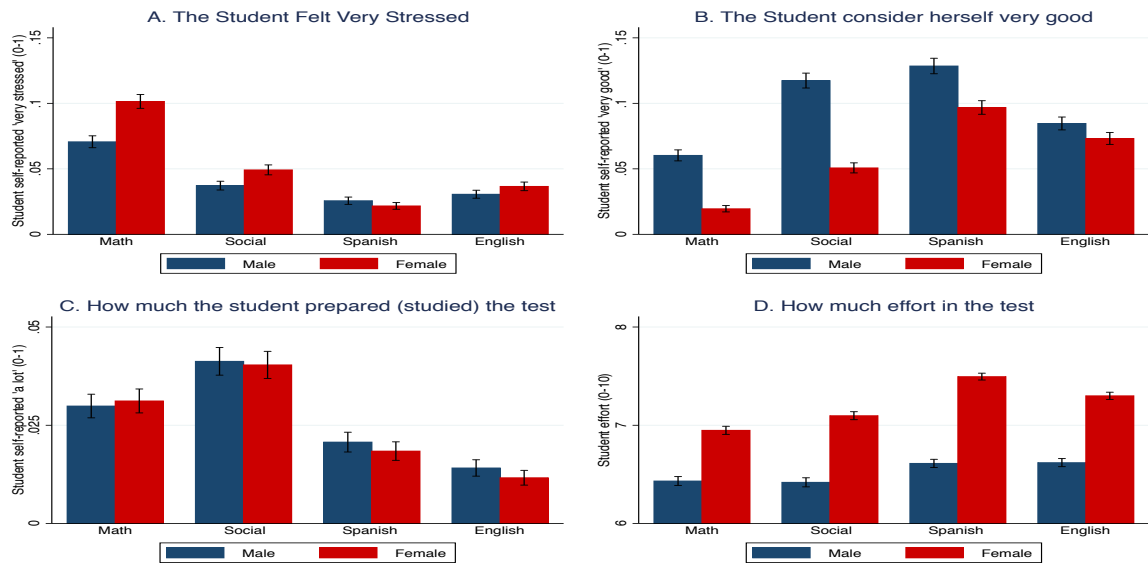


Figure 2.A.5: Survey Questions by gender and subject in 6th Grade.



Note: Blue and Red bars represent mean values. Horizontal lines depict confidence intervals at the 95% level of confidence.

Figure 2.A.6: Survey Questions by gender and subject in 10th Grade.



Note: Blue and Red bars represent mean values. Horizontal lines depict confidence intervals at the 95% level of confidence.

Table 2.A.1: Summary Statistics.

Grade	6th Grade				10th Grade			
Year	2016/2017		2017/2018		2016/2017		2017/2018	
Type of Testing	Internal	External	Internal	External	Internal	External	Internal	External
A. Distribution of Internally/Externally Administered test by Administrative District								
# Schools								
Capital	540	0	453	95	368	0	270	96
North	78	24	0	105	29	23	0	53
South	260	57	0	322	122	58	0	180
East	137	35	137	36	49	34	50	36
West	109	34	109	36	61	35	66	36
Total	1,124	150	699	594	629	150	386	401
# Students								
Capital	25,167	0	20,439	4,697	20,579	0	14,857	5,139
North	3,203	1,127	0	4,422	1,871	1,695	0	3,497
South	12,038	2,608	0	14,87	6,939	3,105	0	10,139
East	6,386	1,705	6,235	1,627	2,982	2,105	2,572	2,394
West	5,611	1,899	5,915	1,86	3,884	2,412	4,384	2,541
Total	52,405	7,339	32,589	27,476	36,255	9,317	21,813	23,71
B. Baseline Sample: Excluding districts with no randomization								
#Schools	584	150	699	167	261	150	386	168
Frac. Schools	0.79	0.21	0.81	0.19	0.64	0.36	0.7	0.3
# Students	27,238	7,339	32,589	8,184	15,676	9,317	21,813	10,074
Frac. Students	0.79	0.21	0.8	0.2	0.63	0.37	0.68	0.32
Frac. Female	0.49	0.48	0.49	0.49	0.49	0.5	0.51	0.48

Notes: Frac. Schools refer to the fraction of schools that are either internally or externally administered in each academic year. Frac. Students refer to the fraction of students who perform the test in either internally or externally administered school in each academic year.

Table 2.A.2: Treatment-Control balance in fixed school characteristics in 6th and 10th Grade.

	Internally (C) Administered (1)	Externally (T) Administered (2)	Difference T – C (3)	2 > <i>p</i> -value (4)
Panel A. 6th Grade				
Public School	0.558	0.502	-0.056	0.063
Semi-Public School	0.266	0.292	0.027	0.347
Private School	0.177	0.206	0.029	0.132
Bilingual School	0.399	0.380	-0.020	0.498
Number of Students	56.808	57.621	0.813	0.640
ESEC Index	0.218	0.294	0.076	0.048
Frac. Students Started After 3yo	0.028	0.029	0.001	0.887
Frac. Non-Spanish Parents	0.135	0.113	-0.022	0.110
Frac. Less than 50 Books at Home	0.183	0.163	-0.019	0.165
N	1,206	286		
Test of joint significance	<i>F</i> -stat: 1.077 (<i>p</i> -value: > 0.376)			
Panel B. 10th Grade				
Public School	0.360	0.413	0.053	0.100
Semi-Public School	0.387	0.329	-0.058	0.068
Private School	0.253	0.258	0.005	0.814
Bilingual School	0.187	0.191	0.004	0.914
Number of Students	73.171	77.954	4.783	0.088
ESEC Index	0.710	0.794	0.084	0.341
Frac. Students Started After 3yo	0.050	0.014	-0.036	0.934
Frac. Non-Spanish Parents	-0.042	0.287	0.329	0.740
Frac. Less than 50 Books at Home	-0.172	0.627	0.800	0.422
N	669	322		
Test of joint significance	<i>F</i> -stat: 1.398 (<i>p</i> -value: > 0.195)			

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.3: Baseline Estimates.

Dependent Variable: Score in:				
A. 6th Grade	Mathematics (1)	Science (2)	Spanish (3)	English (4)
Female*External	-0.048** (0.022)	-0.039* (0.022)	-0.048** (0.022)	-0.043* (0.024)
Female	-0.096*** (0.012)	0.068*** (0.011)	0.380*** (0.012)	0.238*** (0.012)
External	0.033 (0.036)	-0.063* (0.033)	-0.106*** (0.035)	-0.058 (0.042)
Constant	0.197*** (0.046)	0.143*** (0.043)	0.077* (0.041)	0.297*** (0.049)
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	67,841	67,841	67,841	67,841
B. 10th Grade	Mathematics (1)	Social (2)	Spanish (3)	English (4)
Female*External	-0.083*** (0.022)	-0.046* (0.024)	-0.026 (0.020)	-0.009 (0.024)
Female	-0.161*** (0.014)	-0.074*** (0.017)	0.151*** (0.015)	0.114*** (0.016)
External	0.013 (0.032)	-0.096*** (0.035)	-0.001 (0.029)	-0.037 (0.036)
Constant	0.068 (0.057)	0.032 (0.065)	-0.173*** (0.054)	-0.166*** (0.055)
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	56,880	56,880	56,880	56,880

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.4: Average student performance by gender and internally/externally administered test.

Dependent Variable: Score in:								
A. 6th Grade	Mathematics		Science		Spanish		English	
	Female	Male	Female	Male	Female	Male	Female	Male
External	-0.015 (0.035)	0.033 (0.036)	-0.103*** (0.031)	-0.065** (0.033)	-0.153*** (0.032)	-0.105*** (0.035)	-0.101** (0.043)	-0.058 (0.042)
Constant	0.018 (0.073)	-0.040 (0.047)	0.038 (0.057)	-0.082** (0.039)	0.414*** (0.059)	-0.174*** (0.039)	0.370*** (0.071)	-0.143** (0.060)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37,101	38,249	37,101	38,249	37,101	38,249	37,101	38,249
B. 10th Grade	Mathematics		Social		Spanish		English	
	Female	Male	Female	Male	Female	Male	Female	Male
External	-0.069** (0.031)	0.014 (0.032)	-0.143*** (0.038)	-0.095*** (0.035)	-0.033 (0.030)	0.003 (0.029)	-0.050 (0.039)	-0.033 (0.036)
Constant	0.039 (0.059)	0.206*** (0.056)	0.053 (0.069)	0.075 (0.072)	0.245*** (0.050)	0.055 (0.056)	0.392*** (0.090)	0.194** (0.082)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	28,393	28,487	28,393	28,487	28,393	28,487	28,393	28,487

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.5: Robustness Checks I. Baseline Estimates including all DATs.

Dependent Variable: Score in:				
A. 6th Grade	Mathematics (1)	Science (2)	Spanish (3)	English (4)
Female*External	-0.047*** (0.014)	-0.039*** (0.014)	-0.024* (0.014)	-0.024* (0.014)
Female	-0.096*** (0.012)	0.068*** (0.011)	0.380*** (0.012)	0.238*** (0.012)
External	0.033 (0.035)	-0.063** (0.031)	-0.117*** (0.033)	-0.068 (0.042)
Constant	0.127** (0.060)	-0.066 (0.050)	-0.117* (0.060)	0.059 (0.081)
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	112,299	112,299	112,299	112,299
B. 10th Grade	Mathematics (1)	Social (2)	Spanish (3)	English (4)
Female*External	-0.086*** (0.016)	-0.030* (0.017)	-0.006 (0.016)	-0.001 (0.016)
Female	-0.161*** (0.014)	-0.074*** (0.017)	0.151*** (0.015)	0.114*** (0.016)
External	0.015 (0.031)	-0.104*** (0.035)	-0.011 (0.029)	-0.041 (0.036)
Constant	0.203*** (0.055)	0.101 (0.065)	0.074 (0.051)	0.235*** (0.083)
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	91,095	91,095	91,095	91,095

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.6: Robustness Checks II. Baseline Estimates including School Fixed Effects.

Dependent Variable: Score in:				
A. 6th Grade	Mathematics (1)	Science (2)	Spanish (3)	English (4)
Female*External	-0.042*** (0.012)	-0.031** (0.013)	-0.029** (0.014)	-0.017 (0.012)
Female	-0.103*** (0.009)	0.059*** (0.009)	0.387*** (0.010)	0.230*** (0.009)
External	-0.023 (0.031)	-0.103*** (0.032)	-0.168*** (0.040)	-0.090*** (0.032)
Constant	0.059*** (0.018)	0.019 (0.017)	-0.122*** (0.021)	-0.114*** (0.017)
Year FE	Yes	Yes	Yes	Yes
# Schools	1,152	1,152	1,152	1,152
Observations	104,524	104,524	104,524	104,524
B. 10th Grade	Mathematics (1)	Social (2)	Spanish (3)	English (4)
Female*External	-0.080*** (0.014)	-0.030** (0.014)	-0.003 (0.013)	-0.007 (0.012)
Female	-0.163*** (0.010)	-0.081*** (0.011)	0.141*** (0.011)	0.115*** (0.010)
External	0.010 (0.025)	-0.140*** (0.031)	-0.019 (0.025)	-0.046** (0.021)
Constant	0.132*** (0.015)	0.136*** (0.018)	-0.043*** (0.015)	0.004 (0.013)
Year FE	Yes	Yes	Yes	Yes
# Schools	798	798	798	798
Observations	91,095	91,095	91,095	91,095

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.7: Robustness Checks III.Baseline Estimates including school controls.

Dependent Variable: Score in:				
A. 6th Grade	Mathematics (1)	Science (2)	Spanish (3)	English (4)
Female*External	-0.042** (0.020)	-0.028 (0.021)	-0.044** (0.021)	-0.037* (0.021)
Female	-0.098*** (0.010)	0.067*** (0.010)	0.377*** (0.010)	0.235*** (0.009)
External	-0.009 (0.032)	-0.107*** (0.029)	-0.146*** (0.032)	-0.091*** (0.030)
Constant	0.486*** (0.087)	0.356*** (0.081)	0.276*** (0.065)	0.472*** (0.069)
School Controls	Yes	Yes	Yes	Yes
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	55,173	55,173	55,173	55,173
B. 10th Grade	Mathematics (1)	Social (2)	Spanish (3)	English (4)
Female*External	-0.077*** (0.019)	-0.036* (0.021)	-0.016 (0.018)	0.005 (0.018)
Female	-0.160*** (0.012)	-0.073*** (0.014)	0.153*** (0.012)	0.119*** (0.012)
External	0.027 (0.027)	-0.078*** (0.030)	0.006 (0.025)	-0.028 (0.027)
Constant	0.624*** (0.098)	0.598*** (0.091)	0.346*** (0.076)	0.465*** (0.074)
School Controls	Yes	Yes	Yes	Yes
DAT*Year FE	Yes	Yes	Yes	Yes
Observations	71,492	71,492	71,492	71,492

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.8: Heterogeneous Effects by School Ownership.

Dependent Variable: Score in:								
A. 6th Grade	Public School				Semi-Public or Private School			
	Mathematics (1)	Science (2)	Spanish (3)	English (4)	Mathematics (5)	Science (6)	Spanish (7)	English (8)
Female*External	-0.067** (0.027)	-0.031 (0.028)	-0.051** (0.026)	-0.043 (0.029)	-0.025 (0.029)	-0.052* (0.028)	-0.042 (0.032)	-0.027 (0.034)
Female	-0.105*** (0.012)	0.048*** (0.011)	0.393*** (0.013)	0.234*** (0.012)	-0.099*** (0.019)	0.069*** (0.016)	0.366*** (0.017)	0.242*** (0.019)
External	0.084** (0.041)	-0.009 (0.035)	0.004 (0.043)	-0.046 (0.052)	-0.098* (0.051)	-0.157*** (0.048)	-0.212*** (0.049)	-0.119** (0.057)
Constant	-0.246*** (0.055)	-0.224*** (0.046)	-0.364*** (0.057)	-0.271*** (0.067)	0.293*** (0.066)	0.226*** (0.070)	0.007 (0.062)	0.000 (0.055)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,479	40,480	40,481	40,482	34,869	34,870	34,871	34,872
B. 10th Grade	Public School				Semi-Public or Private School			
	Mathematics (9)	Science (10)	Spanish (11)	English (12)	Mathematics (13)	Science (14)	Spanish (15)	English (16)
Female*External	-0.035 (0.025)	-0.050* (0.028)	-0.032 (0.027)	0.022 (0.028)	-0.126*** (0.034)	-0.045 (0.035)	-0.027 (0.028)	-0.042 (0.035)
Female	-0.187*** (0.016)	-0.067*** (0.019)	0.168*** (0.019)	0.093*** (0.017)	-0.142*** (0.019)	-0.079*** (0.023)	0.139*** (0.019)	0.133*** (0.023)
External	0.076** (0.032)	-0.038 (0.042)	0.112*** (0.042)	0.028 (0.047)	-0.010 (0.045)	-0.107** (0.047)	-0.079** (0.037)	-0.062 (0.048)
Constant	-0.105* (0.059)	-0.158** (0.077)	-0.167** (0.077)	-0.169** (0.081)	0.347*** (0.059)	0.386*** (0.053)	0.132*** (0.048)	0.122** (0.049)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	26,187	26,188	26,189	26,190	30,693	30,694	30,695	30,696

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.9: Heterogeneous Effects by School's Fraction of Parents with College Education.

Dependent Variable: Score in:								
A. 6th Grade	Above Median Frac. College				Below Median Frac. College			
	Mathematics (1)	Science (2)	Spanish (3)	English (4)	Mathematics (5)	Science (6)	Spanish (7)	English (8)
Female*External	-0.031 (0.030)	-0.044 (0.028)	-0.026 (0.033)	-0.025 (0.031)	-0.066** (0.031)	-0.032 (0.032)	-0.072*** (0.027)	-0.067** (0.033)
Female	-0.116*** (0.017)	0.051*** (0.015)	0.362*** (0.017)	0.248*** (0.015)	-0.077*** (0.013)	0.085*** (0.013)	0.397*** (0.014)	0.227*** (0.015)
External	0.004 (0.050)	-0.074* (0.044)	-0.198*** (0.049)	-0.071 (0.053)	0.045 (0.043)	-0.065 (0.042)	-0.026 (0.044)	-0.072 (0.050)
Constant	0.266*** (0.086)	0.109 (0.067)	0.179*** (0.064)	0.319*** (0.077)	-0.205** (0.080)	-0.193** (0.084)	-0.369*** (0.055)	-0.365*** (0.107)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,560	33,560	33,560	33,560	34,280	34,280	34,280	31,377
B. 10th Grade	Above Median Frac. College				Below Median Frac. College			
	Mathematics (9)	Science (10)	Spanish (11)	English (12)	Mathematics (13)	Science (14)	Spanish (15)	English (16)
Female*External	-0.117*** (0.030)	-0.026 (0.034)	-0.017 (0.027)	-0.031 (0.031)	-0.035 (0.026)	-0.055** (0.026)	-0.024 (0.026)	0.033 (0.027)
Female	-0.132*** (0.021)	-0.071*** (0.024)	0.164*** (0.020)	0.162*** (0.022)	-0.191*** (0.017)	-0.071*** (0.019)	0.141*** (0.018)	0.071*** (0.016)
External	0.037 (0.045)	-0.065 (0.048)	-0.026 (0.039)	-0.069 (0.046)	-0.023 (0.036)	-0.144*** (0.041)	0.006 (0.035)	-0.020 (0.043)
Constant	0.310*** (0.062)	0.354*** (0.066)	0.157** (0.061)	0.275*** (0.055)	0.065 (0.059)	-0.034 (0.069)	-0.043 (0.059)	-0.327*** (0.060)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30,691	30,691	30,691	30,691	26,189	26,189	26,189	26,189

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.10: Heterogeneous Effects by Year.

Dependent Variable: Score in:								
A. 6th Grade	Year 2016				Year 2017			
	Mathematics (1)	Science (2)	Spanish (3)	English (4)	Mathematics (5)	Science (6)	Spanish (7)	English (8)
Female*External	-0.098*** (0.032)	-0.064** (0.030)	-0.045 (0.030)	-0.047 (0.033)	-0.016 (0.029)	-0.020 (0.029)	-0.044 (0.031)	-0.036 (0.034)
Female	-0.116*** (0.015)	-0.004 (0.014)	0.425*** (0.016)	0.250*** (0.016)	-0.084*** (0.017)	0.117*** (0.015)	0.351*** (0.016)	0.230*** (0.016)
External	0.002 (0.046)	0.005 (0.044)	0.042 (0.046)	0.059 (0.058)	0.054 (0.053)	-0.112** (0.046)	-0.205*** (0.048)	-0.138** (0.062)
Constant	0.155** (0.072)	0.042 (0.057)	-0.047 (0.056)	0.100 (0.075)	-0.016 (0.048)	-0.056 (0.039)	-0.125*** (0.039)	-0.104* (0.060)
DAT FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	27,066	27,067	27,068	27,069	40,774	40,775	40,774	40,775
B. 10th Grade	Year 2016				Year 2017			
	Mathematics (9)	Science (10)	Spanish (11)	English (12)	Mathematics (13)	Science (14)	Spanish (15)	English (16)
Female*External	-0.070** (0.031)	0.006 (0.035)	0.027 (0.025)	0.070** (0.033)	-0.088** (0.039)	-0.094** (0.039)	-0.079** (0.032)	-0.084** (0.041)
Female	-0.192*** (0.023)	-0.085*** (0.024)	0.154*** (0.020)	0.099*** (0.024)	-0.138*** (0.017)	-0.067*** (0.022)	0.148*** (0.020)	0.124*** (0.021)
External	-0.007 (0.048)	-0.091* (0.054)	0.028 (0.042)	0.027 (0.055)	0.029 (0.048)	-0.099** (0.049)	-0.025 (0.042)	-0.089* (0.053)
Constant	0.098 (0.085)	-0.048 (0.085)	-0.106 (0.077)	-0.197*** (0.069)	0.211*** (0.039)	0.261*** (0.046)	0.123*** (0.040)	0.403*** (0.053)
DAT FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,993	24,994	24,995	24,995	31,887	31,887	31,888	31,889

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.11: Questions Analysis: Multiple-choice vs. Non-multiple-choice.

Dependent Variable: Score in:							
A. 6th Grade	Mathematics	Science	Spanish		English		
	Multiple Choice	Multiple Choice	Multiple Choice	Non-Multiple Choice	Multiple Choice	Non-Multiple Choice	
	(1)	(2)	(3)	(4)	(5)	(6)	
Female*External	-0.049** (0.022)	-0.037* (0.021)	-0.030 (0.022)	0.013 (0.023)	-0.012 (0.023)	-0.023 (0.024)	
Female	-0.101*** (0.012)	0.104*** (0.011)	0.226*** (0.011)	0.397*** (0.011)	0.202*** (0.011)	0.223*** (0.012)	
External	-0.119*** (0.036)	-0.104*** (0.032)	-0.008 (0.030)	-0.266*** (0.041)	-0.037 (0.040)	-0.215*** (0.043)	
Constant	0.117* (0.061)	-0.051 (0.044)	-0.050 (0.040)	0.085* (0.044)	0.043 (0.064)	0.313*** (0.048)	
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	
Observations	67,141	67,141	67,141	67,141	67,141	67,141	
B. 10th Grade	Mathematics	Social Sciences		Spanish		English	
	Multiple Choice	Multiple Choice	Non-Multiple Choice	Multiple Choice	Non-Multiple Choice	Multiple Choice	Non-Multiple Choice
	(7)	(8)		(9)	(10)	(11)	(12)
Female*External	-0.063*** (0.022)	-0.038* (0.021)	-0.044* (0.023)	-0.025 (0.020)	-0.022 (0.021)	0.004 (0.023)	-0.006 (0.023)
Female	-0.182*** (0.015)	-0.045*** (0.015)	-0.125*** (0.018)	0.107*** (0.014)	0.221*** (0.015)	-0.021 (0.014)	0.148*** (0.015)
External	-0.019 (0.032)	-0.013 (0.031)	-0.257*** (0.034)	0.031 (0.030)	-0.142*** (0.031)	0.003 (0.033)	-0.078** (0.037)
Constant	0.031 (0.061)	-0.046 (0.054)	0.156** (0.061)	-0.167*** (0.049)	-0.125** (0.054)	-0.102* (0.052)	-0.155*** (0.054)
DAT*Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	56,862	56,862	56,862	56,862	56,862	56,862	56,862

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.12: Heterogeneous Effects by Main Teacher Gender.

Variables	Female Teacher		Male Teacher	
Female*External	-0.083*	-0.083*	-0.087	-0.110**
	(0.045)	(0.044)	(0.060)	(0.055)
Female	-0.095***	-0.099***	-0.148***	-0.129***
	(0.021)	(0.020)	(0.034)	(0.030)
External	-0.015	-0.024	-0.088	-0.053
	(0.056)	(0.054)	(0.088)	(0.079)
Teacher Characteristics				
Age		-0.000		-0.009
		(0.006)		(0.011)
Civil servant		-0.257***		-0.480***
		(0.058)		(0.083)
Experience		0.013**		0.008
		(0.006)		(0.011)
Full time contract		0.058		-0.234
		(0.150)		(0.326)
Constant	-0.045	0.135	0.096	18.997
	(0.046)	(11.467)	(0.101)	(21.417)
Observations	15,489	15,489	8,066	8,066
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.				

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.13: Survey Answers Baseline Estimates for 6th Grade Students.

A. Student Self-Reported Stress					B. The Student consider herself good at this Subject				
	Mathematics (1)	Science (2)	Spanish (3)	English (4)		Mathematics (5)	Science (6)	Spanish (7)	English (8)
Female*External	0.001 (0.010)	-0.003 (0.007)	0.008 (0.005)	0.005 (0.006)	Female*External	0.002 (0.010)	-0.002 (0.011)	0.009 (0.013)	-0.009 (0.010)
Female	0.066*** (0.004)	0.011*** (0.003)	-0.001 (0.002)	0.010*** (0.003)	Female	-0.141*** (0.005)	-0.081*** (0.005)	-0.049*** (0.006)	0.005 (0.004)
External	0.008 (0.007)	-0.002 (0.005)	-0.011*** (0.004)	-0.001 (0.005)	External	-0.009 (0.008)	0.014 (0.010)	-0.007 (0.012)	0.007 (0.008)
Constant	0.108*** (0.004)	0.061*** (0.003)	0.046*** (0.002)	0.053*** (0.003)	Constant	0.226*** (0.004)	0.226*** (0.005)	0.399*** (0.006)	0.122*** (0.004)
DAT FE Observations	Yes 36,171	Yes 36,171	Yes 36,171	Yes 36,171	DAT FE Observations	Yes 36,171	Yes 36,171	Yes 36,171	Yes 36,171
C. How much the student prepared (studied) the test					D. How much attention/interest have the student exert in the test				
	Mathematics (9)	Science (10)	Spanish (11)	English (12)		Mathematics (13)	Science (14)	Spanish (15)	English (16)
Female*External	0.003 (0.012)	0.011 (0.010)	0.008 (0.010)	-0.000 (0.008)	Female*External	0.044 (0.028)	0.070** (0.030)	0.064** (0.031)	0.057* (0.030)
Female	0.046*** (0.005)	0.010** (0.005)	0.015*** (0.005)	0.027*** (0.004)	Female	0.188*** (0.013)	0.212*** (0.014)	0.261*** (0.014)	0.265*** (0.013)
External	0.018* (0.011)	-0.001 (0.010)	-0.001 (0.009)	0.000 (0.008)	External	0.034 (0.026)	0.013 (0.028)	-0.038 (0.030)	-0.013 (0.030)
Constant	0.171*** (0.005)	0.154*** (0.005)	0.140*** (0.005)	0.114*** (0.004)	Constant	-0.143*** (0.014)	-0.140*** (0.014)	-0.144*** (0.014)	-0.153*** (0.014)
DAT FE Observations	Yes 36,171	Yes 36,171	Yes 36,171	Yes 36,171	DAT FE Observations	Yes 35,522	Yes 35,182	Yes 35,797	Yes 35,899

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.A.14: Survey Answers Baseline Estimates for 10th Grade Students.

A. Student Self-Reported Stress					B. The Student consider herself good at this Subject				
	Mathematics (1)	Science (2)	Spanish (3)	English (4)		Mathematics (5)	Science (6)	Spanish (7)	English (8)
Female*External	-0.006 (0.009)	-0.003 (0.006)	0.006 (0.004)	0.003 (0.005)	Female*External	0.010* (0.006)	0.004 (0.008)	0.002 (0.009)	-0.001 (0.008)
Female	0.032*** (0.005)	0.013*** (0.003)	-0.006** (0.003)	0.005 (0.003)	Female	-0.044*** (0.003)	-0.068*** (0.005)	-0.032*** (0.005)	-0.011** (0.005)
External	-0.010* (0.006)	-0.008** (0.004)	-0.009*** (0.003)	-0.004 (0.003)	External	-0.007 (0.005)	0.004 (0.007)	-0.012 (0.008)	-0.000 (0.007)
Constant	0.079*** (0.004)	0.041*** (0.003)	0.029*** (0.002)	0.037*** (0.003)	Constant	0.062*** (0.003)	0.112*** (0.004)	0.124*** (0.005)	0.080*** (0.004)
DAT FE Observations	Yes 24,846	Yes 24,846	Yes 24,846	Yes 24,846	DAT FE Observations	Yes 24,846	Yes 24,846	Yes 24,846	Yes 24,846
C. How much the student prepared (studied) the test					D. How much attention/interest have the student exert in the test				
	Mathematics (9)	Science (10)	Spanish (11)	English (12)		Mathematics (13)	Science (14)	Spanish (15)	English (16)
Female*External	-0.006 (0.006)	-0.011 (0.007)	-0.004 (0.004)	-0.001 (0.003)	Female*External	-0.105*** (0.038)	-0.040 (0.034)	-0.039 (0.032)	-0.040 (0.032)
Female	0.003 (0.003)	0.004 (0.004)	-0.001 (0.002)	-0.002 (0.002)	Female	0.245*** (0.020)	0.291*** (0.019)	0.411*** (0.017)	0.318*** (0.017)
External	0.001 (0.006)	0.044*** (0.006)	0.009** (0.004)	-0.002 (0.002)	External	-0.012 (0.046)	-0.057 (0.041)	-0.043 (0.041)	0.015 (0.037)
Constant	0.031*** (0.003)	0.027*** (0.003)	0.018*** (0.002)	0.015*** (0.001)	Constant	-0.088*** (0.026)	-0.105*** (0.025)	-0.180*** (0.023)	-0.153*** (0.021)
DAT FE Observations	Yes 24,846	Yes 24,846	Yes 24,846	Yes 24,846	DAT FE Observations	Yes 24,212	Yes 23,459	Yes 24,568	Yes 24,566

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 2.B Additional Figures and Tables

The online appendix supplements the paper “The Gender Gap in Student Performance: The Role of the Testing Environment”. It presents details on balancing tests (section A), Description of schools characteristics variables (section B), Description of student survey questions (section C), Heterogeneous effects by teacher experience and age (section D).

