European
University
Institute

DEPARTMENT
OF
ECONOMICS

Essays in Economics of Education

David Martinez De Lafuente

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 12 February 2021

European University Institute
Department of Economics

Essays in Economics of Education

David Martinez De Lafuente

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Andrea Ichino, EUI, Supervisor
Prof. Sule Alan, EUI
Prof. Manuel Bagues, University of Warwick
Prof. Caterina Calsamiglia, Institute of Political Economy and Governance (IPEG)

Researcher declaration to accompany the submission of written work

I, David Martinez de Lafuente, certify that I am the author of the work "Essays in Economics of Education" I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. [quotation from it is permitted, provided that full acknowledgement is made.] This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Statement of inclusion of previous work (if applicable):

I confirm that chapter 3 was jointly co-authored with Lucas Gortazar and Ainhoa Vega-Bayo, and I contributed 70% of the work.

Signature and Date: October 5, 2020

DAVID MARTINEZ DE LAFUENTE

Abstract

This thesis consists of three independent essays in economics of education.

In the **first chapter**, I investigate the connection between cultural identities and parental schooling decisions. By leveraging the case of the Basque Country (Spain), this essay studies how parents trade off academic quality for being educated in the regional language. Using a discrete choice structural model, I show that households display strong preferences for the Basque-monolingual model. Results indicate a willingness to forego a substantial amount of mean academic performance to evade the Spanish and the bilingual models. By means of regression analysis, I find a strong association between nationalistic voting and educational language choices. This suggests that schooling decisions are significantly shaped by parents' affiliation to the regional culture.

In the **second chapter**, I test whether the cultural assimilation efforts of immigrant families mitigate discriminatory attitudes of schools. To this end, I sent fictitious visit requests to more than 2,500 schools located in the Community of Madrid (Spain). I find that Romanian families who gave a Spanish name to their child are 50% less discriminated than those who selected a Romanian name. Emails from families whose members have Romanian names are 12% less likely to receive a response than those from native Spanish-name families. The results show a consistent response pattern across school characteristics.

The **third chapter**, co-authored with Lucas Gortazar and Ainhoa Vega-Bayo, studies the presence of systematic differences between teacher non-blind assessments and external quasi-blindly graded standardized tests. We use a rich administrative database covering two cohorts from publicly-funded schools in the Basque Country. We find that systematic underassessment exists for boys, children with immigrant origin, and poorer students. The results indicate that stereotyping is a consistent mechanism through which our findings can be interpreted.

Nire Gurasoei / To My Parents

Acknowledgements

This thesis marks the end of a long journey. During these four years, I have been immensely lucky for the many people that have accompanied me.

I want to begin by thanking my supervisors. This thesis would not have been possible without your dedication and honesty. Thanks to Andrea Ichino. Your encouraging words and critiques have been an important source of motivation. You have been an excellent *sensei*, and I am grateful for having learned so much from you. Thanks to Sule Alan for your very useful suggestions and comments. You encouraged me at the right times, and inspired me with your enthusiasm.

I am indebted to Jérôme Adda, who provided guidance at the early stages of the thesis; to Juan J. Dolado, for his continuous nearness and helpful suggestions; and to Manuel Bagues, for his thoughtful comments as a member of my Thesis Committee.

I am also very grateful to my co-authors, Lucas Gortazar and Ainhoa Vega-Bayo, for our work together. I want to particularly thank Lucas; an excellent mentor and colleague. You have taught me so much about education. To a great extent, my passion for this field comes directly from you. Special mention goes to Sara de la Rica. Thank you for all the opportunities you have given me. I do not think this thesis could have been completed without your help.

I have been extremely fortunate for the wonderful people I have shared these years with. To Giulia, thank you for all the invaluable support and the countless hours spent together. You brought so much joy into this adventure. Thanks to Gonzalo, for helping me navigate the ups and downs of this roller coaster, for being an excellent friend, and a great flatmate. To Iacopo, for encouraging me to "go bigger", for all the good laughs, and for our ramen dinners. You inspired me with your work ethic and cheerful attitude. I am also thankful to my flatmates in San Marco, David and Carlos; and to Ana, Martina, and Ane Miren, for those weekend nights together. Your company made this experience much better. I cannot help but thank my EUI colleagues Mustafa, Daniele, Flavia, Lazar, and Roger. I feel fortunate for having met you and for having shared many good times.

The final part of this thesis was written in Barcelona. I am especially grateful to my mentor at IPEG, Caterina Calsamiglia, for her time, her kindness, and for being part of my Thesis Committee. Your priceless advice has been fundamental for the first chapter of this dissertation. *Moltes graciès*. I would like to also thank Ane and Enara, who hosted me with great hospitality and made my time there far more enjoyable.

I cannot emphasize enough how important my friends at *Palado* have been throughout these years. For the evenings in *Mume* and the long nights in Bilbao and elsewhere, I am incredibly thankful. The countless laughs I shared with you were my perfect source of disconnection from work. *Eskerrik asko*.

Last, but not least, I want to thank my family. I am indebted to my loving grandparents for their infinite generosity. To my beautiful sister and my beloved *Izeko*, thank you for being there whenever I needed. And, above all, to my parents. This dissertation is dedicated to you. For your constant support, for providing me with the best education, for believing in me, and for your unending affection and care; I am deeply grateful. Thank you so much.

Contents

1 Identity and School Choice: Parental Preferences for Language Educational Models

1.1 Introduction

Identity formation is a lifelong process characterized by the development of associations to different social groups. Although long-lasting, this process is significantly shaped at an early age through the adoption of family socialization decisions and other interactions with peers. Parental choices, like which school to attend, can strongly influence offspring's self-image, core values and identification with certain groups [Akerlof & Kranton (2000, 2002)]. To the extent that parents are biased towards their own identity, they may adopt deliberate actions to restrict children integration into other oppositional cultures [Bisin & Verdier (2001)]. Therefore, families may weigh their preferences for intergenerational transmission of identity with their regard for other desirable outcomes for their children.[1]

School choice provides an interesting setting to study this topic for several identity cleavages. For example, the religiosity, the class sentiment or the national identity can be incentivized through choices between sectarian and nonsectarian schools, private and public schools or, in multilingual communities, between different linguistic models. Specifically, this chapter focuses on the latter. The recognition of various linguistic models in education significantly shapes the dynamics of national identity formation. For instance, Clots-Figueras & Masella (2013) find that students with longer exposure to Catalan teachings, following the compulsory adoption of the local language in schools, had strengthened Catalan sentiments. Therefore, language instruction, through its connection to a cultural identity, can substantially interfere in schooling choices. In particular, if parents have strong preferences for cultural transmission,

[1]This type of behavior can be sustained with *imperfect empathy*. It requires parents to be altruistic with their offspring, but by making choices based on their own preferences. For example, a religious parent may care about their children success, but oppose them embracing secular values [Bisin & Verdier (2011)].

they may concede academic quality to sort their children into the linguistic model associated with their own identity.

In this essay, I examine this trade off by leveraging an ideal case study: that of the Basque Country. This region, located in northern Spain, is constituted by a bilingual community and gathers two identities that differ in their attachment to the Basque language and culture. To investigate the role of cultural identity in schooling choices, the study site brings a setting with several unique features. First, the two officially recognized languages, Basque and Spanish, provide disparate private economic returns outside the region. While the Basque language is a *language isolate* with no well-known connections to other existing languages; Spanish is a global language with more than 500 million speakers in