2.B.1 Additional Balancing Tests

Table 2.B.1: Treatment-Control balance in fixed school characteristics in 6th Grade for 2016/2017 and 2017/2018.

	North 2016		South 2016		Capital 2017		All DATs 2016&2017		All DATs without West in 2016	
	Difference (1)	P-value (2)	Difference (3)	P-value (4)	Difference (5)	P-value (6)	Difference (7)	P-value (8)	Difference (9)	P-value (10)
Public School	-0.045	0.680	-0.021	0.742	-0.075	0.174	-0.050	0.080	-0.056	0.063
Semi-Public School	0.070	0.431	0.014	0.819	0.035	0.534	0.028	0.295	0.027	0.347
Private School	-0.025	0.770	0.007	0.815	0.039	0.241	0.022	0.247	0.029	0.132
Bilingual School	-0.020	0.852	-0.017	0.786	0.033	0.529	-0.026	0.352	-0.020	0.498
Number of Students	5.528	0.349	-1.690	0.547	2.962	0.396	0.986	0.570	0.813	0.640
ESEC Index	0.278	0.144	0.048	0.685	0.072	0.087	0.093	0.012	0.076	0.048
Frac. Students Started After 3yo	-0.015	0.119	-0.001	0.883	0.007	0.652	-0.000	0.955	0.001	0.887
Frac. Non-Spanish Parents	-0.044	0.266	-0.031	0.154	-0.023	0.433	-0.025	0.052	-0.022	0.110
Frac. Less than 50 Books at Home	-0.040	0.309	-0.015	0.564	-0.022	0.460	-0.023	0.070	-0.019	0.165
Test of joint significance	0.627		0.899		0.554		0.143		0.37	
	East				West					
	2016		2017		2016		2017			
	Difference (11)	P-value (12)	Difference (13)	P-value (14)	Difference (15)	P-value (16)	Difference (17)	P-value (18)		
Public School	-0.062	0.442	-0.062	0.442	0.000	1.000	-0.059	0.544		
Semi-Public School	0.021	0.781	0.021	0.781	0.036	0.672	0.008	0.931		
Private School	0.041	0.303	0.041	0.303	-0.036	0.655	0.051	0.517		
Bilingual School	-0.048	0.571	-0.083	0.317	-0.073	0.409	-0.081	0.380		
Number of Students	1.274	0.789	-1.223	0.791	2.427	0.688	-2.768	0.627		
ESEC Index	0.152	0.322	0.040	0.443	0.245	0.073	-0.052	0.358		
Frac. Students Started After 3yo	0.002	0.841	-0.007	0.504	-0.012	0.097	0.003	0.693		
Frac. Non-Spanish Parents	-0.008	0.801	-0.002	0.964	-0.049	0.068	-0.024	0.456		
Frac. Less than 50 Books at Home	-0.037	0.159	0.021	0.553	-0.059	0.063	-0.032	0.312		
Test of joint significance	0.423		0.651		0.310		0.511			

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.B.2 Description of Variables

Table 2.B.2: Description of Schools Characteristics Variables.

Variable	Definition	Range
Public	Dummy variable that takes value 1 if the school is public and 0 otherwise	[0;1]
Semi-Public	Dummy variable that takes value 1 if the school is semi-public and 0 otherwise	[0;1]
Private	Dummy variable that takes value 1 if the school is private and 0 otherwise	[0;1]
Bilingual	Dummy variable that takes value 1 if the school instruction is in English and Spanish, and 0 if instruction and schools with Spanish instruction only	[0;1]
# Students	Number of students at this school-grade	(0,170)
ESEC	Index of socio-economic and cultural background	(-4,2)
Frac. Repeaters	Fraction of students who have repeated a grade at this school-grade	[0-1]
Frac. Start after 3 years old	Fraction of students who started school after 3 years old at this school-grade	[0-1]
Frac. Non-Spanish parents	Fraction of students with none of parents born in Spain at this school-grade	[0-1]
Frac. Less 50 books	Fraction of students who have less than 50 books at home at this school-grade	[0-1]
Frac. College	Fraction of students with parents with college education at this school-grade	[0-1]
Frac. Qualified occupation	Fraction of students with parents with qualified occupation at this school-grade	[0-1]
Frac. Basic Occupation	Fraction of students with parents with basic occupation at this school-grade	[0-1]
Frac. High absent	Fraction of students with one or two absents in two weeks at this school-grade	[0-1]
Frac. Low absent	Fraction of students with never absent at this school-grade	[0-1]

Notes:

Table 2.B.3: Description of Student Survey Questions.

Variable	Definition	Range
Question 1: Stress	Have you felt nervous in the exam?	(Nothing; A bit; Quite; A lot)
Question 2: Confidence	I am good at this subject	(Nothing; A bit; Quite; A lot)
Question 3: Preparation of the test	How much have you prepared (studied) the test?	(Nothing; A bit; Quite; A lot)
Question 4: Effort exerted during the test	How many attention/interest have you exert in the test?	[0-10]

Notes:

2.B.3 Additional Heterogeneous Effects

Table 2.B.4: Heterogeneous Effect by Teacher Experience.

Variables	Teacher Above Med. Experience		Teacher Below Med. Experience	
Female*External	-0.071 (0.046)	-0.085** (0.042)	-0.100** (0.049)	-0.103** (0.048)
Female	-0.141*** (0.025)	-0.141*** (0.023)	-0.089*** (0.019)	-0.086*** (0.019)
External	-0.066 (0.068)	-0.048 (0.069)	-0.007 (0.064)	-0.017 (0.061)
Teacher Characteristics				
Female		-0.066 (0.065)		-0.057 (0.056)
Age		-0.005 (0.005)		-0.015*** (0.005)
Civil servant		-0.455*** (0.075)		-0.215*** (0.065)
Full time contract		0.034 (0.247)		-0.108 (0.170)
Constant	0.114* (0.058)	9.713 (10.091)	-0.106* (0.062)	30.092*** (10.246)
Observations	11,418	11,418	12,137	12,137
R-squared	0.020	0.067	0.012	0.029

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

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.B.5: Heterogeneous Effect by Teacher Experience.

Variables	Teacher Above Med. Age		Teacher Below Med. Age	
Female*External	-0.091** (0.046)	-0.100** (0.043)	-0.079 (0.051)	-0.087* (0.049)
Female	-0.115*** (0.024)	-0.116*** (0.022)	-0.116*** (0.021)	-0.107*** (0.020)
External	-0.078 (0.066)	-0.076 (0.069)	0.009 (0.062)	0.011 (0.059)
Teacher Characteristics				
Female		-0.043 (0.060)		-0.067 (0.054)
Civil servant		-0.450*** (0.074)		-0.291*** (0.068)
Experience		0.007* (0.004)		0.032*** (0.007)
Full time contract		0.006 (0.270)		-0.113 (0.156)
Constant	0.067 (0.049)	0.239 (0.287)	-0.064 (0.071)	-0.074 (0.167)
Observations	11,616	11,616	11,939	11,939
R-squared	0.018	0.064	0.015	0.046

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

Notes: Standard errors are clustered at the school level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 3

School Choice Priorities and School Segregation: Evidence from Madrid*

*We thank Manuel Arellano, Ghazala Azmat, Antonio Cabrales, Caterina Calsamiglia, Sara de la Rica, Gabrielle Fack, Julien Grenet and Olmo Silva for their helpful comments. We are grateful to the Madrid Consejería de Educación staff (Ismael Sanz, Luis Pires and Gerardo Azor) and the Census Office staff (Antonio Bermejo) for the provision of data and their support on this project. We thank seminar participants at Paris School of Economics, CUNEF, and CSIC, as well as conference participants at the I Southern Spain Workshop on Economics of Education (University of Malaga). In addition, José Montalbán thanks to the Bank of Spain for its financial support. An earlier version of this work was circulated under the title: “School Choice, Student Mobility and School Segregation: Evidence from Madrid”. All errors are our own.

Abstract

This work aims at broadening the scope of market design questions to school choice by examining how government-determined school choice priorities affect families' choices and pupil sorting across schools in the context of the Boston Mechanism. We use two large-scale school choice reforms in the school choice priority structure undertaken in the region of Madrid (Spain) as a source of variation. In particular, we exploit an inter-district school choice reform that largely expanded families' choice set of schools. We combine an event study first difference across cohorts and a Difference-in-Difference design to identify the impact of the reforms. Using unique administrative data on parents' applications to schools, this paper shows that families reacted to the reform exerting higher inter-district choice and applying to schools located further away from home than before the reform. We find distributional effects of the reform concluding that parents from the highest education levels and parents of non-immigrant students were those who reacted the most in absolute terms. Interestingly, results support the idea of potential information gaps and the dynamic learning process across immigrant status groups. We find a decrease in school segregation by parental education and an increase in school segregation by immigrant status -though effects on the latest fade out when controlling for residential stratification. Results suggest that when parents' school choices exhibit a strong degree of polarization by social and immigrant background, priority structures need to be carefully designed to achieve diversity objectives.

JEL Codes: I24, I28

Keywords: Education and Inequality; Education Policy; School Choice; School Segregation

3.1 Introduction

School segregation has received substantial attention in the education debate. Excessive school segregation is becoming a growing public concern, and international organizations are starting to warn education authorities about the risks of leading to student disadvantage through school segregation (OECD, 2019b; Musset, 2012). A relatively large literature has been devoted to analyzing the effects of school segregation, with recent empirical evidence showing that school segregation may contribute to create unequal opportunities for pupils of different schools (Hoxby, 2000; Card and Rothstein, 2007; Hanushek, Kain and Rivkin, 2009). However, there is still limited evidence on understanding the mechanisms that give rise to school segregation. Böhlmark, Holmlund and Lindahl (2016) highlight two key mechanisms that may explain school segregation. First, residential segregation across neighbourhoods, which levels may be a result of residential sorting, the so-called Tiebout choice (Tiebout, 1956), or housing policies. Second, the mechanisms and rules for assigning pupils to schools, such as the assignment mechanism, student priorities for schools, or boundaries of catchment areas, which are the critical components of school choice. There is a fast-growing literature on the market design questions to school choice, mainly devoted to analyzing the relative performance and strategic implications of alternative matching allocation mechanisms, taking the inputs of school choice -preferences, priorities, and capacities- as exogenous (Abdulkadiroglu and Sönmez, 2003). The extent through which school choice priorities impact school segregation remains unclear.

In this paper, we broaden the scope of the market design questions to school choice by investigating how government-determined school choice priorities affect households' choices and pupil sorting across schools. We use two large-scale school choice reforms in the school choice priority structure as a source of variation. First, the low-income priorities' to the top-ranked school were reduced and granted an additional point to alumni family members of the school in 2012/2013. Second, the resident-based priorities to assign pupils to schools were almost completely abolished

in 2013/2014. The city of Madrid counts 21 school districts that were almost merged *de facto* into a unique single district.¹ Using unique administrative data on the universe of applicants to the public school system from 2010 to 2016 in the Region of Madrid, along with detailed data on school supply, household socioeconomic characteristics and standardized test scores, we explore the relationship between changes in school priorities and variation in families' out of district assignment and school segregation. We combine two different empirical strategies to identify the impact of the reforms. First, we use an event study first difference approach. We compare families entering the educational system for the first time (pre-school age of 3) before and after the reforms. Second, we use a Difference-in-Difference Analysis (DID). A limitation of the institutional context is that both reforms were implemented at the same time for every school district of the city of Madrid. Thus, the institutional setting does not provide an alternative control group, since all families are "treated" by the reforms. However, we exploit an analysis based on the spirit of treatment intensity by focusing on parents that are closer to the school district boundaries ("treatment group"), and comparing them with those whose primary residence is located at the geometric center (centroid) of each the school district boundaries ("control group").

The context of this study is unique due to the school choice allocation mechanism and the levels of school segregation prior to the reform. First, we exploit changes in student priorities in a system where the student assignment mechanism to assign pupils to school is the so-called Boston Mechanism (BM). This allocation mechanism has shown limitations to capture truthful families' preferences ([Abdulkadiroglu and Sönmez, 2003](#)), and has been proved to foster socioeconomic (strategic) segregation across public schools even with open school choice ([Calsamiglia et al., 2017](#)). In the BM, families incentive to top-rank the school in which they have the highest number of points is remarkably strong that we identify families' choices, which are not necessarily preferences. In other words, under this mechanism families choose what

¹In the Region of Madrid, the Regional Government, enlarged the choice zone to the municipal level, granting a larger set of options to all households. The reform moved from around 2,000 within-municipality catchment areas to 179 single municipal zones. We focus on the city of Madrid for our main analysis.

they can access but not necessarily what they want. The change in the inter-district school choice policy reform promoted a *de facto* increase in families' choice set. Before the reform, their choice set was mainly restricted to those schools located within their school district, and after the reform, it is every school in the city of Madrid. In this context, the BM implies an advantage to identify the effect of the reform (this massive increase in the choice set), since we can make this comparison because of the BM. Otherwise, if the allocation mechanism would be strategy-proof (e.g., Deferred Acceptance instead of BM), the preference-revelation should not depend on priorities. Second, the pre-reform levels of school segregation by socioeconomics characteristics in the Region of Madrid were relatively high, while school segregation by immigrant status was rather low.² Then, we can explore the effects on two margins of school segregation widely studied in the empirical evidence (i.e., parental education and immigrant status) that count with highly different pre-determined levels of school segregation. Therefore, our institutional context allows us to investigate the impact of changing the government-determined priorities that increase households' school choice set on family choices and school segregation, under a system that promotes socioeconomic segregation with different baseline levels of school and residential segregation.

We contribute to the currently existing literature in three main dimensions. First, we are able to compute the contemporaneous effect of increasing choice on school segregation at the earliest schooling stage (pre-school for 3 years old students). Most of the literature has focused on secondary education that entails two different aspects: (i) Segregation may be the result of a combination of factors that are shaped in earlier educational stages; (ii) Priority bonus in secondary education are typically based on student grades, while those of primary education are usually centered on socio-demographic indicators which may potentially have a more direct impact on school segregation. Second, this paper can closely relate changes in school choice priorities with the immediate impact on school segregation. Most of the related literature has been focused either on broader contexts or on the impacts of early-stage

²See [Murillo, Garrido and Belavi \(2017\)](#) and [Murillo and Garrido \(2018\)](#).

policy reforms (e.g., primary education) of later phases of the educational career (e.g., secondary education) -which results may be potentially biased by time-variant confounding factors-. Third, this paper explores variables that some of the previous literature does not consider, such as families' choices or the precise geo-location of household's primary residence and schools. This allows us to control for variables that are determinants of school segregation, such as residential segregation.

The relationship between school choice reforms and student segregation has generated significant policy interest. During the last three decades, there has been a clear pattern of educational authorities have increased the degree of school choice in their educational systems ([Musset, 2012](#)). In the US, many school choice reforms were complemented by busing programs (e.g., Seattle in 1999 or North Carolina in 2002). In particular, school choice reforms involve, among others, zoning and de-zoning policies, changes in admission criteria, and changes in the system of assignment of students to schools. The potential effects of those policies on school segregation are not straightforward. From a theoretical point of view, [Jenkins, Micklewright and Schnepf \(2008\)](#) points out three main mechanisms that give rise to school segregation: residential segregation, parental choice, and school' selection of pupils. Residential segregation may be influenced by Tiebout Choice ([Tiebout, 1956](#)) or the residence-based priorities to access schools. We expect a strong correlation between the level of residential and school segregation when residence-based priorities bonus is high. Relaxing proximity-based priorities may affect school segregation in different directions.

On the one hand, parents with higher levels of information and/or preferences for accessing a better performing school may exert higher degrees of choice. If the lack of information and preferences to choose the closest school are correlated with lower household socioeconomic background and immigrant status, we may expect that the reform increases the levels of school segregation.³ On the other hand, under a system

³Literature describes that parents tend to value peer composition of the school mostly and, only to a lesser extent, the effectiveness of the school in the learning progress of students given their socioeconomic characteristics ([Rothstein, 2006](#); [Mizala and Urquiola, 2013](#)). Preferences for schools are different depending on families' socioeconomic backgrounds since preferences for different dimensions of education vary across types ([Anderson, A. Palma and Thisse, 1992](#); [Burgess et al., 2015](#)). [Hastings, Kane and Staiger \(2009\)](#) find that while high-income families care mainly about

with strong residence-based priorities and high levels of school segregation, relaxing residence-based priorities may help families who live in deprived neighbourhoods to opt-out of the assigned school, contributing to reduce the levels of school segregation.

Most of the previous studies have limited scope to disentangle between segregation that is due to school stratification or residential sorting. An advantage of our paper is that thanks to the richness of the administrative data, we are able to identify both school and residential segregation (neighbourhood stratification). Most of the empirical evidence devoted to analyzing the impact of increasing the level of choice in school segregation finds a positive relationship between choice and segregation. [Epple, Romano and Urquiola \(2017\)](#) review the theoretical, computational, and empirical literature on student vouchers. They argue that large-scale voucher systems are associated with more student sorting by ability or parental income, while the introduction of those programs leads public schools to improve. [Böhlmark, Holmlund and Lindahl \(2016\)](#) investigate the effects of a Swedish universal voucher reform in 1992 on school segregation in upper secondary education, which introduced new independent schools and increased the level of parental choice. Exploiting variation in school choice across municipalities, they find that in those regions where school choice became more prevalent, school segregation by immigrant status and parental education increased the most. However, the increase in school segregation that can be attributed to the reform in the long term is of moderate size.⁴ [Söderström and Uusitalo \(2010\)](#) focus on an admission reform undertaken in 2000 that changed admission criteria to those solely based on grades for access to upper secondary schools in Sweden. They find that segregation by ability increased, and although the

test scores, more impoverished and minority families must trade-off preferences for high-performing schools against preferences for a predominantly minority nearby schools. The authors argue that the difference in choice responses leads to a more stratified school system, as the impact of school choice policies is determined eventually by parents' preferences on education. The provision of information also matters. [Hastings and Weinstein \(2008\)](#) show, using a natural field experiment, that low-socioeconomic parents receiving information about the school performance increase their likelihood of choosing a high-scoring school.

Additionally, beyond income factors, the sociology and psychology literature has identified several mechanisms through which school choice is shaped by own aspirations, behaviors, social capital, and networks. For example, [Teske and Schneider \(2001\)](#) discuss parental involvement and motivation as drivers of differences in school choice.

⁴[Böhlmark and Lindahl \(2007\)](#) find evidence of students being sorted by immigrant origin and parental background shortly after the reform was implemented.

increase in segregation by socioeconomic background is explained by ability sorting, the increase in immigrant segregation may be attributed to the reform.⁵ Other strands of the literature analyze the characteristics of those families who decide to opt-out of their assigned school (Levin, 1998; Hastings, Kane and Staiger, 2005), the “typing points” for schools (Card, Mas and Rothstein, 2008),⁶ the impact of choosing private schools on school segregation (Hsieh and Urquiola, 2006; Figlio and Stone, 2001).⁷, peer effects (Duflo, Dupas and Kremer, 2011), or the impact of segregation on social cohesion (Billings, Deming and Rockoff, 2014).

We find that the inter-district school choice reform undertaken in the city of Madrid is associated with an increase in the fraction of outer school district applications and an increase in the distance to the final assigned school of 3 percentage points and 259 meters respectively (30% and 22% with respect to the baseline mean). This result implies that families change their choices when their choice set is amplified. Results are robust to different specifications using both the event study first difference approach and DID. Even though that families of higher quintiles of parental education reacted more to the reform in absolute terms, the effects were somewhat similar across parental education quintiles (about 30%) in relative terms. Different conditions emerge when comparing the effects by immigrant status. The increase in the outer district assignment and distance to first assigned school is entirely driven by families with Spanish children since parents of immigrant children do not seem to have reacted at all in the first two years of the reform. Interestingly, results support the idea of potential information gaps across immigrant status groups, since immigrant started

⁵Yang Hansen and Gustafsson (2016) find the same results using multilevel models. Burgess, Propper and Wilson (2007) explore the relationship between school segregation and the number of schools available in three cities of the UK, finding an increase in the levels of school segregation by immigrant status, parental income and student ability.

⁶There is evidence of the so-called “white flight” effects, which happens when white students decide to leave certain schools in which the fraction of a minority group is above certain threshold.

⁷In addition, the New Zealand reforms that were implemented in the 90s have been shown to have increased social and immigrant segregation of schools (Ladd and Fiske, 2001). However, other factors interacting with choice settings may be playing a hidden role in the real effects, such as how schools are able to implement explicit or implicit forms of discrimination. For example, Burgess and Briggs (2010) investigate the effect of school choice on social mobility in secondary education in England. They find that children from low-income families are less likely to get places in good schools, and that probability is unaffected by the degree of school choice. This suggests that there must be other additional features belonging (or related) to the educational system that affect student mobility beyond the degree of school choice.

to react to the reform (by the same magnitude as Spanish families) three years after the implementation of the reform. The dynamics seem to point out for a learning process over time of families with non-Spanish children, catching up on the absolute effect of the reform in the last observed two years (2015 and 2016).

We measure school segregation using the Mutual Information Index, which satisfies several desirable properties ([Frankel and Volij, 2011](#)). We find a decreasing trend in school segregation by parental education over time (mostly driven by the decrease in within school district segregation), but an increasing trend in school segregation by immigrant status. The 2012/2013's reform, in which priorities for low-income pupils were reduced and new priority was granted for alumni family members, is associated with a contemporaneous increase in the levels of school segregation by both parental education and immigrant status, controlling for residential segregation, school districts fixed effects, and time-variant district characteristics. On the contrary, we find that the 2013/2014's inter-district school choice reform is associated with a reduction in school segregation by parental education of about 1.8%. However, this reform is associated with an immediate increase in school segregation by the immigrant status of 1.6% in the first year of the reform, although the effect is not statistically significant when controlling for pure residence-based segregation. Results indicate that the potential welfare loss of the inter-district school choice reform caused by this immediate increase in school segregation completely fades out over time potentially due to the learning process of immigrant families.

There is an institutional characteristic that directly relates to the interpretation of the results. In the list of priority points in case of over-demand of schools, a specific point is decentralized to the school principals' decision to break ties. School principals may have stronger preferences for students of higher ability (a characteristic that is correlated with higher socioeconomic background and non-immigrant status). In this case, school principals would use this point to increase school segregation. It implies that we are potentially estimating a lower bound of the effect when we observe a decrease in school segregation by parental education, and an upper bound when school segregation by immigrant status slightly increase in the first year of the

reform.

The inter-district reform was implemented in several municipalities (usually those of medium size) in 2012/2013 (as well as the low-income and alumni pupil bonus criteria), whereas for the larger ones (including Madrid), it took place in 2013/2014. We exploit the gradual implementation of the policy in different municipalities conditional on the population size to estimate changes in school segregation associated with the reform. We show that results on the willingness to commute of households and the increase in school segregation by immigrant status are robust and consistent. Although large-size and middle-size municipalities present similar levels of school segregation by immigrant status, school segregation seems to have increased more on large-size municipalities in the reform years. This pattern is driven by the within municipality school segregation, which increasingly seems to be more salient in large-size municipalities. Interestingly, the between municipality school segregation seems to be almost negligible for large-size and middle-size municipalities, but it is almost equally important as the within municipality segregation in small-size municipalities. Overall, our results are consistent with the theoretical predictions: when the level of school and residential segregation is relatively high, increasing the degree of choice may help to decrease segregation. The opposite happens in a context in which those levels are relatively low.

Organization of the Paper. The remainder of the paper is organized as follows. Section 3.2 describes and contextualizes the school choice reform undertaken in Madrid. Data are detailed in Section 3.3. The empirical strategy and potential identification challenges are discussed in Section 3.4. Results are shown in Section 3.5. Section 3.6 addresses several robustness checks. Finally, Section 3.7 concludes.

3.2 Institutional Background

The Spanish Education System. The Spanish education system consists of 10 years of compulsory education, which starts at age 6 and includes six years of primary school (up to age 12) and four years of lower secondary education (up to age

16). Even though compulsory primary education starts at the age of 6, students are offered free universal access to the public education system from age of 3 onwards. Since most of publicly-funded schools offer pre-school and primary education together, age of 3 is typically the time when families enroll their children to school.⁸

Regarding the access to schools, the 1978 Spanish constitution grants the right to education and the freedom to educate children, an equilibrium of rights stemming from a political pact between progressive and conservative forces.⁹ In the following years, this was accompanied by a decentralization process through which educational policies started to be jointly determined at the national, regional and municipal level.¹⁰ Since then, the central government is responsible for establishing the organic laws (*Organic Laws*) and the royal decrees that the regional governments are allowed to further develop as long as they do not contradict the organic laws. With respect to the Spanish school choice policies in the years around the reform, the national organic law in place at the time of the reform (LOE) established the general regulatory principles to be followed by the regional governments in order to determine the priority criteria of students in over-demanded schools.¹¹

In Spain, the vast majority of school networks are publicly funded. Such a network includes public and semi-public schools.¹² The government fully funds public schools and managed by civil servants and local school boards. Semi-public schools (*centros concertados*) are privately run but mostly financed through public funds. Although tuition fees are not allowed in semi-public schools, in practice, parents pay small quasi-compulsory symbolic donations for essential educational services that can act as a barrier to entry for disadvantaged families. Concerning admissions, all the schools in the public system (public and semi-public) are expected to unconditionally accept all students assigned by the centralized school choice mechanism, provided

⁸Pre-school Education is entirely publicly funded from ages 3 to 6. This right is recognized in the [Organic Law 1/1990 \(LOGSE\)](#).

⁹In terms of education policy, the second principle was translated into the 1985 education act (LODE), which explicitly regulated the freedom of families to choose their children's school. See [Organic Law 8/1985 \(LODE\)](#).

¹⁰See the [Organic Law 9/1992: Ley Orgánica de transferencia de competencias a Comunidades Autónomas que accedieron a la autonomía por la vía del artículo 143 de la Constitución](#).

¹¹[Organic Law 2/2006 \(LOE\)](#).

¹²We follow [Calsamiglia and Güell \(2019\)](#) that refer to the network of privately managed schools as semi-public schools.

demand does not exceed supply.

School Choice in Madrid. In the Region of Madrid, the majority of schools (about 85%) are part of the publicly-funded network of schools. This system includes publicly managed schools (which enroll approximately around 50% of all students) and semi-public schools (which cover around 35% of all students). Semi-public schools tend to be located in urban areas which are larger in size and serve more upper-middle-income and non-immigrant households.¹³

The school choice system is based on a centralized assignment mechanism that is used to allocate students to schools in the publicly-funded system (both public and semi-public) for pre-school (starting at age of 3), primary, lower secondary and special education. More than 96 percent of the students in the Region of Madrid attend pre-school, and the school choices decisions are taken when they are three years old ([Anghel, Cabrales and Carro, 2016](#)). Students who are enrolled in pre-school in a given school have full priority over every student applying to primary education. If all vacancies are filled at the age of 3 years old and no student leaves the school, there would be no available slots for those who are not previously enrolled in that level at that cohort. As a result, changes in school after the age of 3 are not frequent, and the vast majority of families make their schooling decisions at this point in time.

Families are requested to submit a rank-order list of schools up to a total number of choices, and their children are allocated by the centralized and algorithm-based automatic allocation procedure, the so-called Boston Mechanism ([Abdulkadiroglu and Sönmez, 2003](#)).¹⁴ The application timing works as follows. Before the school year starts in September (between the end of April and early May), every participating family is requested to submit the rank-order list of schools to their first-choice school. Applicants are assigned into a school using the Boston Mechanism (BM hereafter), a centralized school choice system that works as follows. First, students are allocated to their first-choice school. For schools where there is an over-demand of students, students are granted priority points (according to several criteria which depend on

¹³Some authors have argued that preferences for education in Spain are mediated by the existence of the semi-public network ([Arellano and Zamarro, 2007](#); [Mancebón et al., 2012](#)).

¹⁴It has been further updated and regulated after the 2006 LOE education act was passed.

student characteristics and location of the household or parental job) which provide them with a rank number that assigns places to students until all available places are filled. Ties are broken conditional on priority bonus points obtained¹⁵. In the second step, students who are rejected from their first-choice are proposed to their second submitted school in the rank-ordered list provided that there are available seats after the first step. If there are more applicants than available places, students are allocated in the same way in the first step with the priority points granted in the first-choice school. In the third step, those students who are rejected from their second choice are proposed to their third choice, and the mechanism continues until all students are assigned a seat or are rejected from all schools in their rank-order list. The final assignment is made public in June, and enrollment must take place at the end of June (for pre-primary and primary education) or July to September (for lower secondary education). A special feature of the system is that students' priority points that are used for tie-breaking at all stages are based on the ones obtained for the first-choice.¹⁶

Until recently, the BM has been very influential in practice (beyond Spanish regions, US school districts which used this mechanism including Boston, Cambridge, Denver, Minneapolis, and Seattle, among others, as well as other cities such as Beijing, Amsterdam or Frankfurt). One of the special features of this assignment system is that the choice of the top-ranked school is highly important since the assignment in each round is final. The probability of a student being admitted in the second round relative to the first is dramatically reduced, and the chances decrease even more in further rounds.¹⁷ Recent empirical evidence supports the theory. In the case of Barcelona, [Calsamiglia and Güell \(2019\)](#) highlight the fact that more than 85 percent of the assignments are resolved in the first round in Barcelona, and

¹⁵See Table 3.C.1 in the Appendix for further details.

¹⁶[Calsamiglia \(2014\)](#) states that the main reason why the government uses this procedure is that it is computationally easier. Alternative assignment mechanisms require computational power that currently the education administration cannot deal with. Section 3.C.8 provides further details on the theoretical properties of the BM assignment mechanism.

¹⁷[Abdulkadiroglu and Sönmez \(2003\)](#) highlight that one of the major difficulties of the BM is the fact that it is not strategy-proof. A student may have a very high priority to enter school s , but if she does not list it as her top-ranked school, she loses her seat in favor of students who have listed s as their top choice. BM provides incentives to families to misreport their preferences by ranking first those schools in which they have higher priorities to be admitted.

this is persistent across different cities worldwide.¹⁸ In Madrid, about 86 percent of children were assigned to the school they ranked first (see Section 3.3).

Another essential feature of the BM is that this mechanism tends to promote segregation across schools. Calsamiglia et al. (2017) shows that the BM fosters socioeconomic (strategic) segregation across public schools, even with open school choice. Therefore, the context of this paper is expected to show high predetermined levels of school segregation, which are confirmed by the empirical evidence. The Region of Madrid is the most socially segregated (socially defined as social, economic, and cultural characteristics of the family) among all autonomous communities in Spain and neighboring countries in secondary education (Murillo and Garrido, 2018). In terms of immigrant-origin segregation, the levels are rather low in comparison to other Spanish regions (Murillo, Garrido and Belavi, 2017). However, as mention in the Section 3.1, this paper studies segregation levels right at the beginning of the schooling age (families applying for pre-school with children of 3 years old), which may be different from those studied in secondary education.¹⁹

Reform of the priority criteria to school access in the city of Madrid.

In the case of over-demand at a specific school, students are assigned to schools based on a government-determined priority criterion, which grants points to students according to their characteristics and their home residence or parental job location. For school choice, the Region of Madrid counts 179 municipalities, with the medium and large-size municipalities being subsequently divided into school choice catchment areas.²⁰ In particular, the city of Madrid (the largest municipality and our main unit of analysis) is divided into 21 school districts, which coincide with such choice catchment areas. Figure 3.A.1 shows a map of the distribution of the 21 school districts in the city of Madrid.

Table 3.B.1 shows the score scale used in the city of Madrid before and after the

¹⁸See also Abdulkadiroglu et al. (2006) for Boston; Hastings, Kane and Staiger (2009) for Charlotte; Lavy (2010) for Tel Aviv; and De Haan et al. (2015) for Amsterdam.

¹⁹Another relevant factor that may have contributed to the increasing levels of segregation is the implementation of the bilingual program in the Region of Madrid. Anghel, Cabrales and Carro (2016) find that observable characteristics of families changed against students with immigrant status (and those with lower socioeconomic background) in schools which became bilingual when the policy was implemented in the first place.

²⁰Called as *zonas de influencia*.

reform.²¹ Before the school year 2012/2013, children living in the (at the boundary of) district of the top-ranked school received 4 points (2 points).²² Regarding individual student characteristics, students were awarded 2 points if their per capita household income was under the IPREM Index (7,236.60 euros), and got 1 point if their per capita household income was between 100% and 200% of this index (between 7,236.60 and 14,473.20 euros).²³ Families which ranked a school where there was a sibling enrolled got 4 points for the first sibling, plus an additional 3 for every one of them enrolled at this school. Students received extra points if they had a family member with a disability (1.5 points), and if they belonged to a large family (1.5 if general - 3 children -, and 2.5 if special- 4 or more children-). Besides, a specific point (1 point) was decentralized to the school principals' decision, which must be decided according to objective criteria that are made public.

In March 2012, the Regional Government announced a reform that aimed at strengthening the principle of school choice by households with children entering pre-primary, primary and lower secondary schools.²⁴ The Regional Government founded its arguments on the constitutional right that parents have to educate their children based on their convictions. The goals of the government policy were to increase families participation to improve the availability of information on schools (through the results on the standardized test scores, schools' educational program, school resources, and services), to simplify the admission process, to promote school competition, and to enhance free school choice. In particular, the reform modified the computation of pupils' priorities in each school (see Table 3.B.1). The changes in priority points and school districts were implemented in two consecutive years:

1. In 2012/2013:

- The criteria to obtain bonus points granted to low-income families was reduced and changed.

²¹Table 3.B.1 also applies to the rest of the municipalities regarding the individual characteristics.

²²Families receive the same scale of points if any of the parent/guardians are working in the district of the top-ranked school.

²³IPREM is the acronym in Spanish for Multiple Effects Income Public Index and represents a minimum annual threshold for social programs and subsidy eligibility. The Index remained constant between 2010 and 2015.

²⁴Order 2939/2012 of March 9 of the Regional Government of Madrid.

- A new priority bonus was granted when a pupil's family member was an alumni student of the top-ranked school.²⁵
2. In 2013/2014: The proximity to the school criterion was relaxed. The Regional Government of Madrid updated the regulatory framework with a regional decree which regulated the single school choice for all the municipalities of the region.²⁶ This change implied moving from 21 school districts (with around 25 schools per district) as choice catchment areas to a virtually single municipal school choice district with more than 500 schools in the city of Madrid.²⁷

In 2012/2013, a sharp decline in the bonus for low-income families was introduced. Children were awarded 2 points if the family received the Minimum Income for Insertion Subsidy (*Renta Mínima de Inserción*), a social program which is granted to a small proportion of low-income households with no earnings (0.9% of the total population in the Region of Madrid). The number of recipients of this subsidy is much smaller (around 30,000 households in a region of more than 6 million population) than the number of families with a per capita household income under the 100% of the IPREM (around 15% of the population)²⁸. Finally, an additional 1.5 point was awarded to students in a school where any family member had been an alumni student, a feature that may potentially limit equality of opportunity of students to access certain schools, given the weight given to the socioeconomic background of parents.²⁹

In 2013/2014, families' incentives to apply for a school inside their residence school district were shifted with the implementation of the inter-district school choice (the 2013/2014 reform is referred to as the inter-district school choice reform

²⁵ Additionally, more points were granted to families with siblings enrolled in the school. Following Calsamiglia and Güell (2019), we consider this change as irrelevant for the analysis, given that families' choice are previously conditioned by their older children's choice, and we do not include pupils with older siblings in our main analysis.

²⁶ Decree 29/2013 (*Decreto del Consejo de Gobierno, de libertad de elección de centro escolar en la Comunidad de Madrid*).

²⁷ Relative to 2012/2013, 2 extra points were awarded to families with siblings enrolled at the school. We do not consider this change relevant for our analysis.

²⁸ The share of households at risk of poverty or social exclusion in Madrid in 2014 was 19.2%. The poverty line in 2014 was established at 7,961 euros, slightly higher than the IPREM index, 6,390 euros.

²⁹ Tie-break criteria were also slightly modified, as can be seen in Table 3.C.1 of Appendix.

henceforth). A pupil living (or with parents working) and applying to a school in the same school district was awarded an additional 0.5 points, plus 4 points (2 points) if the school was located in the same municipality (other municipality in the region) of the household or the parental workplace.³⁰ Overall, the inter-district school choice reform implied a substantial drop in the importance of the proximity criteria for over-demanded schools.

Implementation of the reform in other municipalities. In the Region of Madrid, the number of school catchment areas (school districts in the city of Madrid) was a function of the size of the municipality. The region has 179 municipalities. The smallest 142 municipalities - with a population of less than 10,000/15,000 inhabitants - have always had a unique school catchment area, whereas larger municipalities had more than one. Due to capacity constraints, the expansion of the inter-district school choice in medium and large-size municipalities was conducted in two consecutive years: (i) In 2012/2013, 22 municipalities, mostly of medium size (with a population between 15,000 and 100,000 inhabitants approximately), adopted the inter-district school choice policy; (ii) In 2013/2014, the remaining 15 municipalities (mostly the larger ones, including the city of Madrid) adopted the inter-district school choice policy.³¹ We use this gradual implementation across municipalities to perform robustness checks.

3.3 Data and Summary Statistics

3.3.1 Data

We use a combination of four administrative datasets that provide rich and unique information on the universe of pre-school and primary school applications of each household in the Region of Madrid, the characteristics of the publicly funded schools in the region of Madrid, the education level of households at the census block level

³⁰The weight of going to a school within the same district of household residence/parental job went from 4 points out of 4 to 0.5 points out of 4.5 points after the reform.

³¹Table 3.C.2 in the Appendix provides a summary of the municipalities that joined the single-zone school choice system across years.

in the city of Madrid, and the standardized test scores at the school level. Data on student applications, schools, and school test scores were provided by the Education Ministry (*Consejería de Educación*) of the Regional Government of Madrid, and data on parental education was obtained from the Madrid Census. Information is available for every year from 2010/11 (2010 hereafter) to 2016/2017 (2016 hereafter).

Applications for pre-school and primary education. Our primary source of analysis is a unique administrative database containing information on the universe of students who applied to a primary school in the Region of Madrid. For each applicant, the dataset contains the top-ranked school from the rank-order list, the basic student information regarding family characteristics, home address, total priority points obtained based on such characteristics, and the assigned school. Regarding family information, the data contain the precise geo-location of each pupil's home residence which we link (with the help of a geo-location software) to different geographical areas (census blocks, neighborhoods and districts). Besides, the application contains information about the pupil's country of birth, which we use to construct a proxy for immigrant background status.

School database. We use the universe of schools in the region of Madrid provided by the Regional Government. This database includes the precise geographic coordinates of each school, the school type (public, semi-public or private), whether the school offers bilingual education (schools with English and Spanish instruction), and the levels of education offered.

Household socioeconomic characteristics. We use information from the Census Office of the city of Madrid which provides the distribution of education levels of the population by census block of the city on January 1 2012, 2013 and 2014. The data is accessible for the three geographical levels of disaggregation: districts, neighborhoods, and census blocks. The most disaggregated units are the census blocks (*Sección Censal*), which are constructed for local, regional and national election purposes (assigning each census block to one voting center), and usually contain no more than 2,500 individuals.³² We have access to information on the

³²Figure 3.C.1 in the Appendix includes an example of a block of the Central district of Madrid.

proportion of population in each level of education by age groups at the census block level. We use this to translate the corresponding level of education to an equivalent number of years of schooling, which allows us to compute the average number of years of schooling in each census block.³³ We assign to each family the corresponding value of the census block where the family resides. This proxy for parental education is, therefore affected by measurement error. Potential limitations and unobserved heterogeneity issues are discussed in Section 3.4.³⁴

School Standardized Test Scores. To proxy school quality, we use a standardized exam administered for all 6th Grade students in the region of Madrid between the 2004/2005 and 2014/2015 school years. The exam, known as the Essential Knowledge and Skills test (*CDI-Conocimientos y Destrezas Indispensables*) was designed for education policy measures and did not have any specific academic consequences for students. The goal of this test was to provide information for policymakers, schools, and families about the school’s average performance. The test focused mostly on curriculum content knowledge in the areas of reading and mathematics. The results were publicized every year with the purpose of facilitating school choice for families with new students entering the system.

The population of interest for our main analysis is formed of households who live in the city of Madrid and apply to schools in Madrid. We use families who apply for pre-school at the age of 3 years old, which is the age at which the vast majority make their schooling decisions. We restrict our sample to families that have no siblings in the top-ranked school ([Calsamiglia and Güell, 2019](#)).³⁵ Families with older siblings enrolled at the school have different incentives and behavior compared to the rest of applicants, since their past choice conditions their present choice. The number

³³The construction of this variable is detailed in Appendix Section 3.C.3. Ideally, we would use data from the 25-49 age group to obtain a more accurate proxy for parental education, as this is the most relevant level of education for parents with pre-school and primary school children. Nevertheless, we do not use this measure in the main analysis since the database that includes such desegregation corresponds to the year 2017 onward and not to pre-reform years.

³⁴We assign to each family the corresponding value of the census block in the following way: (i) Those applying in 2010, 2011 or 2012 are assigned the values of January 1 in 2012; (ii) Those applying in 2013 are assigned the values of January 1 in 2013; (iii) Those applying in 2014, 2015 and 2016 are assigned the values of January 1 in 2014. Results are robust to changes in the computation of this variable and are available upon request.

³⁵[Calsamiglia and Güell \(2019\)](#) also restrict their main estimation sample to families who apply for pre-school at 3 years old.

of bonus points that families obtain when applying to a school where a sibling is already enrolled is the highest. As a result, admission to those schools is almost automatically guaranteed. These families may, therefore, react differently to policy changes since they have distinct preferences and incentives which could make their decisions more inelastic to changes in the bonus criteria. Table 3.C.5 presents the restrictions that we use to construct the analysis sample.

3.3.2 Summary statistics of applicants

Table 3.B.2 presents the summary statistics of applicants. First, the majority of applicants (more 60 percent) are applying to start in pre-school education at the age of 3, which is our population of interest. Beyond this, a large fraction of applicants are native students (around 85 percent), whereas the rest represents the foreign-born pupils' population. This table presents a decreasing trend in the number of families of 3 years-old children applying to pre-school over time, with a sharp reduction in 2013. Table 3.C.3 presents the population census of 3 years-old children over the period under study, showing that the observed drops in the total population at that age are not specific to our estimation sample. In 2013, there was a significant rise in the proportion of applicants with immigrant background, from 13% to 16.7%, although this is also consistent with the city demographics.³⁶ The share of female applicants remains constant over the period. We observe that the gradual decrease in the sample size is not correlated with parental education. Therefore, the changes in sample size over time are driven by city demographics. In our empirical strategy, we account for these demographic changes by tracking residential dynamics of the immigrant population.³⁷

³⁶According to official municipal data from the city of Madrid, there was a sharp decrease in the number of births in 2009 (i.e., individuals aged 3 in 2012) and 2010 (who were 3 in 2013) with respect to those born in 2008. In particular, the number of births in the city of Madrid was 36,663 in 2008, 35,147 in 2009 and 33,987 in 2008. Conversely, the number of immigrants increased between 2010 and 2013. Hence, the observed drops in the census are consistent with those observed in Table 3.B.2.

³⁷Pupils' awarded places with each of the two new priority bonus created in the 2012 reform only account for a small part: between 3% and 4% for the new low-income bonus and around 6% for alumni student relatives at the school. We do not drop these students from the main sample analysis due to the fact that we do not have this information for years before the reform. Discarding these observations may lead to a sample selection bias between the period before and after the

Figure 3.A.4 illustrates the distribution of parental education (measured as the fraction of parents with a college education by census block) across school districts and census blocks of the city of Madrid in 2013. Higher quintiles of parental education are concentrated around the center of the city, while lower quintiles are over-represented in the periphery. This suggests a high level of spatial residential segregation in the city of Madrid by parental education.

We find two limitations in the data. First, we do not have the information concerning the location of parents' workplace. Given that the reform reduces the importance of the district of both the household and workplace location, we need to assume that parental mobility to the workplace is constant during the years of the reform. Otherwise, changes in out-of-district assignment could be driven by changes in parents' workplaces during these years. We believe that this is not a reliable assumption since changes in the location of the workplace are limited and challenging to exert by families, particularly over the years when the world economic crisis was particularly strong in Spain. Moreover, households typically choose where to locate their primary residence to take advantage of the supply of public local good (Tiebout, 1956).

The second data limitation relates with the measure of parental education. This variable may potentially suffer from unobserved heterogeneity caused by the measurement error of parental education. While we can measure student immigrant status with a full degree of accuracy (at the student level), the data from the parental background is more limited since we use average values across census blocks which each cover a population of about 2,000 inhabitants. This measure may incur in measurement error, which potentially reverses to the mean tails in each block. For example, in low and middle-income census blocks, highly educated parents are identified with much lower education levels (and not identified through education quintiles, as their observed educational level is averaged with those living in their same census block). A similar phenomenon occurs for low-educated households living in highly educated blocks. If the reaction to the reform is higher in relative

reform.

terms for highly educated parents, our measure of parental education may incur in measurement error that is correlated with the effect of the inter-district school choice reform. If that is the case, our estimates of changes in out-of-district assignment by parental education may suffer from negative bias, so that the “true” effects may be even larger after the reform for relatively highly educated families.³⁸ Besides, our measure of school segregation (through parental education) may be under-estimated, especially in low-educated blocks. This leads to an under-estimation of the gaps in mobility and an over-estimation of the change in school segregation by parental education after the reform.

3.4 Empirical Strategy

In this paper, we first attempt to estimate the effect of changes in pupil priorities on families’ school choices and final school assignment, and second, whether the reform had an impact on the levels of segregation across schools (measured by parental education and immigrant status). Hence, we investigate whether the inter-district school choice reform had a direct impact on families’ willingness to commute to schools located in different school districts and whether this affected pupil sorting across schools. An advantage of the context under study is that we are able to closely relate changes in school choice priorities with contemporaneous school segregation changes at the start of the schooling decisions. There was no other modification of the school system apart from the sizeable inter-district reform in the city of Madrid, which we use as the primary source of variation.

3.4.1 Out of School District Choice and Assignment

3.4.1.1 Event Study: First Difference Approach

The two policy reforms undertaken in the city of Madrid modified the set of feasible schools where families had a high priority of admission. We identify the

³⁸It is unlikely that the sign of the bias would go in the opposite direction, since on average, highly educated families living in districts with high-quality schools do not prefer low-educated areas with low performing schools.

effect of these changes on families' choices by comparing school choices immediately before and after these modifications. First, we investigate whether parents aimed to get admission or were finally assigned to a school located in a different school district from the household's primary residence. Second, we use two different measures for computing the commuting distance between the household across two margins: (i) Parents' top-ranked school; (ii) Their assigned school. We use the Open Source Routing Machine (OSRM) routine, which returns the travel distance using the latitude and longitude coordinates of the household and the school.³⁹ Both variables provide different information. While outer district applications may be seen as a simple "extensive" margin measure of mobility, the average distance to first-choice (and assigned) school may represent a combination of the "intensive" and "extensive" margins of mobility, as the distance traveled is composed of both within and between school district mobility.

To estimate the effect of the changes in priorities on families' out-of-district choices we estimate the following reduced-form specification:

$$D_{ibsd t} = \alpha + \delta_t + X'_{isdt}\beta + B'_{bt}\gamma + \nu_b + \epsilon_{ibsd t} \quad (3.1)$$

where $D_{ibsd t}$ is either a dummy that takes value 1 if pupil i living in census block b applies for a school s located in a different school district d where she resides in year t , and 0 otherwise, or the travel distance from the residence of the pupil i in census block b and school district d to the school s that was top-ranked in year t in her application form; δ_t are year fixed effects; ν_b are census blocks fixed effects; X'_{isdt} is a vector of households characteristics, which includes the number of schools in a radius of 500 meters (i.e., a proxy for school supply), the Euclidean distances (in meters) to the closest school district border and to the nearest school; B'_{bt} is a vector of time-variant census block characteristics, such as population density and average income; and ϵ_{ist} is the error term. The coefficients of interests are the year

³⁹The command computes this distance based on a map: we use OpenStreetMap as it allows to work offline with an unlimited request of distances to be computed and replicated (Huber and Rust, 2016). The database contains the UTM coordinates in ED50 base. The OSRM command needs GPS coordinates and ETRS89 base so we use a Geographical Information System (GIS) to convert them into suitable coordinates.

fixed effects. The fixed effect of the year prior to the reform is excluded, such that δ_t are interpreted as the change in the dependent variable in year t with respect to the baseline year 2011. Therefore, α captures the mean outcome variable in the year prior to the reform (academic year 2011/2012) for the excluded census block. We cluster the standard errors at the census block level to account for the fact that willingness to commute may exhibit spatial and serial correlation within a given census block. We may expect that families' responses to the reform may be different depending on the distance to the boundary of catchment areas, the supply of schools surrounding their primary residence, or the census block in which they reside. Note that our data are constructed as a repeated cross-section of families with 3 years old children who apply for pre-school, which implies that the sample of families is different every year. We also perform different specifications that include neighbourhoods or school district fixed effects, as well as observable characteristics at these two geographical units, finding that our results are robust and point estimates are barely sensitive to these specifications.⁴⁰

Using parents who are already in the system and who apply to a school change would potentially bias the results. First, after the first year of entry in the system (at the age of 3), pupils have priority to remain at the same school (if they plan to continue in the first grade of primary education). Pupils who enter the system after the age of 3 do not face the same set of feasible schools as pupils who enter the system at the age of 3, as they are only left with the available slots due to current students leaving the school, or under-demanded schools. Second, these families may have different preferences for schools. Third, these households may act even more strategically due to the fact that they potentially know better how the system works since they have previously applied. Hence, pupils who enter the system for the first time (at the age of 3) may have different preferences, priorities and behavior than pupils who aim to change the school later on the system, making those groups difficult to compare.⁴¹ Therefore, to avoid bias and take advantage of the cleared

⁴⁰Results are available upon request.

⁴¹In our empirical framework, we need to assume that the distribution of parents' preferences remains constant over time. This seems to be highly plausible at least for the specific years of the reform.

school choice market in the first pre-school year, we use families with 3 years old children as our estimation sample.

3.4.1.2 Difference-in-Difference Analysis

We implement a Difference-in-Difference (DID) estimation strategy. The DID strategy is a mean comparison design that consists of creating a counterfactual outcome for the treatment group using the outcome of the control group. The limitation of this institutional context is that both reforms were implemented at the same time for every school district of the city of Madrid. Thus, the institutional setting does not provide an alternative control group that is not affected by the reforms, since all families are “treated”. However, the school choice literature finds that distance to school enters as a harmful component of parents’ preferences. We can, therefore, exploit an analysis in the spirit of treatment intensity by focusing on parents that are closer to the school district boundaries, and comparing them with those who are further away. Parents whose primary residence locates within 300 meters close to the school district boundaries are defined as “treatment group”. Families whose primary residence locates within 150 meters close to the geometric center of the school district boundaries (centroid) are defined as “control group”.

Figure 3.A.3 shows the trends in out-of-district assignment and average distance to assigned school of defined both groups of families. It seems that the treatment group presents higher levels of out-of-district assignment and average distance to assigned school than families whose primary residence is at the school district centroid. The trends for the treated and the control group are both slightly constant and almost parallel (especially for the out-of-district assignment), which makes the common trend assumption reasonably plausible.

Even if the pre-trends may look fairly similar, the two groups may still systematically differ due to the discretionary nature of the selection rule applied. We compare whether household characteristics are balanced between treatment and control groups in the pre-reform years. Finding that some groups of families are more likely to sort on each of the groups would indicate endogenous sample selection. Linear regressions are

performed using each of the households' observable characteristics (i.e., immigrant status, gender, distance to closest schools, number of schools in a radius of 500 meters) as the dependent variable. Table 3.C.6 shows the treatment-control balance in fixed households characteristics in pre-reform years. Two out of eight contrasts are significant at the 10% level. The fraction of families whose primary residence is located within a census block that is labeled on the highest quintile of education and the euclidean distance to the closest school are higher in families located close to the school choice district boundaries. We also test for the joint significance of the observable characteristics for treatment and control groups. We cannot reject the null hypothesis that the F-statistic is equal to zero.

To estimate the effect of the changes in priorities on families' out-of-district assignment we estimate the following reduced-form specification:

$$D_{ibsd t} = \alpha + \delta T_i + \lambda Y_{2012 t} + \zeta \text{Post}_{2013 t} + \kappa T_i * Y_{2012 t} + \pi T_i * \text{Post}_{2013 t} + X'_{isdt} \beta + B'_{bt} \gamma + \theta_d + \epsilon_{ibsd t} \quad (3.2)$$

where $D_{ibsd t}$ is either a dummy that takes value 1 if pupil i living in census block b applies for a school s located in a different school district d where she resides in year t , and 0 otherwise, or the travel distance from the residence of the pupil i in census block b and school district d to the school s that was top-ranked in year t in her application form; T_i is a dummy variable equal to 1 when the households belong to the treatment group, and zero otherwise; $Y_{2012 t}$ is a dummy variable equal to 1 when the year is 2012, and zero otherwise; $\text{Post}_{2013 t}$ is a dummy variable equal to 1 when the year is 2013 or later, and zero otherwise; θ_d are school district fixed effects; X'_{isdt} is a vector of households characteristics, which includes the number of schools in a radius of 500 meters (i.e., a proxy for school supply), the euclidean distances (in meters) to the nearest school; B'_{bt} is a vector of time-variant census block characteristics, such as population density and average income; and ϵ_{ist} is the error term. We cluster the standard errors at the census block level to account for the fact that willingness to commute may exhibit spatial and serial correlation within

a given census block. The coefficients of interests are κ and π , that provide the effect of the 2012's reform and the 2013's inter-district school choice reform respectively.

3.4.2 School Segregation

The first goal of this study is to analyze whether the school choice reform had an impact on the outer district top-ranked and assigned school. The second goal of this study is to investigate whether changes in the outer district assigned school were translated into different school segregation levels among new pupils entering the school system. We test whether changes in the levels of segregation across schools by parental education and by immigrant status are associated with the policy reform. As explained in Section 3.3, we proxy parental education by the average years of schooling at the census block level, and we divide families by quintiles of parental education. We measure immigrant status through a dummy variable that takes value one when the children are non-Spanish and zero otherwise.⁴²

We use the Mutual Information Index (M hereafter) as our main measure of segregation across schools. We complement this measure using other global segregation indices widely used in the literature, such as Dissimilarity Index (Duncan and Duncan, 1955), the Adjusted R-squared, and the Normalized Exposure Index. Results are robust to the use of these global segregation indices.⁴³ We define M as:

$$M = 1 - \left[\left(\sum_{\text{School } s} \left(\frac{\# \text{ Pupils in school } s}{\# \text{ Total pupils}} \right) * E_s \right) / E_g \right] \quad (3.3)$$

where s is the school, E_s and E_g are the Entropy Diversity Index distribution of each school and group g (defined as parental education quintiles or non-immigrant/immigrant status).⁴⁴

We use the M index since it has several desirable properties. This index satisfies Scale Invariance, Independence, the School Division Property, Symmetry, the Group

⁴²Note that we observe parental education at the census block level and immigrant status at the student level.

⁴³Results are available upon request.

⁴⁴The entropy of s is defined as $\sum_{k=1}^K q_s \log \frac{1}{q_s}$ where q_s is the proportion of the students in school s within the city of Madrid. The entropy of g is defined as $\sum_{k=1}^K q_g \log \frac{1}{q_g}$ where q_g is the proportion of the students in group g within the city of Madrid.

Division Property, and a technical continuity property (Frankel and Volij, 2011).⁴⁵ However, there are two limitations of this index. First, it is not Composition Invariant, which is a specific drawback of this index. Second, we can not directly interpret the values of this index, which corresponds to the general trade-off between intuitiveness and performance in choosing an index. Therefore, the values do not provide an intuitive meaning, but we can directly interpret changes in segregation associated with the policy reform.

A graphical analysis of the evolution of the levels of school segregation over years would provide evidence of the association between the school choice reform and school segregation. However, trends in school segregation may be correlated with other important factors, such as residential trends or demographic pressure. In order to investigate whether the reform is statistically significantly associated with changes in school segregation, we use the following model:

$$Seg_{dt}^s = \alpha + \delta_t + \nu_d + \gamma Seg_{dt}^{nm} + D'_{dt}\beta + \epsilon_{dt} \quad (3.4)$$

where Seg_{dt}^s is the school segregation index (either by parental education or by immigrant status) of school district d in year t ; δ_t are year fixed effects; ν_d are school districts fixed effects; Seg_{dt}^{nm} is the school segregation index (either by parental education or immigrant status) of district d in year t under pure residence-based assignment (i.e., if it were entirely due to residential segregation); D'_{dt} is a vector of time-variant school district characteristics, such as population density and average income; and ϵ_{ist} is the error term. We use robust standard errors. The variable Seg_{dt}^{nm} can be seen as a counterfactual measure of an extreme situation in which every student attends to the assigned school based on their primary residence (i.e., out-of-district assignment). This variable control for trends in residential segregation over time by groups, and improves the precision of our estimates. We abstract from causal interpretations of the year fixed effects in Equation 3.1 and Equation 3.4, and view these parameters as providing conditional associations.

⁴⁵For a summary on the properties satisfied by the different indices see Table 1 of Frankel and Volij (2011).

3.4.3 Identification Threats

In this subsection, we identify several potential identification challenges for our main specification and discuss their implications for our results.

First of all, we cannot capture parents' full response since we only observe the top-ranked school in the submitted list. The school choice assignment mechanism plays a vital role. A unique characteristic of the BM is the fact that first-ranked school is crucial for final allocation. [Abdulkadiroğlu, Che and Yasuda \(2011\)](#) highlight that one of the features of the BM is that it better takes the parents' cardinal preferences into account, compared to other algorithms with more desirable ordinal properties (e.g., Gale Shapley or Top Trading Cycles). Moreover, a specific feature of the school choice in Madrid (and Barcelona) is the fact that pupils' priority points are always measured for the top-ranked school, which provides an additional incentive to families to think carefully about their first-choice and potentially to apply to the school where they have high chances of being admitted. Table [3.B.3](#) shows the percentage of pupils who were assigned to their top-ranked school, those who went to a school they did not rank first, and those not assigned to any of their listed schools. It shows that about 86 percent of the students were assigned to the school they ranked first, and around 3 percent were not assigned to any of the schools that families listed. Hence, despite not being able to capture the full extent of families' behavioral responses, the top-ranked school is crucial under the BM to understand the most essential choice and provides a sizeable amount of information.

Another important threat to our identification is the residential sorting of families as a consequence of an anticipation to the reform, that may confound the out of school district application effect. The announcement of the reforms play a key role in order to explore this identification challenge. The inter-district school choice reform was a proposal in the electoral program of the political party that won the elections in the region of Madrid in May 2011.^{[46](#)} The information on the 2012 (and 2013) priority changes was disclosed to the press on February 2012 and announced

⁴⁶In the programme, the winning party in the 2011 election (the conservative party *Partido Popular*) included the following statement: *to establish full freedom of school choice, implementing a single zone for parents to bring their children to whichever school they want.*

formally through an administrative order in March 2012. Applications are handled between April and May of the academic year, implying that families did not have much time from March to May of 2012 to change their decisions. However, they had about one year between the formal announcement of the reform and the 2013 round of applications, which could raise concerns about anticipation. Such effects would create a problem for identification if families reacted before the implementation of the reform (e.g., changing their primary household residence). [Hastings, Kane and Staiger \(2009\)](#) find that the preferences for high ranked schools increase with proximity and that parents with higher income are more willing to commute farther away for such schools. This increases the demand for school districts with higher average public school performance, leading to a rise in the housing prices of those districts.⁴⁷ A relaxation of the proximity priority criteria may reduce those incentives, and potentially provide incentives for families to change their primary residence to districts with lower housing prices (typically positively correlated with lower school performance), since living in the school district of the desired school is not as decisive as previously in terms of admission probabilities. In this case, a potential positive significant effect on out of district application (and distance to school) may be correlated with this dispersion effect, leading to an upward bias in our estimates that would reflect an upper bound of the “true” effect. Table [3.C.4](#) is aimed to test this potential response. We observe that housing prices are positively correlated with school district average performance. However, we find no evidence of differential changes in housing prices associated with the policy reform, which seems to rule out housing anticipation effects.

⁴⁷See, for instance [Fack and Grenet \(2010\)](#).

3.5 Results

3.5.1 Out of District Choice and Assignment

3.5.1.1 Event Study: First Difference Approach

Average Effects. Figure 3.A.4 and Figure 3.A.5 plot the point estimates and confidence intervals of year fixed effects in Equation 3.1 for out of district applications and assignment, and average travel distance to the top-ranked and assigned school respectively. The inter-district school choice reform, and (to a lesser extent) the 2012 reform seem to be associated with a sharp increase in these two variables. These changes seem to be reasonably stable over time. The fraction of out of district applications before the reform was 0.10. This relatively low pre-reform level may be explained by the fact that BM induces high residential sorting, though most of the families exert school choices through the primary household residence choice. The 2012's reform is associated with an increase in out of district school applications of 1.1 percentage point, which is equivalent to an increase of 11% with respect to the baseline mean of 2011. The 2013's inter-district reform is associated with a further increase in the fraction of out of district applications of 3 percentage points (30% with respect to the baseline mean). The figure shows that the effects on the probability of top-ranking out of district school almost equals the impact on the probability of being assigned to an out of district school. This strong association may be explained mechanically by the fact that the vast majority of families get their first-choice under the BM. The baseline average travel distance to the top-ranked school was 1,2 kilometers in 2011. The 2012 reform is associated with an increase in this distance of 33 meters (3% with respect to the baseline mean), and the inter-district school choice reform further increased it by 259 meters on average (22% with respect to the baseline mean).

Heterogeneous Effects by Parental Education and Immigrant Status. Figure 3.A.6 shows the effect of the reform on out-of-district assignment by parental education quintiles. The inter-district school choice reform seems to be associated

with a heterogeneous increase in the out of district assignment by parental education. Interestingly, despite the fact that families with higher levels of education reacted more to the reform in absolute terms, the effects were fairly similar across parental education quintiles in relative terms. The baseline probabilities of being assigned to an out of district school in 2011 were substantially different: 0.04, 0.11, and 0.14 for parental education quintile one, three, and fifth, respectively. The inter-district school choice reform was associated with an increase in the out-of-district assignments of 1, 3, and 5 percentage points, but the effect size of these increases was 26, 31, and 31 percent respectively. A different situation occurs when considering heterogeneous effects by immigrant status (see Figure 3.A.7). The increase in the out of district assignment is entirely driven by families with Spanish children since parents of immigrant children do not seem to have reacted at all in the first years of the reform. However, the figure shows an increase in the out-of-district assignment in 2015 and 2016 for immigrants, which may suggest information gaps between immigrants versus native families groups.

Heterogeneous Effects by Distance to the Closest Border. Figure 3.A.8 presents the effect of the reform on out-of-district assignment by distance to the closest school district border. We find substantial heterogeneity in the probability of being assigned to an out of district school conditional on the distance of the households' primary residence to the closest school district border. While families who live within 850 meters or more get an out-of-district assignment in four percent of the cases, parents whose primary residence is located within 300 meters get 21 percent. The effect of the reform is larger for parents who live closer to the school district boundaries in absolute terms. However, families who lived further away reacted more in relative terms (40% and 14% with respect to the baseline mean for families living within more than 850 and less than 300 meters respectively). These results imply that the reform expanded the out of district assignment throughout the entire spacial dimension, although with larger relative impacts for those families living further away from school district boundaries.

To which schools are pupils assigned? Families with lower levels of education

quintiles managed to get admission in schools with higher average performance (although the size of the effect is quite small), middle parental education quintiles increased their propensity to get access to schools in districts with higher average income levels. Most of the families managed to get access to districts with lower predetermined levels of school segregation (measured by parental education status). There are no substantial differences between immigrants and natives on admission by school average performance and predetermined school segregation (measured by immigrant status), but immigrants tend to get admission in schools in which the average income is lower on average.⁴⁸ The fact that most of the families got access to a slightly higher performing school is potentially correlated with the constant supply of schools, but decreasing demographic trends of pupils of the age of 3, during the period under study in the city of Madrid.

3.5.1.2 Difference-in-Difference Analysis

Table 3.B.4 shows the main results of the DID strategy. It is the result of estimating equation Equation 3.2. Columns (1) to (3) report three specifications for out-of-district assignment, while columns (4) to (6) provide the same three specifications for distance to assigned school. Column (1) and (3) present the baseline estimates including school district fixed effects, columns (2) and (4) include census block level demographic controls, and columns (3) and (6) add households observable characteristics. The effect of the inter-district school choice reform is positive and statistically significant in all specifications. The effect size is an increase of 3.6 percentage points for out-of-district assignment 311 meters for distance to the assigned school. Considering that the average fraction of out-of-district assignment (distance to assigned school) in 2011 was 0.05 (1,259 meters) for families whose primary residence locates within 150 meters to the school district centroid, an increase of 3.6 percentage points (311 meters) is equivalent to a 70% increase (25%). Results are robust to the sensitivity choice in the definition of the treatment group, such as parents whose primary residence locates within 200, 250, 300, 350 and 450

⁴⁸See Section 3.C.11 for details.

meters close to the school district boundaries (see Table 3.C.7).

Overall, there are two significant findings. First, on average, the reform implied a positive households' response for the distance traveled and out-of-district school assigned. Second, the effect was not homogeneous across parental education status and between natives and non-natives. In particular, students from the most educated households and native students were the ones that reacted the most to the reform in absolute terms, but in relative terms, the effect size of the reform was fairly similar. These results are consistent with previous evidence from reforms of other countries (e.g., Sweden⁴⁹), showing that less advantaged and immigrant households tend to exert less choice than more socially advantaged and natives families when choosing schools (which may reflect differences in preferences, information). This could have distributional implications that are further analyzed in the next section. Finally, the absolute magnitude of the change in terms of the total student population is modest. It is probably due to the weight of other key contextual factors, such as the 0.5 point that families continue to obtain when they reside in the district of their first-choice application, the disutility from commuting, and the fact that the extent of the effect is expected to be small due to the low risk behavior in the top-ranked school induced by the BM.

3.5.2 School Segregation

We have analyzed the impact of the reform on the probability that the top-ranked and assigned school are located in a different school district from where families reside, finding an increase in the out-of-district assignment. This subsection is devoted to investigate whether this reshuffle of pupils had a contemporaneous impact on school segregation by parental education and immigrant status. We explore the trends in school sorting over time, being able to compare the within vs. between district school segregation, and the school vs. residential segregation. We conclude with an econometric analysis to explore conditional associations between the school choice reform and student sorting.

⁴⁹See Böhlmark and Lindahl (2007).

Trends in School Segregation. Figure 3.A.9 plots the M Index of school segregation by parental education and immigrant status. Standard errors are computed using the bootstrap resampling method with 100 replications. Despite the small increase in the level of segregation between schools by parental education in 2012, there is a decreasing trend over time following the inter-district school choice reform. School segregation by immigrant status shows the opposite pattern. There is a sharp increase in the level of segregation in 2012 and 2013 with a decreasing thereafter that ends at slightly higher levels than in 2010 and 2011.

Within vs. Between School District Segregation. One of the advantages of the M Index of segregation is that it is additively decomposable. This property allows to divide the value of the index into two components: (i) The share of segregation that comes from the school segregation between different groups (e.g., between school districts), which we refer as between district school segregation; (ii) A component that measures the part of segregation that comes from within groups school segregation (e.g., within school districts). The addition of between and within district school segregation adds up to the total school segregation. Most of the indices that are widely used in the literature, such as the Dissimilarity Index, do not have this property. Figure 3.A.10 investigates the extent through which the dynamics of the total school segregation are driven by within and between school district segregation. Interestingly, whereas the level of between district segregation is higher than the within district segregation by parental education, the opposite pattern is right when we consider segregation by immigrant status. It seems that the decreasing trend in school segregation by parental education is driven mostly by the decrease in within district segregation, while between district segregation remains flat after the inter-district school choice reform. Both between and within district segregation by immigrant status display a similar trend, but within district, segregation shows a larger increase.

School vs. Residential Segregation. An important confounding factor when addressing the effects of the reform on school segregation is the trends in residential segregation. Changes in segregation between schools may be correlated

with changes in the social composition of neighborhoods, especially when analyzing a cross-cohort comparison. An advantage of this study is that we count with the precise geo-location of families' home address, allowing to compute an accurate measure of residential sorting. Residential segregation is measured as the M index of families' primary residence segregation by parental education and immigrant status between neighbourhoods. Figure 3.A.11 plots the trends in school and residential segregation. Although residential segregation is higher than school segregation by parental education, the opposite occurs when we consider segregation by immigrant status. This stylized fact is potentially correlated with the findings of Figure 3.A.7, in which we find that immigrants are less likely to top-rank a school located in a different district than natives. This may create a situation in which the levels of school segregation by immigrant status are higher than the residential ones. It is not as clear for segregation by parental education. Since every quintile of parental education move to some extent, the equilibrium outcome is more uncertain. Interestingly, we find that the decrease in school segregation after the inter-district school choice reform is not followed by a change in residential sorting since the latter remains flat after 2013. However, the increasing pattern in school segregation by immigrant status seems to be highly correlated with changes in residential segregation, which may act as a confounding factor when identifying the effect of the reform. Controlling for this variable when analyzing associations between the inter-district school choice reform and school segregation would improve the precision of our estimates, especially when considering school segregation by immigrant status.

The Effect of the School Choice Reform on School Segregation. In order to perform a more formal test of the impact of the reform on the levels of school segregation, we estimate the model described by Equation 3.4, which controls for school districts' fixed effects, segregation under pure residence-based assignment, and time-variant school district characteristics. Table 3.B.5 shows the regression results. The 2012 reform is associated with a mild increase in school segregation by parental education of 3%, but a sizeable increase in segregation by immigrant status by 9% (see Figure 3.A.12 and Figure 3.A.13). Both impacts are statistically

significant at the 1 and 5 percent confidence levels, respectively. By contrast, the inter-district school choice reform of 2013 is associated with a reduction in the level of segregation by parental education of about 1.8%. Although the estimates are not statistically significant for 2013 and 2016, they are statistically significant at the 1 percent confidence level for 2014 and 2015 (depending on the specification), but more importantly, the point estimates are always negative and of similar size. On the contrary, the inter-district school choice reform is associated with an increase in the level of segregation by immigrant status of 1.6%, although the estimates are not statistically significant and point estimates for years after that 2013 are close to zero. It seems that residential segregation explains a large part of the variation in school segregation by immigrant status (as suggested by Figure 3.A.11).

To sum up, it seems that the 2012 reform, in which priorities for low-income pupils were reduced, and new priority was granted for alumni school family members, is associated with a contemporaneous increase in the levels of school segregation by both parental education and immigrant status. However, the inter-district school choice reform is associated with a decrease in school segregation by parental status, and an increase in the school segregation by immigrant status, though results in the latest are not robust to different specifications.

3.6 Robustness Check: Phasing-in of the Reform in other Municipalities

As we described in Section 3.2, the school choice reform was gradually implemented in other municipalities beyond the city of Madrid. The inter-district reform was implemented in several municipalities (usually those of medium size) in 2012 (as well as the low-income and alumni pupil bonus criteria), whereas for the larger ones (including Madrid), it took place in 2013. Figure 3.A.14 shows a map of the phase-in of the reform in the entire set of municipalities in the region of Madrid.⁵⁰

We exploit this phasing-in of the inter-district school choice to analyze the effects on

⁵⁰Table 3.C.2 in the Appendix describes which municipalities joined the inter-district criterion in 2012 and 2013.

the other municipalities as a robustness check. In particular, we compare the results between three groups of municipalities: small-size (no reform change), medium-size municipalities (the reform took place in 2012/2013) and large-size municipalities (the reform took place in 2013/2014)

3.6.1 Out-of-Municipality Assignment and Distance to Assigned School

In terms of the out of district assignment, we measure the distance to the assigned school given that we do not have access to geographical data on the catchment areas in other municipalities than Madrid city before the reform.⁵¹ Besides, we compute the out-of-municipality assignment as a proxy of an extensive margin measure. This variable is defined as a dummy variable that takes value 1 if the pupil is assigned to a school located in a different municipality, and zero otherwise. Figure 3.A.15 shows the average out-of-municipality assignment and the distance to assigned school by municipality size. Medium-size municipalities seem to increase both their out-of-municipality assignment and their distance to the assigned school in 2012 when both reforms were implemented. Interestingly, households living in large-size municipalities do not seem to change their out-of-municipality assignment but react increasing their distance the assigned school. Families living in small-size municipalities present the highest values of out-of-district assignment and distance to the assigned school. These households seem to be also changing their pattern, increasing both margins with the 2012's reform. The change in these trends for small-size municipalities can be explained by the fact that school choice reform relaxed the proximity bonus for intra-municipality school choice, but also for inter-municipality.⁵² As the educational markets at the municipal level are not independent, the relaxation of the inter-municipality bonus points may generate

⁵¹In Madrid, the choice catchment areas coincide with the city district. However, in other municipalities, the catchment areas were defined using other criteria. We do not count with data on catchment areas in other municipalities than Madrid, which implies that we cannot compute the out of school district assignment. However, we can compute the average commuting distance in other municipalities.

⁵²See Table 3.B.1.

interactions between several municipalities.

Figure 3.A.16 shows the effect of the school choice reforms on these variables, confirming the detailed results found in Figure 3.A.15. Panel A shows that the low-income and alumni reform and the inter-district school choice reform increased the out-of-municipality assignment (average distance traveled to the top-ranked school) in 50% (12%) with respect to the baseline mean for those municipalities where the inter-district choice was implemented in 2012 (i.e., those of medium-size municipalities). The school choice reforms increased the out-of-municipality assignment (average distance traveled to the top-ranked school) in 37% (8%) with respect to the baseline mean for those municipalities that had a single school choice zone (i.e., those of small-size municipalities). We observe no effect of the low-income and alumni bonus points reform in the out-of-municipality assignment, and a slight increase in the average distance traveled to the top-ranked school (13%) for large-size municipalities, that is consistent with the results from in Figure 3.A.4 and Figure 3.A.5. In 2013, the reform leads to a 30% increase (with respect to the baseline mean) in the commuting distance to the top-ranked school in large-size municipalities, which cached up with the absolute effect of those of medium-size municipalities, remaining constant over the following years. Overall, we find that the inter-district school choice reform is associated with an increase in the commuting distance irrespective of whether families reside in large, medium, or small size municipalities.

3.6.2 School Segregation by Immigrant Status

We can also exploit the gradual implementation of the policy to estimate changes in school segregation associated with the reform. It seems that despite the fact that large-size and middle-size municipalities present similar levels of school segregation by immigrant status, school segregation seems to have increased more on large-size municipalities in the reform years (see Figure 3.A.17). Figure 3.A.18 shows that this raise in school segregation is driven by the within municipality school segregation, which increasing seems to be more salient in large-size municipalities. Interestingly, the between municipality school segregation seems to be almost negligible for large-size

and middle-size municipalities, but it is almost equally important as the within municipality segregation in small-size municipalities.

To determine whether the inter-district school choice reform had a differential impact on school segregation by immigrant status, we regress school segregation by immigrant status at the municipal level on year and municipality fixed effect. Results are summarized in Figure 3.A.19. The implementation of the inter-district school choice reform seems to be associated with a small increase in school segregation by immigrant status, but results are clearer and more precise in municipalities of larger size. We cannot reject the null hypothesis of no change in school segregation in small-size municipalities. Results are robust to those found in the previous section for the city of Madrid.

3.7 Conclusion

In this paper, we use novel administrative data of the region of Madrid to analyze the impact of two school choice reforms, a large-scale inter-district school choice reform and a reduction in low-income pupils' priorities to the top-ranked school, on school choice and school segregation. We find that the inter-district school choice reform increased the probability of applying and of being assigned to an out-of-district school of parents, with a larger response for families with higher levels of education and non-immigrant children. We show that parents of immigrant children did not react at all to the reform, while parents from every education quintile increased their willingness to travel to the assigned school in absolute and relative terms. We find that the inter-district school choice reform is associated with a decrease in school segregation by parental education and an increase in school segregation by immigrant status. However, the decrease in low-income pupils' priorities and a new priority granted for alumni family members to the top-ranked school, is associated with a contemporaneous increase in the levels of school segregation by both parental education and immigrant status.

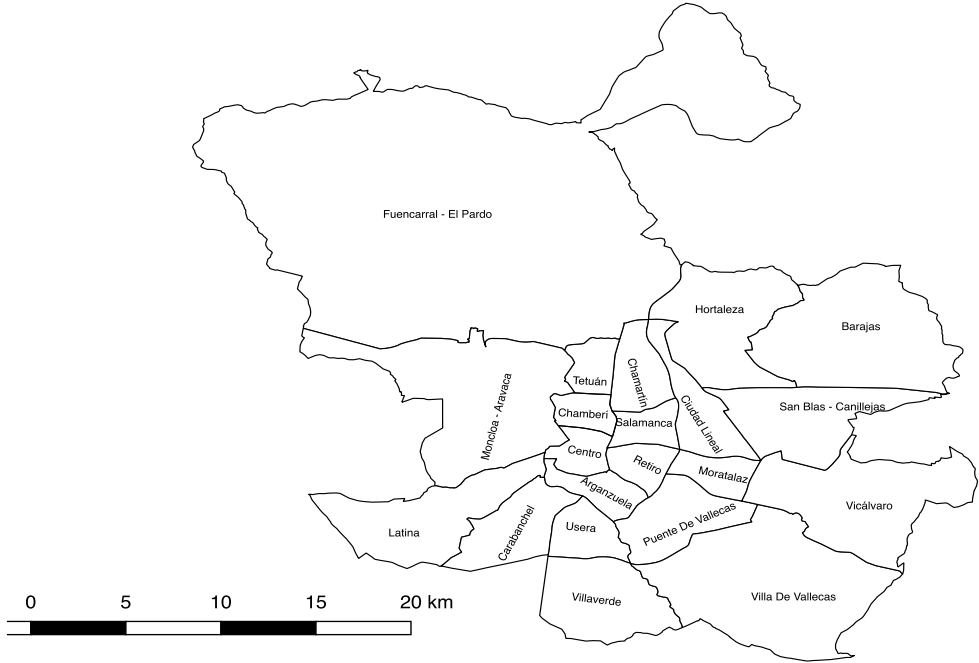
The levels of school segregation before the reform may matter to predict the

effects of relaxing residence-based priorities. The context of this paper is one of relatively high levels of school and residential segregation by socioeconomic background, and of relatively low levels of segregation by immigrant status. In a system with complete residential (and school) segregation, we may expect that increasing choice opportunities lower segregation ([Jenkins, Micklewright and Schnepf, 2008](#)), which is consistent with our results for school segregation by parental education. The opposite may be expected when predetermined levels of segregation are relatively low, and the different subgroups of the population under study differ in their levels of information or preferences for what constitutes a good school. Immigrants may lack the networks and information that are necessary to make an informed choice, or they may prefer to choose the nearest school for cultural reasons. The different responses in out-of-district assignment by immigrant status may explain the increase in the level of school segregation by immigrant status.

We find evidence suggesting that relaxing residence-based priorities may be an effective policy to increase pupils' inter-district school assignment, but the heterogeneous behavioral responses, especially for immigrant parents, needs to be taken into account. Even under a context that induces to low levels of inter-district choice and non-risky behavior, in which we expect a low elasticity of response with respect to changes in school priorities, school segregation may, in fact, be affected, and these effects may have opposite directions depending on the predetermined level of sorting. Results imply that when parents' school choices exhibit a strong degree of polarization by social and immigrant background, priority structures needs to be carefully designed to achieve diversity objectives. Further research need to be undertaken to understand under which conditions, such as the predetermined levels of school (and residential) segregation, subgroups of the population considered, school choice allocation mechanism or parents' preferences, school priorities may contribute to reduce or to increase school segregation.

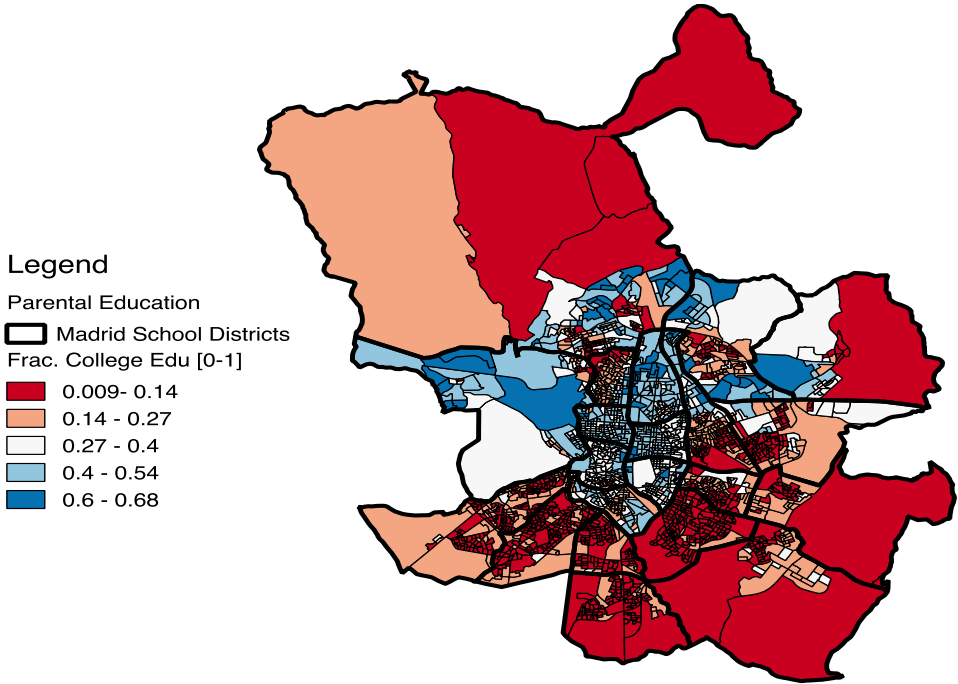
Appendix 3.A Main Figures

Figure 3.A.1: School Districts in the city of Madrid.



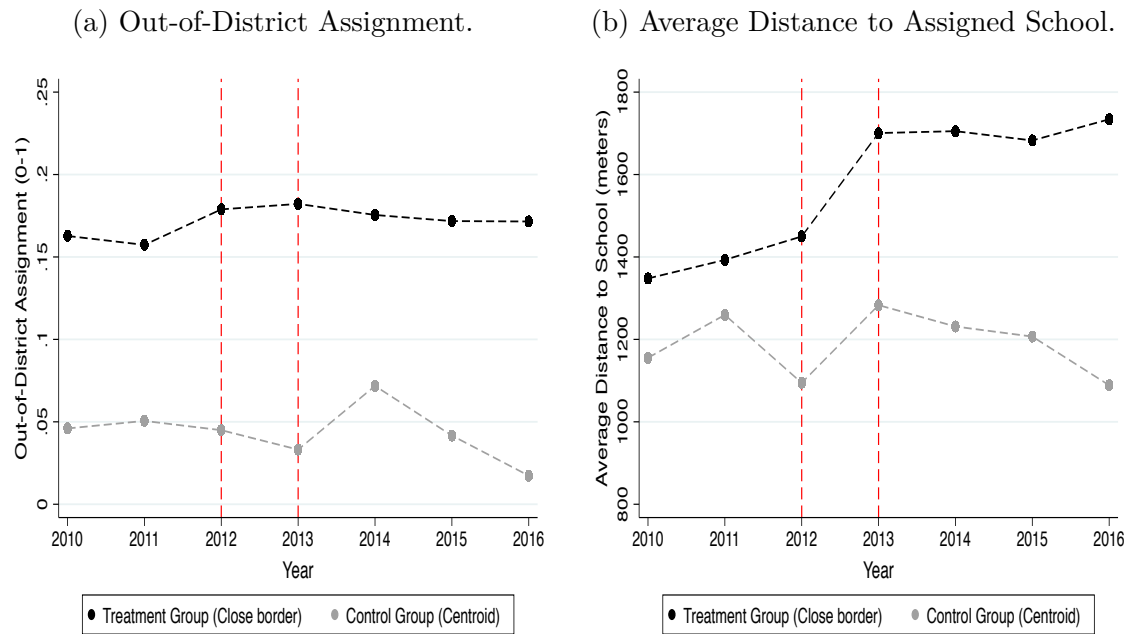
Note: Own computation using shapefiles data from the 21 school districts of the city of Madrid.

Figure 3.A.2: Parental Education by census block in 2013.



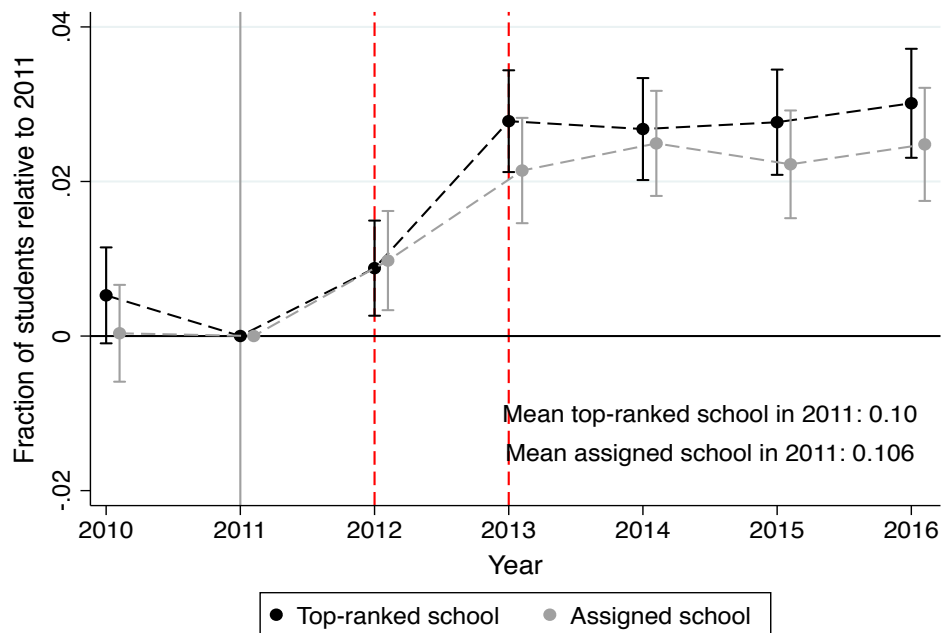
Note: Own computation using shapefiles data from the 21 school districts of the city and Census Office data of Madrid.

Figure 3.A.3: Out-of-District Assignment and Average Distance to Assigned School by group of treatment.



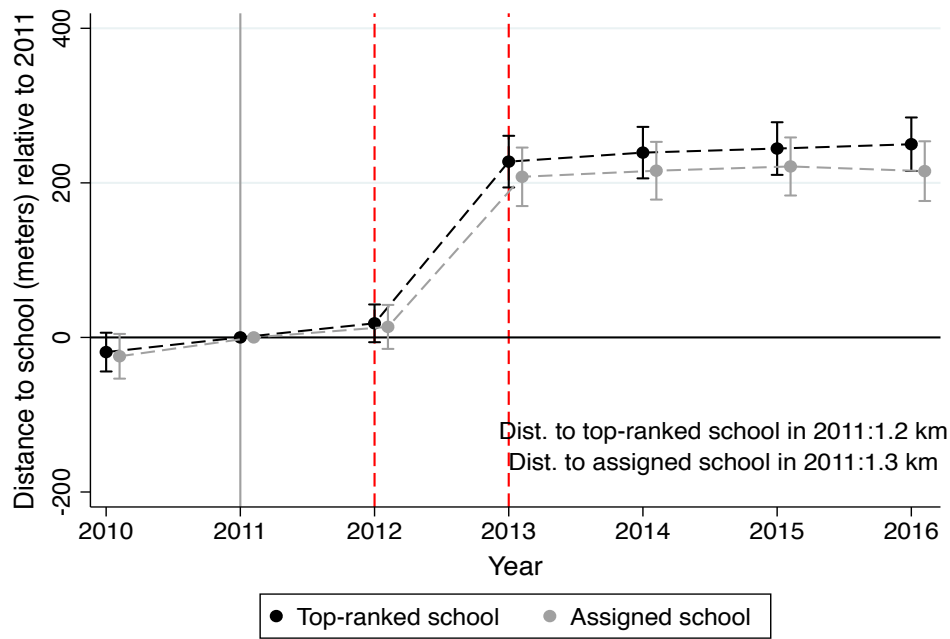
Note: Black and grey dots represent averages. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.A.4: Outer School District Applications: Top-Ranked School vs. Assigned School.



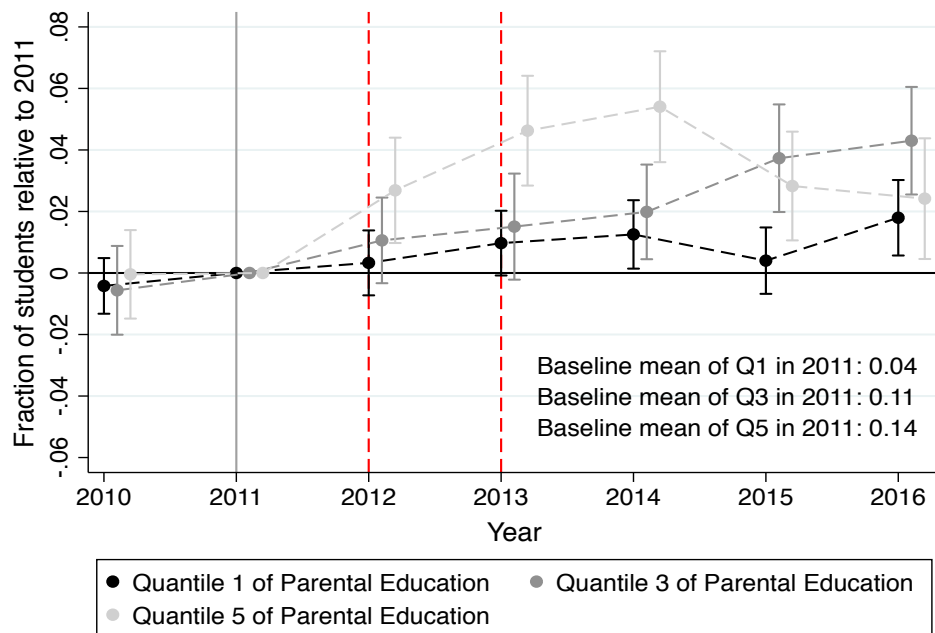
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.5: Average Distance to Top-Ranked School vs. Assigned School.



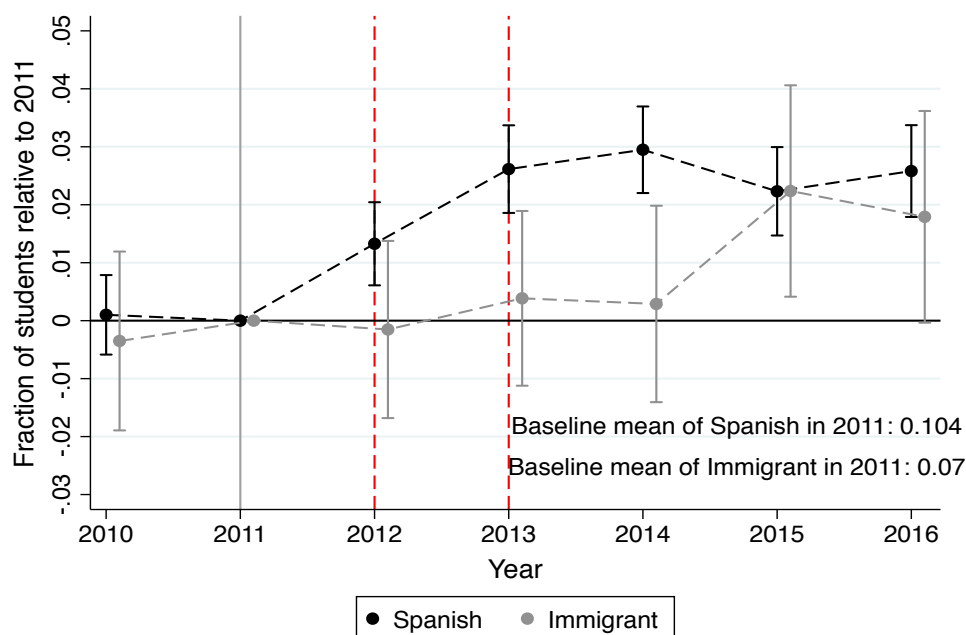
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.6: Impact of the Reform on Out-Of-District Assignment by Parental Education.



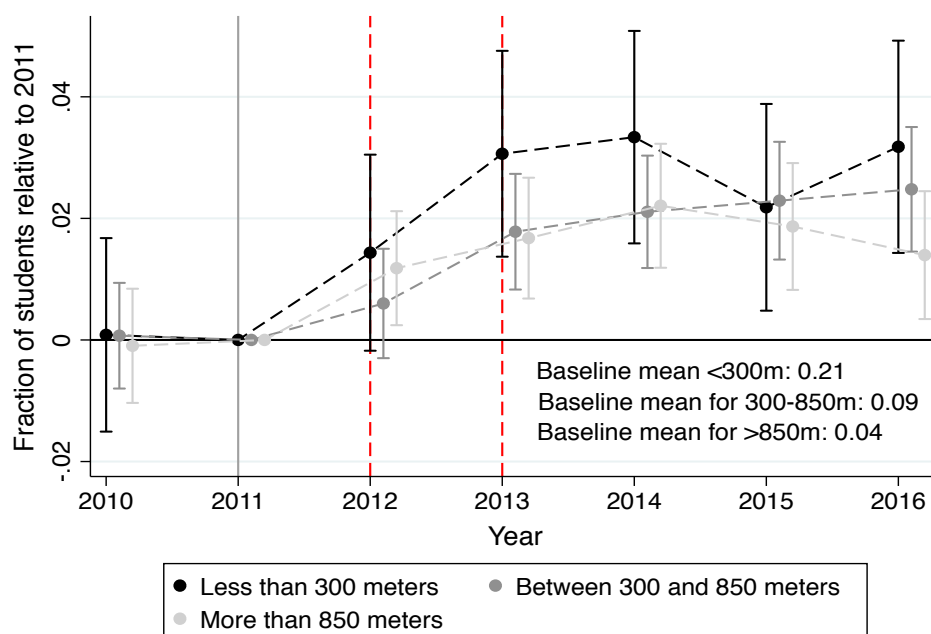
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.7: Impact of the Reform on Out-Of-District Assignment by Immigrant Status.



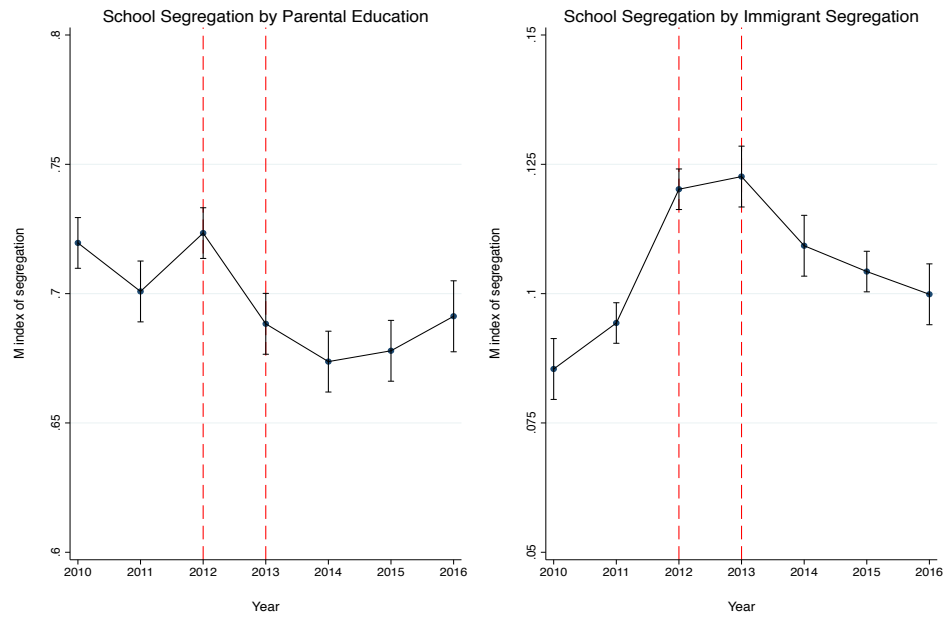
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.8: Impact of the Reform on Out-Of-District Assignment by Distance to the Closest School District Border.



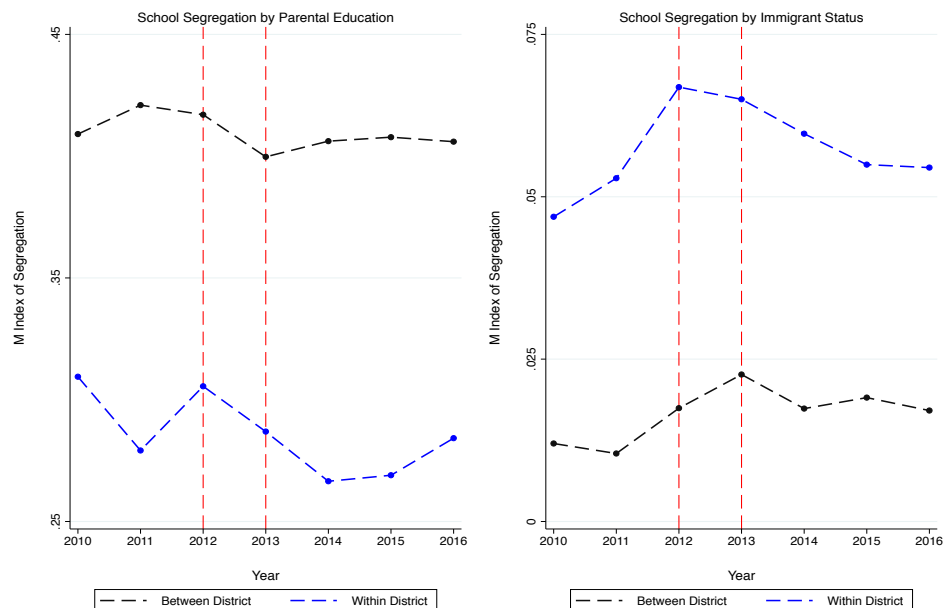
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.9: School Segregation (M Index).



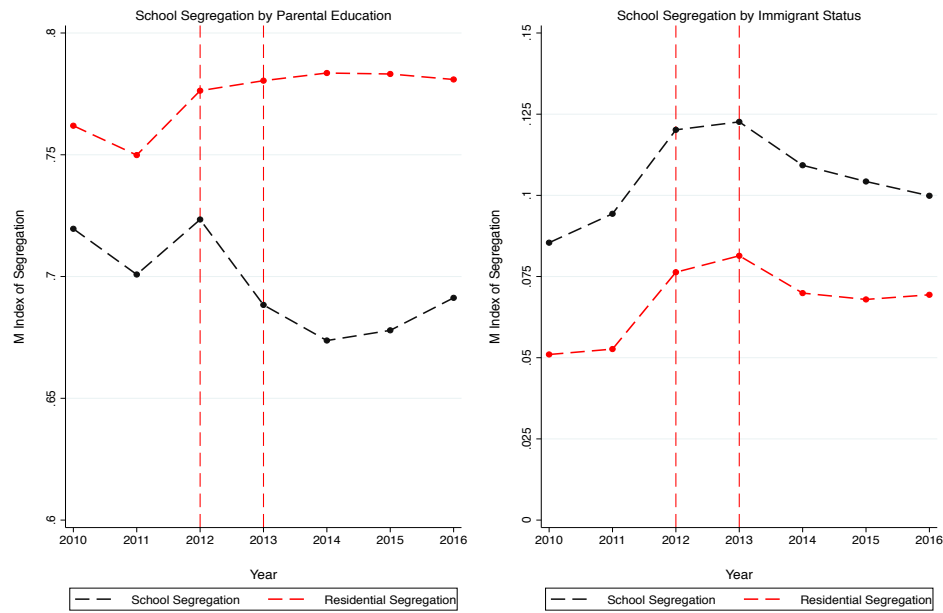
Note: The Figure shows the M index of segregation. The left hand side Figure represents the school segregation measured by parental education. The right hand side of the Figure plots the school segregation by immigrant status. Blue dots are the estimates of school segregation using the Mutual Information Index of Segregation. Horizontal lines depict confidence intervals at the 95% level of confidence using bootstrap with 100 replications. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.A.10: Within vs. Between School District Segregation.



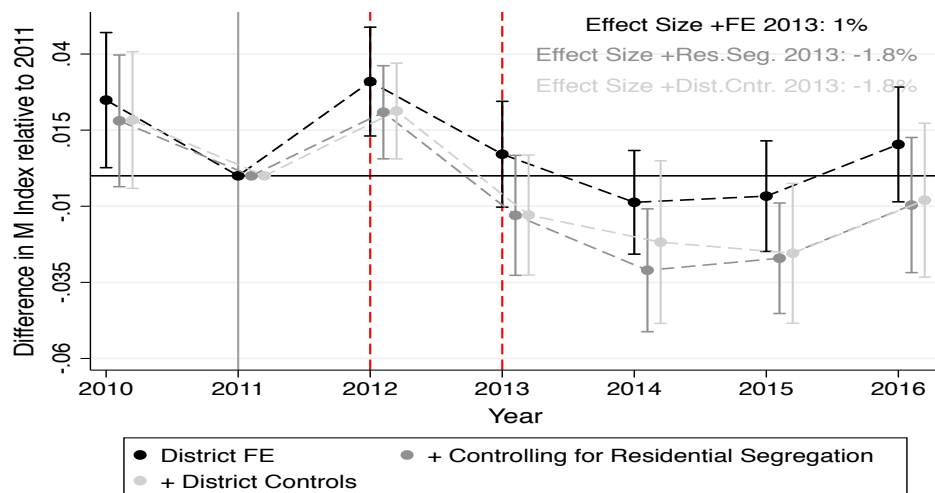
Note: The Figure shows the M index of segregation. The left hand side Figure represents the school segregation measured by parental education. The right hand side of the Figure plots the school segregation by immigrant status. Blue dots are the estimates of within school segregation using the Mutual Information Index of Segregation. Black dots are the estimates of between school segregation using the Mutual Information Index of Segregation. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.A.11: School vs. Residential Segregation.



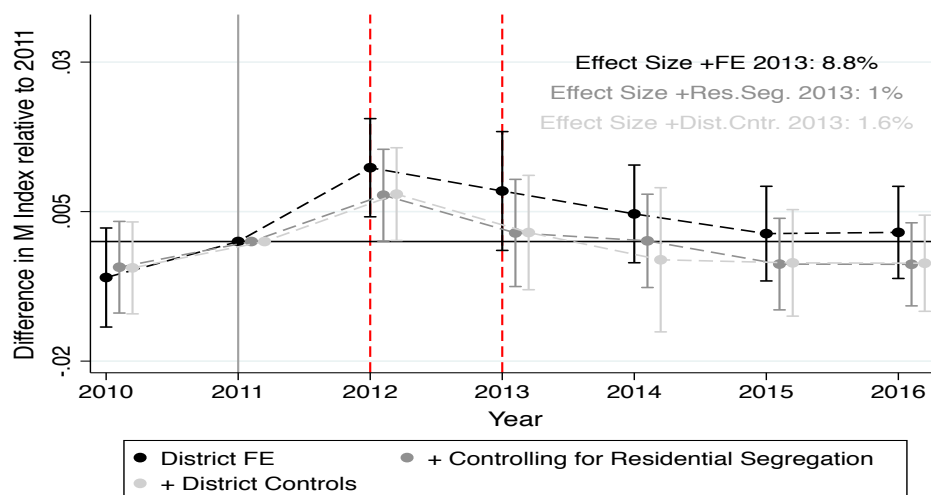
Note: The Figure shows the M index of segregation. School segregation is measured as the M index of pupils segregation by parental education and immigrant status between schools. Residential segregation is measured as the M index of families' primary residence segregation by parental education and immigrant status between neighbourhoods. The left hand side Figure represents the school segregation measured by parental education. The right hand side of the Figure plots the school segregation by immigrant status. Black dots are the estimates of school segregation using the Mutual Information Index of Segregation. Red dots are the estimates of residential segregation using the Mutual Information Index of Segregation. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.A.12: The Effect of the School Choice Reform on School Segregation by Parental Education.



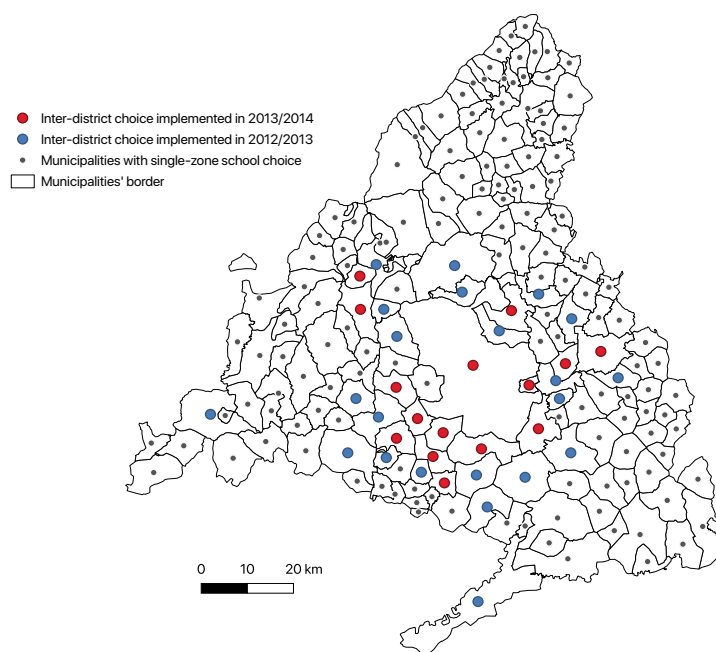
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.13: The Effect of the School Choice Reform on School Segregation by Immigrant Status.



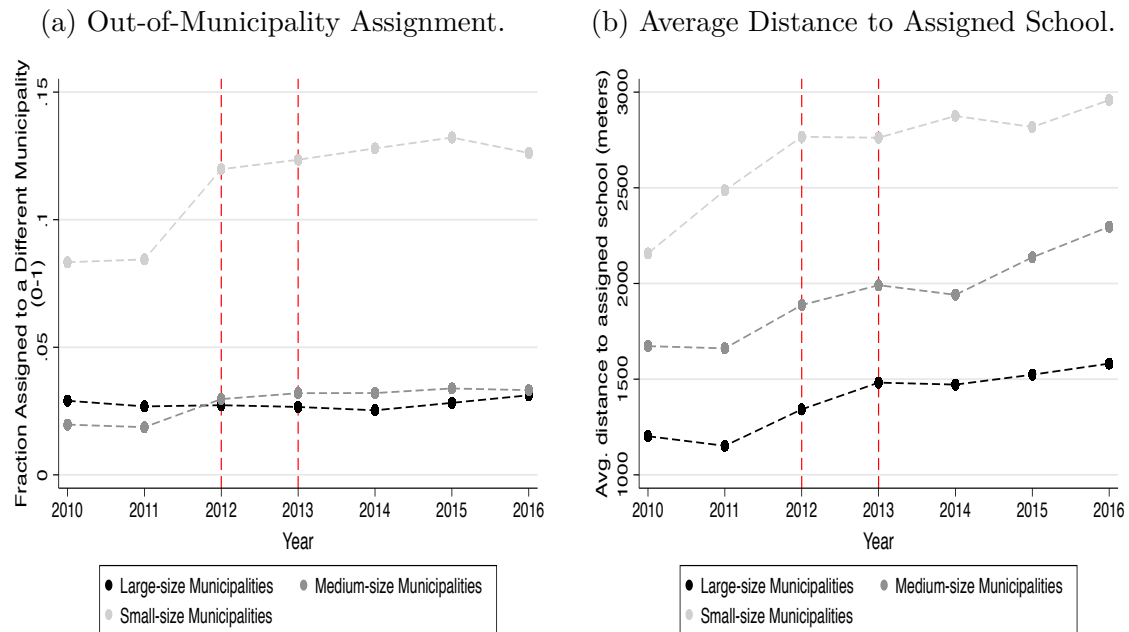
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.14: Phasing-in of the Reform in the Municipalities of the Region of Madrid.



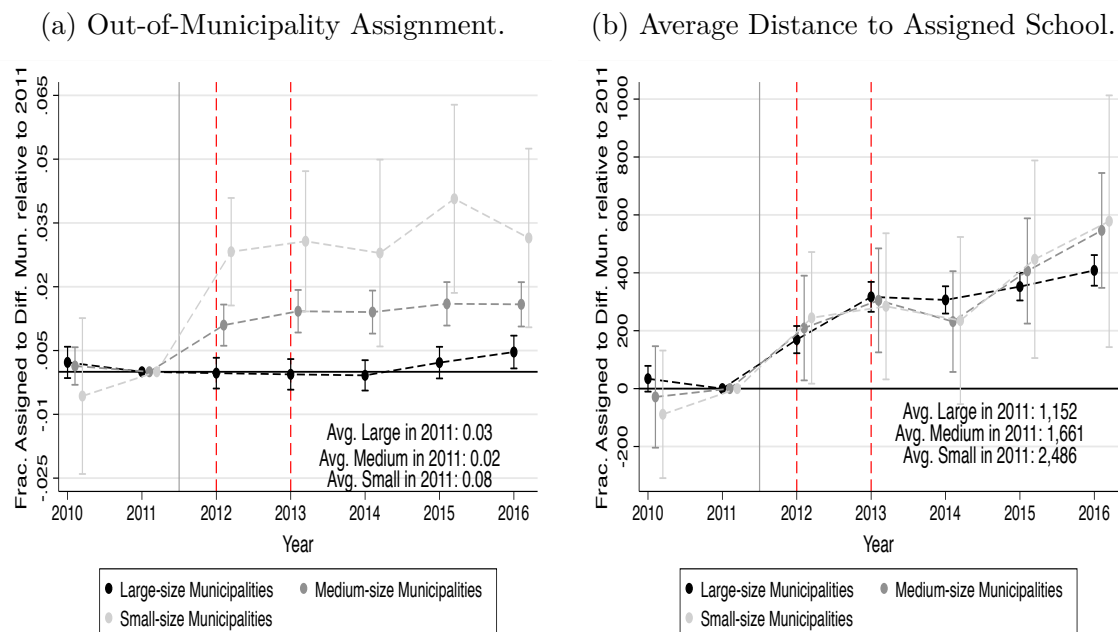
Note: Own computation using shapefiles data from the 179 municipalities of the region of Madrid.

Figure 3.A.15: Average Out-of-Municipality Assignment and Average Distance to Assigned School by municipality size.



Note: Black and grey dots represent averages. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

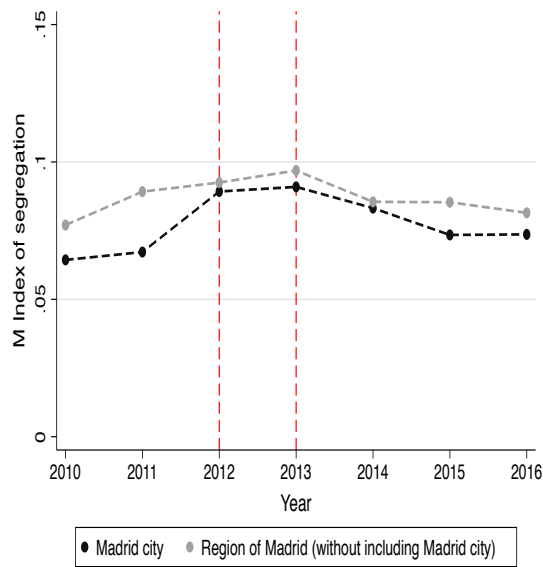
Figure 3.A.16: Effect of the School Choice Reform on Out-of-Municipality Assignment and Average Distance to Assigned School by municipality size.



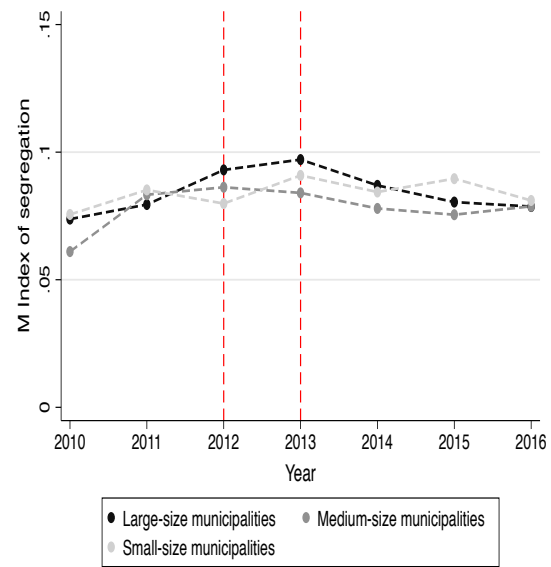
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Figure 3.A.17: School Segregation by Immigrant Status by Municipality Size.

(a) Madrid city vs. Rest of Municipalities.

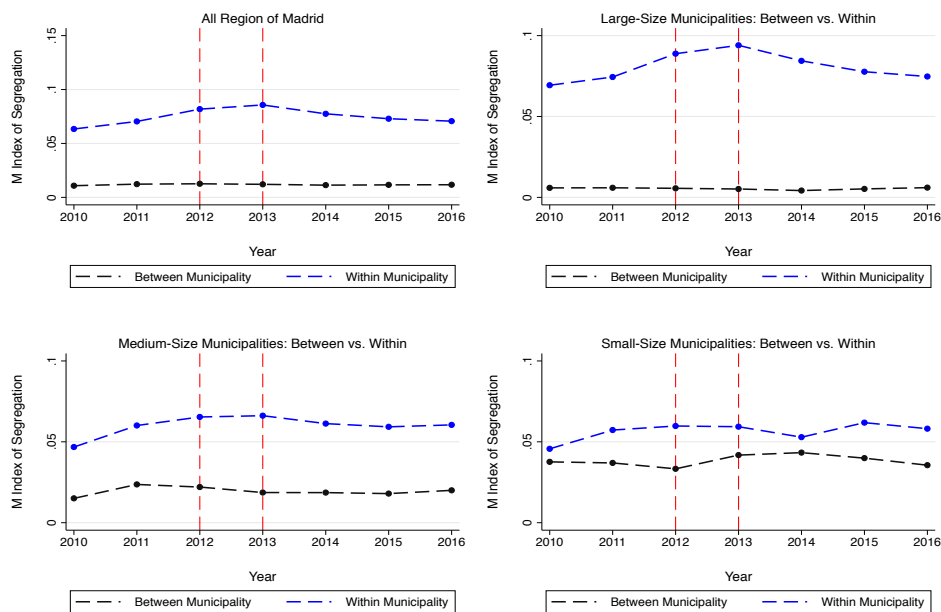


(b) Municipality Size



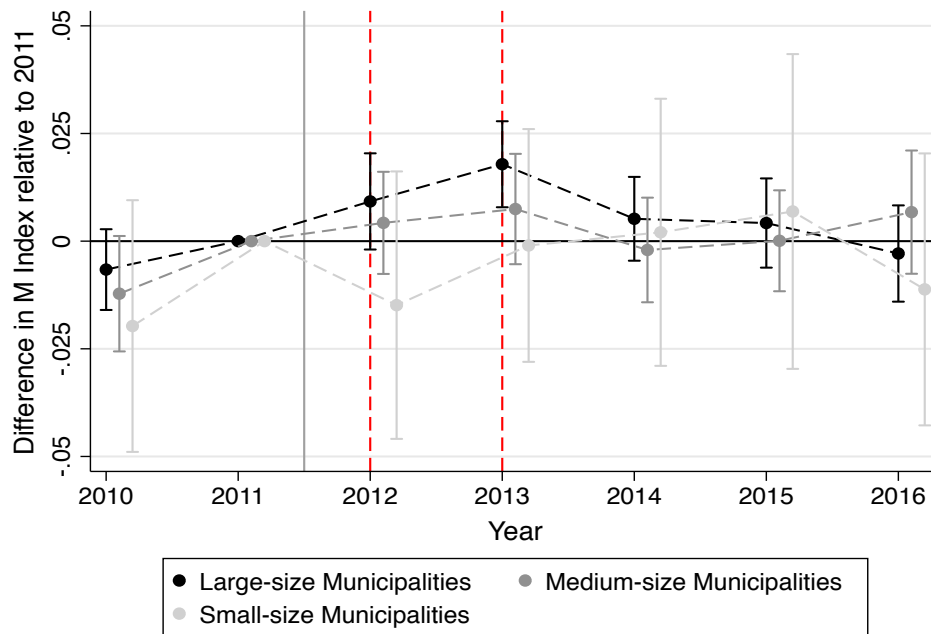
Note: The Figure shows the M index of segregation. The dots represents the school segregation measured by immigrant status. The left hand side Figure represents the school segregation in Madrid city and in the rest of the Region of Madrid. The right hand side of the Figure plots the school segregation by municipality size. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.A.18: School Segregation by Immigrant Status: Within vs. Between Municipality.



Note: The Figure shows the M index of segregation. Blue dots are the estimates of within school segregation using the Mutual Information Index of Segregation. Black dots are the estimates of between school segregation using the Mutual Information Index of Segregation. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.A.19: The Effect of The School Choice Reform on School Segregation by Immigrant Status and Municipality Size.



Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. The coefficients are normalized to zero in 2011.

Appendix 3.B Main Tables

Table 3.B.1: Priority Points in case of over-demand of schools in the city of Madrid.

BONUS	CRITERIA	NUMBER OF POINTS		
		Before 2012/2013	2012/2013	2013/2014
Proximity -Madrid city-	Family house or parents' work in:			
	School district	4	4	
	Boundary school district	2	2	
	Family house or parents' work in:			
	Same municipality			4
	School district			0.5
	Region of Madrid			2
Low-income	Income \leq IPREM	2		
	IPREM $<$ Income \leq 2IPREM	1		
	Minimum Insertion Subsidy		2	2
Siblings	First one 4pts, and additional 3pts	4		
	One or more		8	10
Disability	Parents, students or siblings	1.5	1.5	1.5
Large Family	General	1.5	1.5	1.5
	Special	2.5	2.5	2.5
Alumni family member	Family member is alumni student		1.5	1.5
School discretionary		1	1	1

Notes: The changes beyond the proximity criteria were applied together across all medium and large municipalities. IPREM is the acronym in Spanish for the Multiple Effects Income Public Index, which was €7,455.14 in the period of study. The Minimum Insertion Subsidy (*Renta Mínima de Inserción*) is a special provision granted for people with lower income than IPREM. School discretionary is a point that the schools have freedom to assign based on “public and objective” criteria.

Table 3.B.2: Summary Statistics: School applicants in the city of Madrid over 2010-2016.

Variable	2010	2011	2012	2013	2014	2015	2016
A. Grades in City of Madrid							
pre-school age of 3	18,391 [0,62]	18,289 [0,64]	18,006 [0,62]	16,970 [0,6]	16,323 [0,59]	16,266 [0,62]	15,696 [0,62]
pre-school Age 4	2,738 [0,09]	2,571 [0,09]	2,850 [0,1]	2,746 [0,1]	2,556 [0,09]	2,268 [0,09]	1,418 [0,09]
pre-school Age 5	2,087 [0,07]	1,925 [0,07]	2,156 [0,07]	2,306 [0,08]	2,109 [0,08]	1,769 [0,07]	1,000 [0,07]
Primary 1th grade	4,254 [0,14]	3,946 [0,14]	3,973 [0,14]	4,121 [0,14]	4,315 [0,16]	3,899 [0,15]	2,688 [0,15]
Primary 2 to 5th grade	1,602 [0,05]	1,373 [0,05]	1,539 [0,05]	1,618 [0,06]	1,559 [0,06]	1,451 [0,06]	1,123 [0,06]
Primary 6th grade	664 [0,02]	596 [0,02]	681 [0,02]	717 [0,03]	717 [0,03]	618 [0,02]	542 [0,02]
B. Students characteristics - pre-school age of 3- City of Madrid							
Immigrant	0.116	0.131	0.167	0.158	0.145	0.137	0.132
Female	0.491	0.493	0.484	0.487	0.500	0.489	0.487
Quintile 1	0.206	0.211	0.206	0.203	0.203	0.208	0.203
Quintile 2	0.208	0.200	0.205	0.203	0.199	0.202	0.199
Quintile 3	0.199	0.206	0.208	0.203	0.204	0.201	0.204
Quintile 4	0.199	0.200	0.196	0.197	0.202	0.197	0.202
Quintile 5	0.188	0.181	0.184	0.191	0.187	0.189	0.187
Distance to Closest District Border (meters)	463.7	454.2	627.0	626.5	632.4	633.0	625.3
Distance to Closest School (meters)	177.3	164.3	231.7	230.5	231.7	234.2	239.9
Number of Schools in a Radius of 500 meters	3.696	3.757	3.686	3.678	3.661	3.649	3.623
Observations	18,391	18,289	18,006	16,970	16,323	16,266	15,696
C. Students characteristics - pre-school age of 3- Region of Madrid (without Madrid city)							
Immigrant	0.143	0.158	0.166	0.164	0.153	0.140	0.147
Female	0.487	0.486	0.486	0.490	0.485	0.483	0.489
Distance to Closest Municipality Border (meters)	1,454	1,439	1,436	1,431	1,438	1,416	1,431
Distance to Closest School (meters)	323.5	319.1	333.0	330.1	331.2	336.6	342.1
Number of Schools in a Radius of 500 meters	2.792	2.788	2.674	2.654	2.677	2.597	2.578
Observations	26,261	27,175	27,039	26,099	25,299	24,856	23,620

Notes: Each year corresponds to the year of application and the school year starting in September of that year. Data on bonus information for alumni family members and RMI are not available before 2012 given that the reform was implemented that year. Quintiles of parental education are defined at the census block level.

Table 3.B.3: School Assignment in the city of Madrid between 2010/2011 and 2016/2017.

Variable	2010	2011	2012	2013	2014	2015	2016
Assigned to first-choice of school	15,640 [0.85]	15,286 [0.84]	15,703 [0.87]	14,845 [0.87]	14,253 [0.87]	1,4049 [0.87]	13,669 [0.87]
Assigned to other school	2,226 [0.12]	2,203 [0.12]	1,961 [0.11]	1,693 [0.1]	1,604 [0.1]	1,539 [0.09]	1,486 [0.09]
Not assigned to any ranked school	525 [0.03]	800 [0.04]	342 [0.02]	432 [0.02]	466 [0.03]	678 [0.04]	623 [0.04]
Total Students	18,391	18,289	18,006	16,970	16,323	16,266	15,696

Notes: Each year corresponds to the year of application and the school year starting in September of that year. Fraction of students are shown in square brackets.

Table 3.B.4: Effect of the reform on Out-of-District Assignment and Distance to Assigned School. Difference-in-Difference Approach.

Dependent Variable	Out-of-District Assignment			Distance to Assigned School		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.143*** (0.025)	0.117*** (0.027)	0.118*** (0.026)	174.249** (80.861)	16.606 (97.967)	0.281 (91.522)
Year 2012	-0.005 (0.023)	-0.008 (0.024)	-0.004 (0.023)	-63.671 (73.993)	-87.925 (82.657)	-42.383 (78.108)
Post 2013	-0.008 (0.015)	-0.010 (0.016)	-0.007 (0.015)	-1.960 (75.096)	-20.473 (80.212)	16.393 (76.834)
Treatment*Year 2012	0.024 (0.024)	0.025 (0.025)	0.021 (0.025)	134.096* (79.564)	149.392* (87.770)	91.449 (82.325)
Treatment*Post 2013	0.036** (0.016)	0.040** (0.017)	0.036** (0.017)	331.053*** (79.802)	351.699*** (85.263)	311.179*** (81.671)
School District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	No	Yes	Yes	No	Yes	Yes
Families Controls	No	No	Yes	No	No	Yes
Observations	29,518	29,518	29,518	29,518	29,518	29,518

Notes: Each year corresponds to the year of application and the school year starting in September of that year. Standard errors are clustered at the census block level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.B.5: The Effect of the School Choice Reform on School Segregation by Parental Education and Immigrant Status.

	Dependent Variable: Segregation by					
	Parental Education			Immigrant Status		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.251*** (0.010)	0.105*** (0.036)	0.065 (0.115)	0.076*** (0.009)	0.042*** (0.010)	0.053 (0.042)
Year 2010	0.025** (0.011)	0.018 (0.011)	0.018 (0.011)	-0.006 (0.004)	-0.004 (0.004)	-0.004 (0.004)
Year 2012	0.031*** (0.009)	0.021*** (0.008)	0.021*** (0.008)	0.012*** (0.004)	0.008** (0.004)	0.008** (0.004)
Year 2013	0.007 (0.009)	-0.013 (0.010)	-0.013 (0.010)	0.008* (0.005)	0.001 (0.005)	0.001 (0.005)
Year 2014	-0.009 (0.009)	-0.031*** (0.010)	-0.022 (0.013)	0.005 (0.004)	0.000 (0.004)	-0.003 (0.006)
Year 2015	-0.007 (0.009)	-0.027*** (0.009)	-0.025** (0.012)	0.001 (0.004)	-0.004 (0.004)	-0.004 (0.004)
Year 2016	0.010 (0.010)	-0.010 (0.011)	-0.008 (0.013)	0.002 (0.004)	-0.004 (0.004)	-0.004 (0.004)
Residential Segregation		0.344*** (0.080)	0.331*** (0.082)		0.644*** (0.118)	0.637*** (0.122)
School district FE	Yes	Yes	Yes	Yes	Yes	Yes
Time-Variant District Characteristics	No	No	Yes	No	No	Yes
Segregation level in 2011	0.7	0.7	0.8	0.06	0.06	0.06
R-squared	0.911	0.927	0.927	0.598	0.673	0.676
Observations	147	147	148	147	147	147

Notes: Each year corresponds to the year of application and the school year starting in September of that year. Robust standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively.

Appendix 3.C Additional Figures and Tables

The online appendix supplements the paper “School Choice Priorities and School Segregation: Evidence from Madrid”. It presents details on priority criteria, an examples of the census block section, years of schooling, population trends in the region of Madrid, housing prices and school average performance, characteristics of assigned school, sample restrictions, theoretical properties of the Boston Mechanism, Difference-in-Difference Approach, alternative measure of segregation by parental education, and school classification.

3.C.1 Description of priority criteria

Table 3.C.1: Tie-break criteria in the region of Madrid.

Ties are broken in favor of the pupils who has higher points on			
	Before 2012/2013	2012/2013	2013/2014
1	Siblings	Siblings	Siblings
2	Proximity	Proximity	Proximity
3	Disability	Disability	Disability
4	Large family	Alumni family member	Alumni student
5	Low annual income per capita	Large family	Large family
6	Random lottery	Low-income	Low-income
7		Random lottery	School discretionary
8			Random lottery

Notes: IPREM is the acronym in Spanish for the Multiple Effects Income Public Index, which was 7,455.14 euros in the period of study. The Minimum Insertion Subsidy (*Renta Mínima de Inserción*) is a special subsidy granted to people with lower income than IPREM. School discretionary is a point that the schools have freedom to assign by “public and objective” criteria.

1. Cannot read, cannot write
2. No studies
3. Incomplete Primary
4. Middle school, Primary or Compulsory Secondary Education
5. Vocational Training (Elemental)
6. Vocational Training (Advanced)
7. High school
8. Other Intermediate Graduates
9. University School Graduates
10. Technical Engineer
11. College Graduated
12. No-university Graduated
13. PhD and other Post-graduates

We group all these 13 categories into 6 new categories. Categories 1 and 2 are grouped as No studies. Category 3 remains as Incomplete Primary Education. Categories 4 and 5 form the Lower Secondary Education. Categories 6, 7 and 8 are gathered together in Upper Secondary Education. Categories 9 and 10 form the Lower Tertiary Education. Groups 11, 12 and 13 are joined in category Post-Graduate education. In order to assign an equivalent amount of years of schooling, we make the assumption that parents were educated under the legal framework of the Education Act LGE (*Ley General de Educación*), which was in place for pupils born before 1985. We do this as the average maternity age for the first offspring was 29.5 years old in 2007 (therefore those students aged 3 in 2010) and 30.5 for 2013 (therefore those students aged 3 in 2016). This means that with a large probability, mothers of students born in 2007 and 2013 were at school as part of the LGE framework.

The LGE framework consisted of 8 years of basic primary schooling, with 5 years of primary school and 3 years of lower secondary school. We assume category *No Studies* as only 3 years of primary and *Incomplete primary* as just 5 years of schooling. After that Basic schooling, there was vocational training with degrees of 2 to 4 years. We assume Categories 4,5 and 6 to be in between Basic schooling (8 years and some vocational training), averaging 9 years of schooling. Categories 7 and 8 pertain to High School and Other intermediate graduates, which corresponded to 12 years of schooling. Finally, university graduates and technical engineers are assumed to do 18 years of schooling, and post-graduate studies are given an average of 3 additional years of education. The equivalence of years of schooling for a specific census block is given by the following formula, where each percentage of census population is multiplied by the equivalent years of schooling.

$$YS_s = Non-studies_s * 3 + Primary_s * 5 + LowSecondary_s * 9 + UpperSecondary_s * 12 + LowTertiary_s * 17 + College_s * 20 \quad (3.5)$$

3.C.4 Population trends in the Region of Madrid

Table 3.C.3: Total population with 3-year old children in the Region of Madrid by Immigrant status.

	2010	2011	2012	2013	2014	2015	2016
A. Region of Madrid							
Spanish	62,293	63,636	65,905	62,266	60,796	60,296	59,042
Non-Spanish	8,800	9,762	10,771	11,759	10,992	9,394	91,64
Total	71,093	73,398	76,676	74,025	71,788	69,690	68,206
B. City of Madrid							
Spanish	27,202	27,919	29,126	27,045	26,053	25,666	NA
Non-Spanish	3,601	3,881	4,118	4,921	4,495	4,000	NA
Total	30,803	31,8	33,244	31,966	30,548	29,666	NA

Notes: This Table is based on own computations from INE (*Padron*).
NA indicate that data are not available for this particular year.

3.C.5 Housing Prices and School Average Performance.

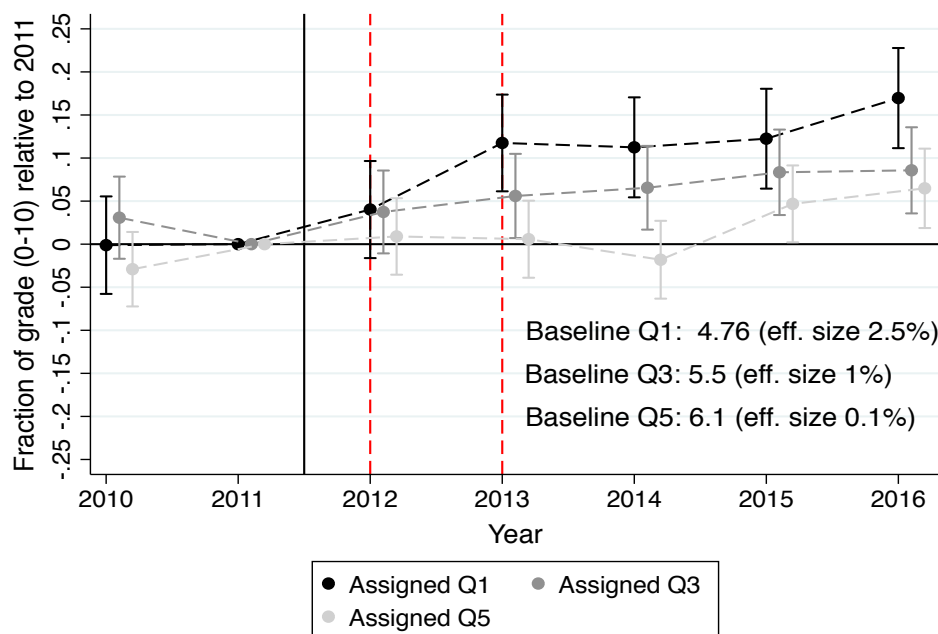
Table 3.C.4: Housing Prices and School Average Performance.

Dependent variable: Log Housing Prices	(1)	(2)	(3)
Constant	7.887*** (0.019)	8.035*** (0.033)	8.037*** (0.034)
Avg. school performance of district	0.502*** (0.033)	0.502*** (0.027)	0.450*** (0.059)
Year 2010		0.063 (0.045)	0.064 (0.046)
Year 2012		-0.109** (0.046)	-0.110** (0.048)
Year 2013		-0.214*** (0.051)	-0.217*** (0.053)
Year 2014		-0.247*** (0.059)	-0.248*** (0.063)
Year 2015		-0.279*** (0.054)	-0.284*** (0.057)
Year 2016		-0.250*** (0.057)	-0.256*** (0.058)
Year 2010*Avg. school performance of district			-0.018 (0.079)
Year 2012*Avg. school performance of district			0.014 (0.082)
Year 2013*Avg. school performance of district			0.066 (0.092)
Year 2014*Avg. school performance of district			0.040 (0.110)
Year 2015*Avg. school performance of district			0.116 (0.097)
Year 2016*Avg. school performance of district			0.145 (0.101)
Observations	147	147	147
R-squared	0.588	0.725	0.733

Notes: Log Housing prices are housing prices at the district level by year t . Avg. school performance district refers to the average peer performance in the *CDI* standardized test by school district previous to the 2012's reform. Years refer to year fixed effects. Robust standard errors are displayed in parenthesis.

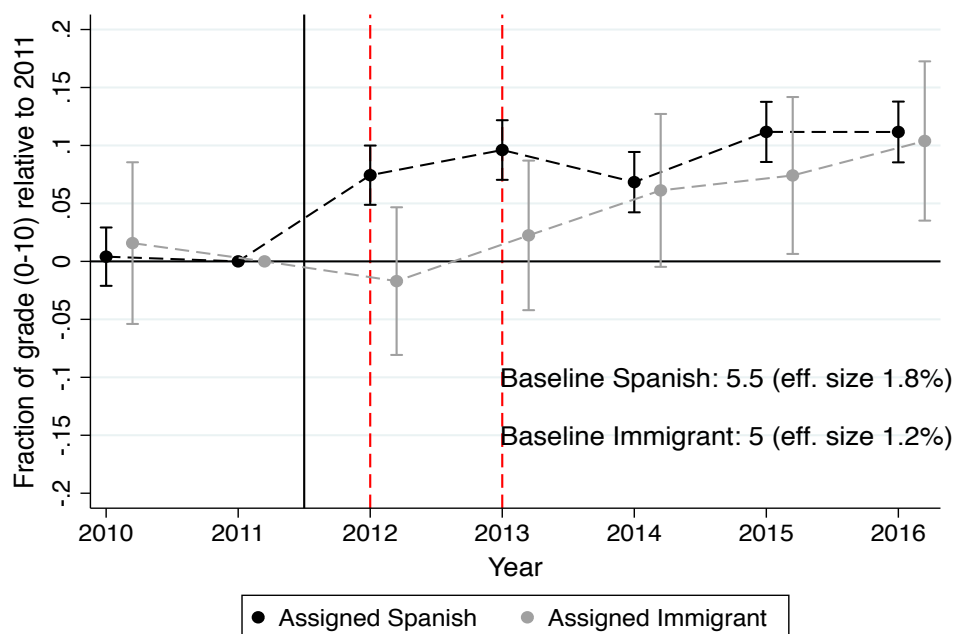
3.C.6 To which schools are pupils assigned?

Figure 3.C.2: Average Performance of Assigned School (0-10) by Parental Education.



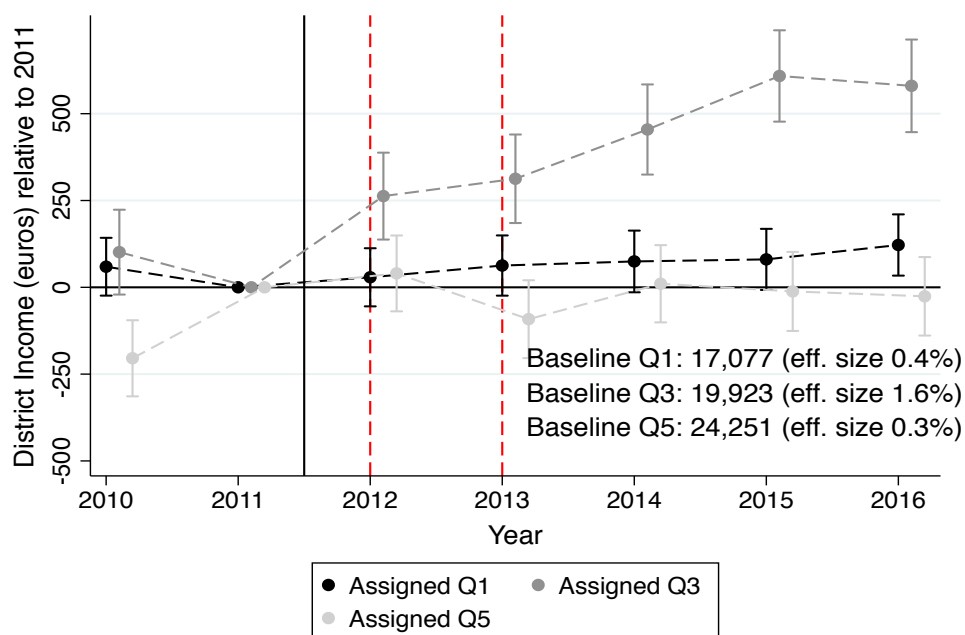
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.C.3: Average Performance of Assigned School (0-10) by Immigrant Status.



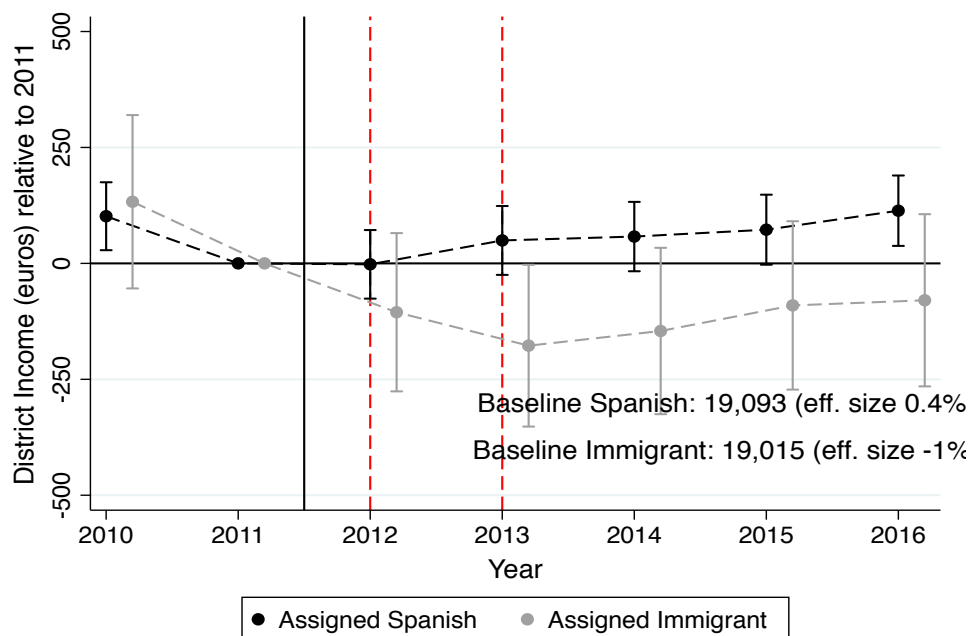
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.C.4: District Income (euros) of Assigned School by Parental Education.



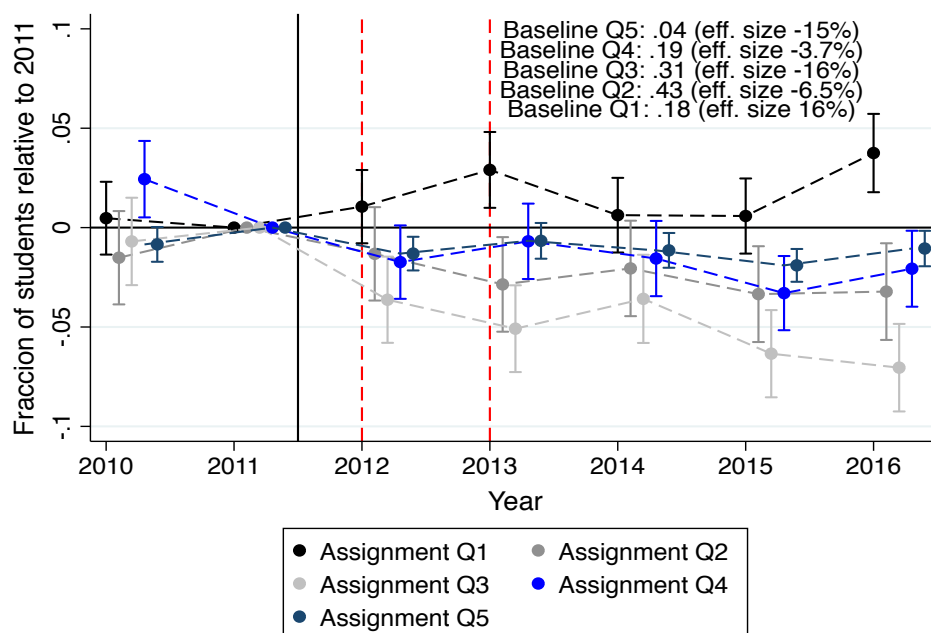
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.C.5: District Income (euros) of Assigned School by Immigrant Status.



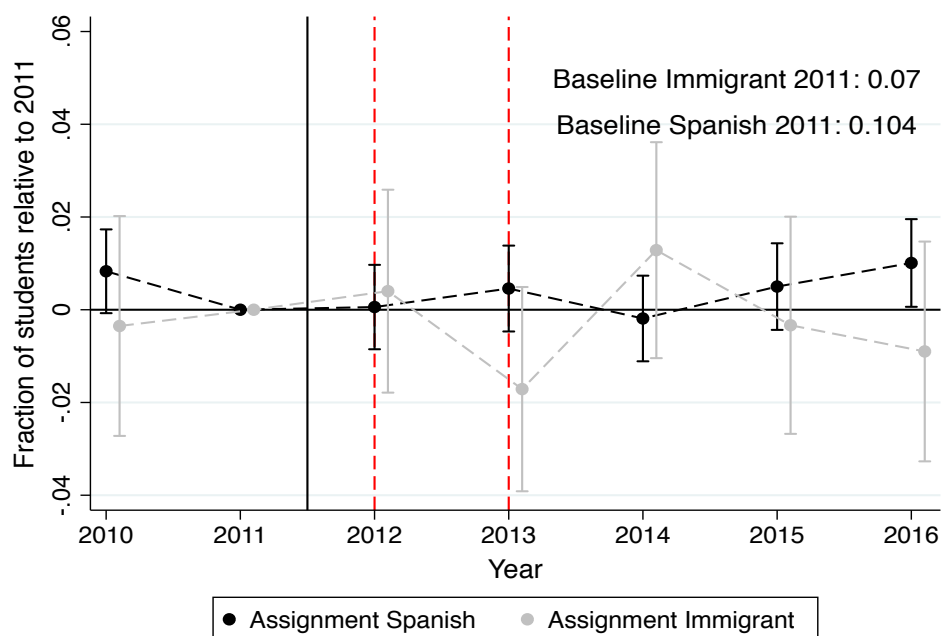
Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.C.6: Predetermined School Segregation of District of Assigned School by Parental Education.



Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

Figure 3.C.7: Predetermined School of District of Assigned School by Immigrant Status.



Note: Black and grey dots represent point estimates. Horizontal lines depict confidence intervals at the 95% level of confidence. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013.

3.C.7 Sample Restrictions

Despite the fact that the administrative dataset of applications is unique and exhaustive, adding other sources of information for the analysis implies some observations which are missing. For example, our analysis focuses on all students who apply for a school where information on the test scores (our measure of school quality) is available. We are able to identify the school test score measure for about 90% to 95% of the observations. Given the importance of identifying geographical mobility patterns, we discard students whose address information is missing or not valid. School application forms are manually submitted to the school or the central administration which then introduce the information into the digital centralized system (Calsamiglia and Güell, 2019), which may result in invalid information. We are able to identify the address of about 99% percent of the final population of interest. The analysis is based on a population of about 26,000 to 29,000 observations, depending on the school year.

Table 3.C.5: Sample Restrictions: School Applicants in the city of Madrid over 2010-2016.

Number of Applicants	2010	2011	2012	2013	2014	2015
Total unique applications	37,146	39,986	37,300	36,273	35,727	34,124
Students with no siblings	32,009	30,594	30,930	30,022	29,341	27,762
School test scores	29,977	28,884	29,488	28,714	27,810	26,487
Valid Address	29,666	28,720	29,338	28,577	27,807	26,491
Total Final Sample	29,736	28,700	29,205	28,478	27,579	26,271

Notes: Each year corresponds to the year of application and the school year starting in September of that year. This Tables does include all applications irrespective of the pre-school or school year in weich families applied for.

3.C.8 Theoretical Properties of the Boston Mechanism

The theoretical properties are extracted from Chen and Sönmez (2006). The Boston mechanism (BM) which works as follows: For each school, a priority ordering of students is determined based on state and local laws/policies. Each student

submits a preference ranking of the schools. The final phase is student assignment based on submitted preferences and priorities. The outcome is obtained in several rounds. In Round 1, only the first-choices of the students are considered. For each school, consider the students who have listed it as their first-choice and assign seats of the school to these students one at a time following their priority order until either there are no seats left or there is no student left who has listed it as her first-choice. In general, at Round k consider the remaining students. In Round k only the k th choices of students are considered. For each school with available seats, consider the students who have listed it as their k th choice and assign the remaining seats to these students one at a time following their priority order until either there are no seats left or there is no student left who has listed it as her k th choice.

The major difficulty with the Boston student assignment mechanism is that it is not strategy-proof; that is, students may benefit from misrepresenting their preferences. Abdulkadiroglu, Atila and Sönmez (2003) point out that the Boston mechanism gives students and their parents a strong incentive to misrepresent preferences by improving the ranking of schools which they have a high priority.

3.C.9 Difference-in-Difference Approach

Table 3.C.6: Treatment-Control balance in fixed households characteristics in pre-reform years.

	Control Group (1)	Treatment Group (2)	Difference T – C (3)	$2 > p$ -value (4)
Female	0.485	0.500	0.015	0.644
Immigrant	0.135	0.110	-0.025	0.133
Immigrant Mother	0.281	0.266	-0.015	0.623
Immigrant Father	0.372	0.324	-0.048	0.168
Lowest Quintile Education Block	0.103	0.121	0.018	0.521
Highest Quintile Education Block	0.110	0.242	0.131	0.053
Euclidean distance to closest school	217.804	250.967	33.163	0.061
Number of schools in a radius of 500 meters	3.583	3.393	-0.190	0.396
N	285	8,528		
Test of joint significance	F -stat: 1.206 (p -value: > 0.292)			

Notes: Standard errors are clustered at the census block level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.C.7: Sensitivity Analysis: Effect of the reform on Out-of-District Assignment and Distance to Assigned School by treatment group definition.

Dependent Variable	Out-of-District Assignment					Distance to Assigned School				
	200m (1)	250m (2)	300m (3)	350m (baseline) (4)	400m (5)	200m (6)	250m (7)	300m (baseline) (8)	350m (9)	400m (10)
Treatment	0.154*** (0.029)	0.136*** (0.027)	0.118*** (0.026)	0.107*** (0.025)	0.097*** (0.024)	24.998 (95.820)	14.136 (94.826)	0.281 (91.522)	16.270 (89.348)	13.445 (86.619)
Year 2012	-0.006 (0.024)	-0.004 (0.023)	-0.004 (0.023)	-0.002 (0.023)	-0.002 (0.023)	-41.802 (78.933)	-41.503 (78.949)	-42.383 (78.108)	-41.190 (77.265)	-42.193 (76.335)
Post 2013	-0.008 (0.016)	-0.007 (0.016)	-0.007 (0.015)	-0.006 (0.015)	-0.006 (0.015)	16.821 (77.083)	15.491 (77.708)	16.393 (76.834)	17.555 (76.362)	17.749 (76.091)
Treatment*Year 2012	0.016 (0.026)	0.018 (0.025)	0.021 (0.025)	0.018 (0.024)	0.018 (0.024)	68.346 (86.749)	88.614 (84.593)	91.449 (82.325)	69.240 (80.789)	78.577 (79.430)
Treatment*Post 2013	0.036** (0.018)	0.034** (0.017)	0.036** (0.017)	0.036** (0.016)	0.036** (0.016)	356.800*** (85.259)	323.511*** (83.623)	311.179*** (81.671)	284.180*** (80.393)	279.650*** (79.373)
School District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Families Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,995	23,566	29,518	35,676	41,537	17,995	23,566	29,518	35,676	41,537

Notes: Each year corresponds to the year of application and the school year starting in September of that year. Standard errors are clustered at the census block level and displayed in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.C.10 Alternative measure of segregation by parental education

We use a two-stage method to compute the change in within-school heterogeneity regarding student's parental education. Hence, what we are computing is an inverse measure of social segregation of schools. In the first step, we regress student's parental education on school dummies and neighborhood fixed effects of the school in a joint regression for 2011, 2012 and 2013. This can be seen as:

$$EDU_{ijt} = \alpha + \phi_j + \delta_n + \epsilon_{ijt} \quad (3.6)$$

where EDU_{ijt} are the number of parental years of education of student i applying for a school j in year t , ϕ_j are school fixed effects, and δ_n are neighborhood fixed effects. We take the absolute value of the student residual ($|\hat{\epsilon}_{ij}|$), and for a better interpretation, we use the logarithm of the residual in the second step. This provides a measure of the student parental education variation that cannot be explained by the school that the student is applying for and the neighborhood heterogeneity in parental education where the student resides.

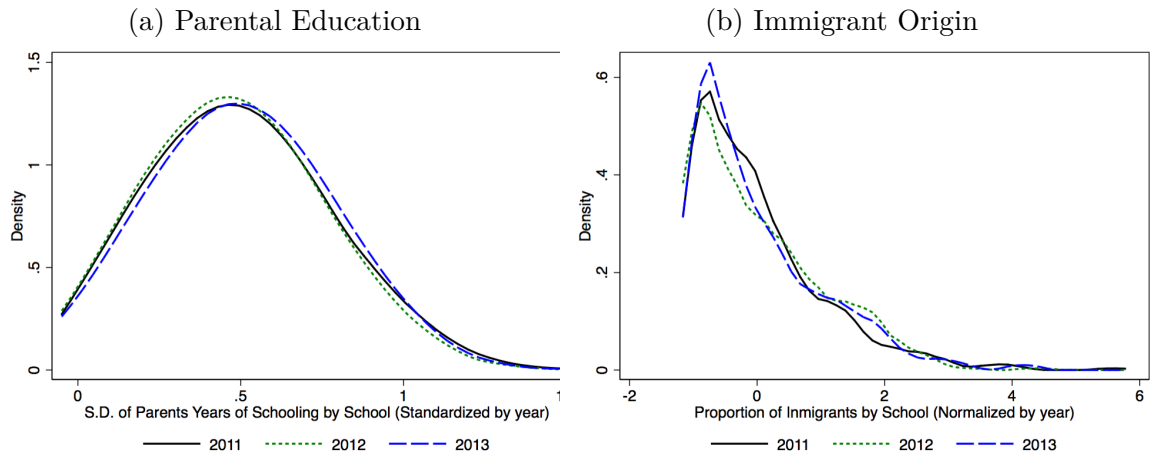
In the second step, we regress this value on a year dummy to measure the average effect of the reform on school heterogeneity. In other words, to establish whether the unexplained variance of student parental education has changed due to the reform. Hence, the estimation can be described as follows:

$$\log|\hat{\epsilon}_{ijt}| = \alpha + \theta Year2012_{ij} + \kappa Year2013_{ij} + \varepsilon_{ij} \quad (3.7)$$

where the coefficients θ and κ are the average change in the student social diversity or heterogeneity in 2012 and 2013. A positive value of the estimates would be interpreted as an increase in social heterogeneity (decrease in segregation), and a negative one as decrease in the level of social heterogeneity (raise in social segregation).

The residual $|\hat{\epsilon}_{ij}|$ of student i for school j reflects the level of student heterogeneity that may not be explained by the school fixed effect and the residential dynamics of the Madrid households between 2011 and 2013. The second estimation step regresses this measure of student heterogeneity on the post-reform year dummies net of other contextual factors such as the school and the neighborhood.

Figure 3.C.8: Densities of within-school socioeconomic variation and immigrant share, by year.



We first present descriptive evidence of the dynamics in student composition at schools. Figure 3.C.8 displays the density of within-school variation of student background for 2011, 2012 and 2013. The solid line represents the density for 2011, while the short-dashed and long-dashed lines refer to 2012 and 2013 respectively. The left-hand side graph shows the density of within-school standard deviation of parents' years of education. The curve shifts mildly to the left in 2012 (with respect to 2011), whereas the 2013 density curve shifts to the right in the low and middle part of the distribution indicating a mild increase in within-school student heterogeneity in terms of parental education. The right-hand side graph (Figure 3.C.8a) shows the density of within-school standard deviation of parents' years of education. The curve shifts mildly to the left in 2012 (with respect to 2011), whereas the 2013 density curve shifts to the right in the low and middle part of the distribution indicating a mild increase in within-school student heterogeneity in terms of parental education. The right-hand side graph (Figure 3.C.8a) plots the distribution of the proportion of

immigrants within school normalized by the total share of immigrants in the city.⁵³ This Figure shows a polarization of the distribution in 2012 and 2013: while the share of schools with very low proportion of immigrants slightly increases, the proportion of schools with a median proportion of immigrants decreases in 2012 and 2013. At the same time, there was an increase in the share of schools with higher proportion of immigrants (usually more than 30% of immigrants). Whereas the evidence in terms of social segregation is mild, the second graph describes a process of immigrant segregation across schools at the time of the reform.

Table 3.C.8 presents the results for the estimations of the model of Equation 3.7 model. We estimate changes in the within-school heterogeneity of parental education of students. It compares the marginal effect of 2012 and 2013 relative to 2011, controlling for school fixed effects and neighborhood dynamics of residential segregation. We conduct three different specifications: (i) we compute the residuals in a joint regression for the three years, controlling for neighborhood heterogeneity of parental education; (ii) we calculate residuals in separate year regressions, controlling for neighborhood heterogeneity of parental education; (iii) we measure residuals in a joint regression for all three years, controlling for district heterogeneity of parental education. The Table shows that the changes in within-school heterogeneity of parental education did not change in 2012, but increased significantly in 2013 with respect to 2011. The magnitude of the effect is between 4% and 5%. In other words, it seems that the reform increased diversity within school by 4 to 5%. Nevertheless, the measurement error of the parental education variable may be correlated with the direction of the reform, inducing a positive (negative) bias in our heterogeneity (segregation) estimates that we discuss in 3.6.

⁵³We standardize the standard deviation of parental education by dividing by the total yearly standard deviation of education. Conversely, we normalize the proportion of immigrants by subtracting the yearly mean and dividing by the average mean of immigrants in the system. This is because the socioeconomic characteristics of parents may not be constant in the 2011-2013 period.

Table 3.C.8: Change in within-school student heterogeneity by parental education.

Dependent variable: $\log \hat{\epsilon}_{i,j} $	(2012)	(2013)
Specification 1	0.0059 (0.0127)	0.0517*** (0.0129)
Specification 2	-0.134 (0.0128)	0.043*** (0.013)
Specification 3	0.0059 (0.0127)	0.0517*** (0.0129)

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively.

3.C.11 School Classification

Table 3.C.9: Districts ranked by income.

Low Income District	Middle-Income District	High-Income District
Puente de Vallecas	San Blas - Canillejas	Centro
Villaverde	Moratalaz	Barajas
Usera	Ciudad Lineal	Moncloa - Aravaca
Carabanchel	Tetúan	Retiro
Latina	Fuencarral - El Pardo	Chamberí
Vicálvaro	Hortaleza	Chamartín
Villa de Vallecas	Arganzuela	Salamanca

Notes: Data from district average income is derived from Municipal census data in the city of Madrid.

Bibliography

Abdulkadiroglu, Atila, and Tayfun Sönmez. 2003. "School choice: A mechanism design approach." *The American Economic Review*, 93(3): 729–747.

Abdulkadiroglu, Atila, Parag Pathak, Alvin E Roth, and Tayfun Sonmez. 2006. "Changing the Boston school choice mechanism." National Bureau of Economic Research.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda. 2011. "Resolving conflicting preferences in school choice: The" boston mechanism" reconsidered." *American Economic Review*, 101(1): 399–410.

Amrein, Audrey L, and David C Berliner. 2002. "High-stakes testing & student learning." *Education policy analysis archives*, 10: 18.

Anderson, S, A. Palma, and J Thisse. 1992. *Discrete Choice Theory of Product Differentiation*. The MIT Press.

Anghel, Brindusa, Antonio Cabrales, and Jesus M. Carro. 2016. "Evaluating a bilingual education program in spain: The impact beyond foreign language learning." *Economic Inquiry*, 54(2): 1202–1223.

Angrist, Joshua D. 1993. "The effect of veterans benefits on education and earnings." *ILR Review*, 46(4): 637–652.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and services for college achievement: Evidence from a randomized trial." *American Economic Journal: Applied Economics*, 1(1): 136–163.

- Angrist, Joshua, Philip Oreopoulos, and Tyler Williams.** 2014. “When opportunity knocks, who answers? New evidence on college achievement awards.” *Journal of Human Resources*, 49(3): 572–610.
- Arellano, Manuel, and Gemma Zamarro.** 2007. “The choice between public and private schools with or without subsidies in Spain.” *Unpublished*.
- Ariely, Dan, Uri Gneezy, George Loewenstein, and Nina Mazar.** 2009. “Large stakes and big mistakes.” *The Review of Economic Studies*, 76(2): 451–469.
- Ariño, Antonio.** 2011. *¿ Universidad sin clases? Condiciones de vida de los estudiantes universitarios en España (Eurostudent IV)*. Ministerio de Educación.
- Azmat, Ghazala, Caterina Calsamiglia, and Nagore Iriberry.** 2016. “Gender differences in response to big stakes.” *Journal of the European Economic Association*, 14(6): 1372–1400.
- Bailey, Martha J, and Susan M Dynarski.** 2011. “Gains and gaps: Changing inequality in US college entry and completion.” National Bureau of Economic Research.
- Baron-Cohen, Simon.** 2004. *The essential difference*. Penguin UK.
- Barrera-Orsorio, Felipe, and Deon Filmer.** 2016. “Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches.” *Journal of Human Resources*, 51(2): 461–499.
- Barrera-Orsorio, Felipe, Andreas De Barros, and Deon Filmer.** 2018. “Long-term impacts of alternative approaches to increase schooling: evidence from a scholarship program in Cambodia.”
- Barrow, Lisa, Lashawn Richburg-Hayes, Cecilia Elena Rouse, and Thomas Brock.** 2014. “Paying for performance: The education impacts of a community college scholarship program for low-income adults.” *Journal of Labor Economics*, 32(3): 563–599.

- Baumeister, Roy F.** 1984. "Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance." *Journal of personality and social psychology*, 46(3): 610.
- Baum, Sandy, and Judith Scott-Clayton.** 2013. "Redesigning the Pell Grant program for the twenty-first century." *The Hamilton Project, Discussion Paper*, 4.
- Baum, Sandy, Jennifer Ma, and Kathleen Payea.** 2013. "Education Pays, 2010: The Benefits of Higher Education for Individuals and Society. Trends in Higher Education Series." *College Board Advocacy & Policy Center*.
- Becker, Gary S.** 1962. "Investment in human capital: A theoretical analysis." *Journal of political economy*, 70(5, Part 2): 9–49.
- Beilock, Sian.** 2010. *Choke: What the secrets of the brain reveal about getting it right when you have to*. Simon and Schuster.
- Beilock, Sian L.** 2008. "Math performance in stressful situations." *Current Directions in Psychological Science*, 17(5): 339–343.
- Bénabou, Roland, and Jean Tirole.** 2000. "Self-confidence and social interactions." National bureau of economic research.
- Bénabou, Roland, and Jean Tirole.** 2002. "Self-confidence and personal motivation." *The Quarterly Journal of Economics*, 117(3): 871–915.
- Beneito, P., J. E. Bosca, and J. Ferri.** 2018. "Tuition Fees and Student Effort at University." *Economics of Education Review*, 64: 114–128.
- Berg, Gary A.** 2016. *Low-income students and the perpetuation of inequality: Higher education in America*. Routledge.
- Bettinger, Eric.** 2015. "Need-based aid and college persistence: The effects of the Ohio College Opportunity Grant." *Educational Evaluation and Policy Analysis*, 37(1_suppl): 102S–119S.

- Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu.** 2012. “The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment.” *The Quarterly Journal of Economics*, 127(3): 1205–1242.
- Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes.** 2017. “Life-cycle earnings, education premiums, and internal rates of return.” *Journal of Labor Economics*, 35(4): 993–1030.
- Billings, Stephen B., David J. Deming, and Jonah Rockoff.** 2014. “School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg.” *Quarterly Journal of Economics*, 129(1): 435–476.
- Black, Dan A, Amelia M Haviland, Seth G Sanders, and Lowell J Taylor.** 2008. “Gender wage disparities among the highly educated.” *Journal of Human Resources*, 43(3): 630–659.
- Blau, Francine D, and Lawrence M Kahn.** 2017. “The gender wage gap: Extent, trends, and explanations.” *Journal of Economic Literature*, 55(3): 789–865.
- Board, College.** 2017. “Trends in student aid 2017.” College Board Publications.
- Böhlmark, A., and M. Lindahl.** 2007. “The impact of school choice on pupil achievement, segregation and costs: Swedish evidence.” *IZA Discussion Paper Series*, , (2786): 65.
- Böhlmark, Anders, Helena Holmlund, and Mikael Lindahl.** 2016. “Parental choice, neighbourhood segregation or cream skinning? An analysis of school segregation after a generalized choice reform.” *Journal of Population Economics*, 29(4): 1155–1190.
- Breda, T., Ly S. T.** 2015. “Professors in core science fields are not always biased against women: Evidence from France.” *American Economic Journal: Applied Economics*, 7(4): 53–75.

- Brown, Charles, and Mary Corcoran.** 1997. “Sex-based differences in school content and the male-female wage gap.” *Journal of Labor Economics*, 15(3): 431–465.
- Burgess, Simon, and Adam Briggs.** 2010. “School Assignment, School Choice and Social Mobility.” *Economics of Education Review*, 29: 639–649.
- Burgess, Simon, Carol Propper, and Deborah Wilson.** 2007. “the Impact of School Choice in England.” *Policy Studies*, 28(2): 129–143.
- Burgess, Simon, Ellen Greaves, Anna Vignoles, and Deborah Wilson.** 2015. “What Parents Want: School Preferences and School Choice.” *The Economic Journal*, 125(587): 1262–1289.
- Bursztyn, Leonardo, and Robert Jensen.** 2015. “How does peer pressure affect educational investments?” *The Quarterly Journal of Economics*, 130(3): 1329–1367.
- Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik.** 2014. “Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs.” *Econometrica*, 82(6): 2295–2326.
- Calsamiglia, Caterina.** 2014. “Matching Practices for Elementary and Secondary Schools – Spain.” MiP Country Profile 17.
- Calsamiglia, Caterina, and Maia Güell.** 2019. “Priorities in School Choice: The case of the Boston mechanism in Barcelona.” *Journal of Public Economics*, *Forthcoming*.
- Calsamiglia, Caterina, Francisco Martínez-Mora, Antonio Miralles, et al.** 2017. “Sorting in public school districts under the Boston Mechanism.” *Unpublished working paper*.
- Camerer, Colin, George Loewenstein, and Drazen Prelec.** 2005. “Neuroeconomics: How neuroscience can inform economics.” *Journal of economic Literature*, 43(1): 9–64.

- Card, David, Alexandre Mas, and Jesse Rothstein.** 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics*, 123(1): 177–218.
- Card, David, and A Abigail Payne.** 2017. "High school choices and the gender gap in STEM." National Bureau of Economic Research.
- Card, David, and Jesse Rothstein.** 2007. "Racial segregation and the black–white test score gap." *Journal of Public Economics*, 91(11–12): 2158–2184.
- Carr, Priyanka B, and Claude M Steele.** 2009. "Stereotype threat and inflexible perseverance in problem solving." *Journal of Experimental Social Psychology*, 45(4): 853–859.
- Castleman, Benjamin L, and Bridget Terry Long.** 2016. "Looking beyond enrollment: The causal effect of need-based grants on college access, persistence, and graduation." *Journal of Labor Economics*, 34(4): 1023–1073.
- Castleman, Benjamin L, Terry Long, Avery Bridget, Murnane Chris, Goodman Richard, and L Benjamin.** 2014. "The impact of partial and full merit scholarships on college entry and success: Evidence from the Florida Bright Futures Scholarship program." *EdPolicyWorks Working Paper Series No. 17*.
- Cha, Paulette, and Reshma Patel.** 2010. "Rewarding progress, reducing debt: Early results from the performance-based scholarship demonstration in Ohio."
- Chetty, Raj, and Nathaniel Hendren.** 2018*a*. "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects." *The Quarterly Journal of Economics*, 133(3): 1107–1162.
- Chetty, Raj, and Nathaniel Hendren.** 2018*b*. "The impacts of neighborhoods on intergenerational mobility II: County-level estimates." *The Quarterly Journal of Economics*, 133(3): 1163–1228.
- Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz.** 2016. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." *American Economic Review*, 106(4): 855–902.

- CJE.** 2015. “Observatorio de la Emancipación del primer semestre de 2015 (n.º 10).” *Madrid. www. cje. org.*
- Coleman, James S.** 1968. “Equality of educational opportunity.” *Integrated Education*, 6(5): 19–28.
- Cornwell, Christopher, David B Mustard, and Deepa J Sridhar.** 2006. “The enrollment effects of merit-based financial aid: Evidence from Georgia’s HOPE program.” *Journal of Labor Economics*, 24(4): 761–786.
- Cornwell, Christopher, David B Mustard, and Jessica Van Parys.** 2013. “Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school.” *Journal of Human Resources*, 48(1): 236–264.
- Cunha, Flavio, and James Heckman.** 2007. “The technology of skill formation.” *American Economic Review*, 97(2): 31–47.
- Daly, Herman, and Victor Lavy.** 2009. “The effects of high stakes high school achievement awards: Evidence from a randomized trial.” *The American Economic Review*, 99(4): 1384–1414.
- Deci, Edward L.** 1972. “The effects of contingent and noncontingent rewards and controls on intrinsic motivation.” *Organizational behavior and human performance*, 8(2): 217–229.
- de España, Banco, et al.** 2017. “Encuesta Financiera de las Familias (EFF) 2014: métodos, resultados y cambios desde 2011.” *Boletín Económico*, , (MAR).
- De Haan, Monique, Pieter A Gautier, Hessel Oosterbeek, and Bas Van der Klaauw.** 2015. “The performance of school assignment mechanisms in practice.”
- Deming, David, and Susan Dynarski.** 2009. “Into college, out of poverty? Policies to increase the postsecondary attainment of the poor.” National Bureau of Economic Research WP. 15387.

- Deming, David J.** 2017. “Increasing college completion with a federal higher education matching grant.” *The Hamilton Project, Policy Proposal*, 3.
- Deming, David J, and Christopher R Walters.** 2017. “The impact of price caps and spending cuts on US postsecondary attainment.” National Bureau of Economic Research.
- Denning, Jeffrey T.** 2018. “Born under a lucky star: Financial aid, college completion, labor supply, and credit constraints.” *Journal of Human Resources*.
- Denning, Jeffrey T, Benjamin M Marx, and Lesley J Turner.** 2017. “ProPelled: The Effects of Grants on Graduation, Earnings, and Welfare.” National Bureau of Economic Research.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer.** 2011. “Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya.” *American Economic Review*, 101(5): 1739–1774.
- Duncan, Otis Dudley, and Beverly Duncan.** 1955. “A Methodological Analysis of Segregation Indexes.” *American Sociological Review*, 20(2): 210.
- Dynarski, Susan.** 2008. “Building the stock of college-educated labor.” *Journal of Human Resources*, 43(3): 576–610.
- Dynarski, Susan, and Judith Scott-Clayton.** 2013. “Financial aid policy: Lessons from research.” *Future of Children*, 23(1): 67–91.
- Dynarski, Susan M.** 2003. “Does aid matter? Measuring the effect of student aid on college attendance and completion.” *American Economic Review*, 93: 279–288.
- Ehrlinger, Joyce, and David Dunning.** 2003. “How chronic self-views influence (and potentially mislead) estimates of performance.” *Journal of personality and social psychology*, 84(1): 5.
- Ellison, Glenn, and Ashley Swanson.** 2018. “Dynamics of the Gender Gap in High Math Achievement.” National Bureau of Economic Research.

- Ellwood, David, Thomas J Kane, et al.** 2000. "Who is getting a college education? Family background and the growing gaps in enrollment." *Securing the future: Investing in children from birth to college*, 283–324.
- Epple, Dennis, Richard E Romano, and Miguel Urquiola.** 2017. "School vouchers: A survey of the economics literature." *Journal of Economic Literature*, 55(2): 441–92.
- Fack, Gabrielle, and Julien Grenet.** 2010. "When do better schools raise housing prices? Evidence from Paris public and private schools." *Journal of Public Economics*, 94(1-2): 59–77.
- Fack, Gabrielle, and Julien Grenet.** 2015. "Improving college access and success for low-income students: Evidence from a large need-based grant program." *American Economic Journal: Applied Economics*, 7(2): 1–34.
- Figlio, David N, and Joe A Stone.** 2001. "Can public policy affect private school cream skimming?" *Journal of Urban Economics*, 49(2): 240–266.
- Frankel, David M, and Oscar Volij.** 2011. "Measuring school segregation." *Journal of Economic Theory*, 146(1): 1–38.
- Fryer Jr, Roland G, and Steven D Levitt.** 2010. "An empirical analysis of the gender gap in mathematics." *American Economic Journal: Applied Economics*, 2(2): 210–40.
- Fryer, Roland G.** 2011. "Financial incentives and student achievement: Evidence from randomized trials." *The Quarterly Journal of Economics*, 126(4): 1755–1798.
- García-Brazales, Javier.** 2019. "Internal versus External Evaluation of Standardized Exams: Evidence from a Randomized Control Trial."
- Gibbons, Stephen, Olmo Silva, and Felix Weinhardt.** 2017. "Neighbourhood turnover and teenage attainment." *Journal of the European Economic Association*, 15(4): 746–783.

- Gneezy, Uri, and Aldo Rustichini.** 2000. "Pay enough or don't pay at all." *The Quarterly Journal of Economics*, 115(3): 791–810.
- Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel.** 2011. "When and why incentives (don't) work to modify behavior." *The Journal of Economic Perspectives*, 25(4): 191–209.
- Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko.** 2006. "The homecoming of American college women: The reversal of the college gender gap." *The Journal of Economic Perspectives*, 20(4): 133–133.
- Goldrick-Rab, Sara, Robert Kelchen, Douglas N Harris, and James Benson.** 2016. "Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion." *American Journal of Sociology*, 121(6): 1762–1817.
- Goodman, Joshua, Michael Hurwitz, Jonathan Smith, and Julia Fox.** 2015. "The relationship between siblings' college choices: Evidence from one million SAT-taking families." *Economics of Education Review*, 48: 75–85.
- Goux, Dominique, and Eric Maurin.** 2007. "Close neighbours matter: Neighbourhood effects on early performance at school." *The Economic Journal*, 117(523): 1193–1215.
- Hanushek, Eric A, John F Kain, and Steven G Rivkin.** 2009. "New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement." *Journal of Labor Economics*, 27(3): 349–383.
- Hastings, Justine, and Jeffrey Weinstein.** 2008. "Formation, School Choice, and Academic Achievement: Evidence from Two Experiments." *The Quarterly Journal of Economics*, 123: 1373–1414.
- Hastings, Justine S, Thomas J Kane, and Douglas O Staiger.** 2005. "Parental preferences and school competition: Evidence from a public school choice program." National Bureau of Economic Research.

- Hastings, Justine, Thomas Kane, and Douglas Staiger.** 2009. "Heterogeneous preferences and the efficacy of school choice." *NBER Working Papers No. 2145*.
- Hoxby, Caroline.** 2000. "Peer effects in the classroom: Learning from gender and race variation." National Bureau of Economic Research.
- Hoxby, Caroline, Sarah Turner, et al.** 2013. "Expanding college opportunities for high-achieving, low income students." *Stanford Institute for Economic Policy Research Discussion Paper*, , (12-014).
- Hsieh, C, and M Urquiola.** 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program." *Journal of Public Economics*, , (90): 1477–1503.
- Huber, Stephan, and Christoph Rust.** 2016. "Calculate Travel Time and Distance with OpenStreetMap Data Using The Open Source Routing Machine (OSRM)." *The Stata Journal*, 16(2): 1–8.
- Imbens, Guido, and Karthik Kalyanaraman.** 2012. "Optimal bandwidth choice for the regression discontinuity estimator." *The Review of economic studies*, 79(3), pp. 933–959.
- Imbens, Guido W, and Thomas Lemieux.** 2008. "Regression discontinuity designs: A guide to practice." *Journal of econometrics*, 142(2): 615–635.
- Iriberri, Nagore, and Pedro Rey-Biel.** 2011. "The role of role uncertainty in modified dictator games." *Experimental Economics*, 14(2): 160–180.
- Iriberri, Nagore, and Pedro Rey-Biel.** 2017. "Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision." *Journal of Economic Behavior & Organization*, 135: 99–111.
- Jencks, Christopher, and Susan E Mayer.** 1990. "The social consequences of growing up in a poor neighborhood." *Inner-city poverty in the United States*, 111: 186.

- Jenkins, Stephen P, John Micklewright, and Sylke V Schnepf.** 2008. "Social segregation in secondary schools: how does England compare with other countries?" *Oxford Review of Education*, 34(1): 21–37.
- Joensen, Juanna Schrøter, and Helena Skyt Nielsen.** 2018. "Spillovers in education choice." *Journal of Public Economics*, 157: 158–183.
- Kahn, Shulamit, and Donna Ginther.** 2017. "Women and STEM." National Bureau of Economic Research.
- Kane, Thomas J.** 1995. "Rising public college tuition and college entry: How well do public subsidies promote access to college?" National Bureau of Economic Research.
- Kane, Thomas J, and Cecilia Elena Rouse.** 1995. "Labor-market returns to two-and four-year college." *The American Economic Review*, 85(3): 600–614.
- Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad.** 2016. "Field of study, earnings, and self-selection." *The Quarterly Journal of Economics*, 131(3): 1057–1111.
- Kohn, Alfie.** 1996. "By all available means: Cameron and Pierceâs defense of extrinsic motivators." *Review of educational research*, 66(1): 1–4.
- Ladd, Helen F., and Edward B. Fiske.** 2001. "The uneven playing field of school choice: Evidence from New Zealand." *Journal of Policy Analysis and Management*, 20(1): 43–64.
- Lance, Lochner.** 2011. "Nonproduction benefits of education: Crime, health, and good citizenship." In *Handbook of the Economics of Education*. Vol. 4, 183–282. Elsevier.
- Lavy, Victor.** 2008. "Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment." *Journal of Public Economics*, 92(10): 2083–2105.

- Lavy, Victor.** 2010. "Effects of free choice among public schools." *The Review of Economic Studies*, 77(3): 1164–1191.
- Lavy, Victor, Olmo Silva, and Felix Weinhardt.** 2012. "The good, the bad, and the average: Evidence on ability peer effects in schools." *Journal of Labor Economics*, 30(2): 367–414.
- Lazear, Edward P.** 2000. "Performance pay and productivity." *American Economic Review*, 90(5): 1346–1361.
- Leventhal, Tama, and Jeanne Brooks-Gunn.** 2000. "The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes." *Psychological bulletin*, 126(2): 309.
- Levin, Henry M.** 1998. "Educational vouchers: Effectiveness, choice, and costs." *Journal of Policy Analysis and Management*, 17(3): 373–392.
- Lindo, Jason M, Nicholas J Sanders, and Philip Oreopoulos.** 2010. "Ability, gender, and performance standards: Evidence from academic probation." *American Economic Journal: Applied Economics*, 2(2): 95–117.
- Mancebón, M. J., J. Calero, Á Choi, and D. P. Ximénez-De-Embún.** 2012. "The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006." *Journal of the Operational Research Society*, 63(11): 1516–1533.
- Marx, Benjamin M, and Lesley J Turner.** 2018. "Borrowing trouble? human capital investment with opt-in costs and implications for the effectiveness of grant aid." *American Economic Journal: Applied Economics*, 10(2): 163–201.
- Mayer, Alexander, Reshma Patel, Timothy Rudd, and Alyssa Ratledge.** 2015. "Designing scholarships to improve college success: Final report on the Performance-based scholarship demonstration."
- McCrary, Justin.** 2008. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of econometrics*, 142(2): 698–714.

- Miller, Cynthia, Melissa Binder, Vanessa Harris, and Kate Krause.** 2011. "Staying on track: Early findings from a performance-based scholarship program at the University of New Mexico."
- Mincer, Jacob.** 1974. "Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2."
- Mizala, Alejandra, and Miguel Urquiola.** 2013. "School markets: The impact of information approximating schools' effectiveness." *Journal of Development Economics*, 103(1): 313–335.
- Montalvo, José Garcia.** 2018. "The impact of progressive tuition fees on dropping out of higher education: a regression discontinuity design."
- Mora, Toni, and Philip Oreopoulos.** 2011. "Peer effects on high school aspirations: Evidence from a sample of close and not-so-close friends." *Economics of Education Review*, 30(4): 575–581.
- Muller, Chandra.** 1997. "The minimum competency exam requirement, teachers' and students' expectations and academic performance." *Social Psychology of Education*, 2(2): 199–216.
- Murillo, Francisco Javier, and Cynthia Garrido.** 2018. "Magnitud de la Segregación escolar por nivel socioeconómico en España y sus Comunidades Autónomas y comparación con los países de la Unión Europea." *RASE: Revista de la Asociación de Sociología de la Educación*, 1(11): 37–58.
- Murillo, Francisco Javier, Cynthia Garrido, and Guillermina Belavi.** 2017. "Segregación escolar por origen nacional en España." *OBETS. Revista de Ciencias Sociales*, , (12): 395–423.
- Murphy, Richard, and Gill Wyness.** 2016. "Testing Means-Tested Aid. CEP Discussion Paper No. 1396. Revised." *Centre for Economic Performance*.
- Musset, Pauline.** 2012. "Shool choice and equity: current policies in OECD countries and literature review." *Education Working Paper*, , (66): 1–51.

- Niederle, M.** 2017. “A gender agenda: a progress report on competitiveness.” *The American Economic Review*, 107(5): 115–119.
- Niederle, Muriel, and Lise Vesterlund.** 2007. “Do women shy away from competition? Do men compete too much?” *The Quarterly Journal of Economics*, 122(3): 1067–1101.
- Niederle, Muriel, and Lise Vesterlund.** 2010. “Explaining the gender gap in math test scores: The role of competition.” *The Journal of Economic Perspectives*, 24(2): 129–144.
- Nollenberger, Natalia, Núria Rodríguez-Planas, and Almudena Sevilla.** 2016. “The math gender gap: The role of culture.” *The American Economic Review*, 106(5): 257–261.
- OECD.** 2015. “Skills Strategy Diagnostic Report. Spain.” *Editions OECD*.
- OECD.** 2016. “Education at a glance 2016.” *Editions OECD*.
- OECD.** 2019a. “Balancing School Choice and Equity. An international perspective based on PISA.” *PISA in Focus*, Éditions OECD(96).
- OECD.** 2019b. “How are school-choice policies related to social diversity in schools?” *PISA in Focus*, Éditions OECD(96).
- Oreopoulos, Philip, and Kjell G Salvanes.** 2011. “Priceless: The nonpecuniary benefits of schooling.” *Journal of Economic perspectives*, 25(1): 159–84.
- Ors, Evren, Frédéric Palomino, and Eloic Peyrache.** 2013. “Performance gender gap: does competition matter?” *Journal of Labor Economics*, 31(3): 443–499.
- Ouazad, A., Page L.** 2013. “Students’ perceptions of teacher biases: Experimental economics in schools.” *Journal of Public Economics*, 105(6): 116–130.
- Patel, Reshma, and Timothy Rudd.** 2012. “Can Scholarships alone help students succeed?”

Reauthorization, HEA. 2013. “Inequality Matters.”

Richburg-Hayes, Lashawn, Colleen Sommo, and Rashida Welbeck. 2011. “Promoting Full-Time Attendance among Adults in Community College: Early Impacts from the Performance-Based Scholarship Demonstration in New York.” *MDRC*.

Richburg-Hayes, Lashawn, Thomas Brock, Allen LeBlanc, Christina H Paxson, Cecilia E Rouse, and Lisa Barrow. 2009. “Rewarding persistence: Effects of a performance-based scholarship program for low-income parents.”

Rothstein, Jesse. 2006. “Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions.” *The American Economic Review*, 96(4): 1333–1350.

Sampson, Robert J, Jeffrey D Morenoff, and Thomas Gannon-Rowley. 2002. “Assessing “neighborhood effects”: Social processes and new directions in research.” *Annual review of sociology*, 28(1): 443–478.

Schudde, Lauren, and Judith Scott-Clayton. 2016. “Pell grants as performance-based scholarships? An examination of satisfactory academic progress requirements in the largest need-based aid program.” *Research in Higher Education*, 57(8): 943–967.

Scott-Clayton, Judith. 2011. “On money and motivation a quasi-experimental analysis of financial incentives for college achievement.” *Journal of Human Resources*, 46(3): 614–646.

Scott-Clayton, Judith. 2017*a*. “Simplifying and Modernizing Pell Grants to Maximize Efficiency and Impact.” *Urban Institute*.

Scott-Clayton, Judith, and Lauren Schudde. 2019. “The Consequences of Performance Standards in Need Based Aid: Evidence From Community Colleges.” *Journal of Human Resources*, 0717–8961r2.

- Scott-Clayton, Judith Elvira.** 2017b. “Undergraduate Financial Aid in the United States.” American Academy of Arts & Sciences.
- Seftor, Neil S, and Sarah E Turner.** 2002. “Back to school: Federal student aid policy and adult college enrollment.” *Journal of Human Resources*, 336–352.
- Shurchkov, Olga.** 2012. “Under pressure: gender differences in output quality and quantity under competition and time constraints.” *Journal of the European Economic Association*, 10(5): 1189–1213.
- Sjoquist, David L, and John V Winters.** 2012. “Building the stock of college-educated labor revisited.” *Journal of Human Resources*, 47(1): 270–285.
- Söderström, Martin, and Roope Uusitalo.** 2010. “School choice and segregation: Evidence from an admission reform.” *Scandinavian Journal of Economics*, 112(1): 55–76.
- Stanley, Marcus.** 2003. “College education and the midcentury GI Bills.” *The Quarterly Journal of Economics*, 118(2): 671–708.
- Steele, C. M.** 1997. “A threat in the air: How stereotypes shape intellectual identity and performance.” *American psychologist*, 52(6): 613.
- Terrier, Camille.** 2016. “Boys Lag Behind: How Teachers’ Gender Biases Affect Student Achievement.” *Browser Download This Paper*.
- Teske, Paul, and Mark Schneider.** 2001. “What research can tell policymakers about school choice.” *Journal of Policy Analysis and Management*, 20(4): 609–631.
- Tiebout, Charles M.** 1956. “A pure theory of local expenditures.” *Journal of political economy*, 64(5): 416–424.
- Turner, Lesley J.** 2017. “The Economic Incidence of Federal Student Grant Aid.” *University of Maryland, College Park, MD*.
- Yang Hansen, Kajsa, and Jan Eric Gustafsson.** 2016. “Causes of educational segregation in Sweden – school choice or residential segregation.” *Educational Research and Evaluation*, 22(1-2): 23–44.

List of Tables

1.B.1	Number of BCG applicants (2010–2015).	74
1.B.2	Descriptive Statistics on Undergraduate Applicants for Different Treatment Samples (2010–2015).	75
1.B.3	Balance of Baseline Characteristics for Different Treatment Samples (2010–2015).	76
1.B.4	Discontinuities in Allowance Amounts, GPA and Fraction of Credits Earned at Different Income Eligibility Thresholds and Periods. . .	77
1.B.5	Discontinuities in Official Dropout from higher education at T1 and T2 grants by period.	78
1.B.6	Discontinuities in Average GPA and Fraction of Credits Earned at Different Income Eligibility Thresholds and Periods by Term. . . .	79
1.B.7	Discontinuities in Average GPA (0-10) at T1 and T2 grants by period and subgroup of applicants.	80
1.B.8	Discontinuities in the Fraction of Credits Earned (0-1) at T1 and T2 grants by period and subgroup of applicants.	81
1.B.9	Discontinuities for the mechanisms variables at T1 and T2 grants by period.	82
1.B.10	Discontinuities in Degree Completion at T1 and T2 grants by period.	83
1.C.1	External Validity.	86
1.C.2	McCrary (2008) Test for Manipulation of the Forcing Variable for Different Treatment Samples in FW and T3 grant.	88
1.C.3	McCrary (2008) Test for Manipulation of the Forcing Variable for Different Treatment subamples.	90

1.C.4	Average Allowance Amounts (in euros) at T1 and T2 Discontinuities by period and subgroup sample.	92
1.C.5	Discontinuities in Average Allowance Amounts at t+1 for T1 and T2 grants by period.	93
1.C.6	Discontinuities in Average GPA at T1 and T2 grants by period. . .	98
1.C.7	Discontinuities in Fraction of Credits Passed at T1 and T2 grants by period.	99
1.C.8	Discontinuities in Average Awarded Grant and Average GPA at Fee Waiver (FW) and Displacement and other needs grant (T3) by period.	100
1.C.9	Discontinuities in Average Awarded Grant, Average GPA, Fraction of Credits Earned and Dropout at Residence Grant and Compensate Grant in 2012.	101
1.C.10	RDD-DID Estimates for T1 Grant.	102
1.C.11	Discontinuities for the mechanisms variables at T1 and T2 grants by period and term.	103
1.C.12	RDD-DID Estimates for T1 Grant.	105
1.C.13	Difference in Baseline Means by period and treatment sample. . .	106
1.C.14	Logistic models of net dropout rate: Academic year 2011-12 vs. 2012-13.	109
1.C.15	Logistic models of yearly dropout rate (2010–2015).	110
1.C.16	DID model of yearly dropout rate (2010–2015).	111
1.C.17	BCG grant academic requirements.	113
2.A.1	Summary Statistics.	142
2.A.2	Treatment-Control balance in fixed school characteristics in 6th and 10th Grade.	143
2.A.3	Baseline Estimates.	144
2.A.4	Average student performance by gender and internally/externally administered test.	145
2.A.5	Robustness Checks I. Baseline Estimates including all DATs. . . .	146

2.A.6	Robustness Checks II. Baseline Estimates including School Fixed Effects.	147
2.A.7	Robustness Checks III. Baseline Estimates including school controls.	148
2.A.8	Heterogeneous Effects by School Ownership.	149
2.A.9	Heterogeneous Effects by School's Fraction of Parents with College Education.	149
2.A.10	Heterogeneous Effects by Year.	150
2.A.11	Questions Analysis: Multiple-choice vs. Non-multiple-choice.	150
2.A.12	Heterogeneous Effects by Main Teacher Gender.	151
2.A.13	Survey Answers Baseline Estimates for 6th Grade Students.	152
2.A.14	Survey Answers Baseline Estimates for 10th Grade Students.	153
2.B.1	Treatment-Control balance in fixed school characteristics in 6th Grade for 2016/2017 and 2017/2018.	154
2.B.2	Description of Schools Characteristics Variables.	155
2.B.3	Description of Student Survey Questions.	155
2.B.4	Heterogeneous Effect by Teacher Experience.	156
2.B.5	Heterogeneous Effect by Teacher Experience.	157
3.B.1	Priority Points in case of over-demand of schools in the city of Madrid.	210
3.B.2	Summary Statistics: School applicants in the city of Madrid over 2010-2016.	211
3.B.3	School Assignment in the city of Madrid between 2010/2011 and 2016/2017.	211
3.B.4	Effect of the reform on Out-of-District Assignment and Distance to Assigned School. Difference-in-Difference Approach.	212
3.B.5	The Effect of the School Choice Reform on School Segregation by Parental Education and Immigrant Status.	212
3.C.1	Tie-break criteria in the region of Madrid.	213
3.C.2	Municipalities with the single-zone school choice setting.	214

3.C.3	Total population with 3-year old children in the Region of Madrid by Immigrant status.	216
3.C.4	Housing Prices and School Average Performance.	217
3.C.5	Sample Restrictions: School Applicants in the city of Madrid over 2010-2016.	221
3.C.6	Treatment-Control balance in fixed households characteristics in pre-reform years.	222
3.C.7	Sensitivity Analysis: Effect of the reform on Out-of-District Assignment and Distance to Assigned School by treatment group definition. . .	223
3.C.8	Change in within-school student heterogeneity by parental education.	227
3.C.9	Districts ranked by income.	227

List of Figures

1	Inter-generational Mobility, Gender Differences, and School Choice.	5
1.A.1	Income eligibility thresholds for the different levels of the BCG grant.	69
1.A.2	Amount of annual cash allowance awarded to applicants with 4 family members, as a function of their parents' taxable income by period	69
1.A.3	McCrary (2008) test for 2010–2015	70
1.A.4	Fraction of Awarded a Conditional Grant for T1 and T2 Discontinuities (2010–2015).	70
1.A.5	Average Grant Amounts for T1 and T2 Discontinuities by Period. .	71
1.A.6	Average GPA (0-10) for T1 and T2 Discontinuities by Period. . . .	71
1.A.7	Average GPA (0-10) for T1 and T2 Discontinuities in Period II by Term	72
1.A.8	Fraction of Credits Earned (0-1) for T1 and T2 Discontinuities over Period I and II by Term	72
1.A.9	Mechanisms for Spring term at T1 Discontinuity (2013–2015) . . .	73
1.C.1	Fraction of re-applicants and McCrary (2008) test for re-applicants density.	89
2.A.1	Distribution of internally vs. externally schools in 6th Grade. . . .	139
2.A.2	Distribution of internally vs. externally schools in 10th Grade. . . .	139
2.A.3	Permutation Test for 6th Grade Students.	140
2.A.4	Permutation Test for 10th Grade Students.	140
2.A.5	Survey Questions by gender and subject in 6th Grade.	141
2.A.6	Survey Questions by gender and subject in 10th Grade.	141

3.A.1	School Districts in the city of Madrid.	200
3.A.2	Parental Education by census block in 2013.	200
3.A.3	Out-of-District Assignment and Average Distance to Assigned School by group of treatment.	201
3.A.4	Outer School District Applications: Top-Ranked School vs. Assigned School.	201
3.A.5	Average Distance to Top-Ranked School vs. Assigned School. . . .	202
3.A.6	Impact of the Reform on Out-Of-District Assignment by Parental Education.	202
3.A.7	Impact of the Reform on Out-Of-District Assignment by Immigrant Status.	203
3.A.8	Impact of the Reform on Out-Of-District Assignment by Distance to the Closest School District Border.	203
3.A.9	School Segregation (M Index).	204
3.A.10	Within vs. Between School District Segregation.	204
3.A.11	School vs. Residential Segregation.	205
3.A.12	The Effect of the School Choice Reform on School Segregation by Parental Education.	205
3.A.13	The Effect of the School Choice Reform on School Segregation by Immigrant Status.	206
3.A.14	Phasing-in of the Reform in the Municipalities of the Region of Madrid.	206
3.A.15	Average Out-of-Municipality Assignment and Average Distance to Assigned School by municipality size.	207
3.A.16	Effect of the School Choice Reform on Out-of-Municipality Assignment and Average Distance to Assigned School by municipality size. . . .	207
3.A.17	School Segregation by Immigrant Status by Municipality Size. . . .	208
3.A.18	School Segregation by Immigrant Status: Within vs. Between Municipality.	208

3.A.19	The Effect of The School Choice Reform on School Segregation by Immigrant Status and Municipality Size.	209
3.C.1	Census blocks in the city of Madrid.	214
3.C.2	Average Performance of Assigned School (0-10) by Parental Education.	218
3.C.3	Average Performance of Assigned School (0-10) by Immigrant Status.	218
3.C.4	District Income (euros) of Assigned School by Parental Education.	219
3.C.5	District Income (euros) of Assigned School by Immigrant Status. .	219
3.C.6	Predetermined School Segregation of District of Assigned School by Parental Education.	220
3.C.7	Predetermined School of District of Assigned School by Immigrant Status.	220
3.C.8	Densities of within-school socioeconomic variation and immigrant share, by year.	225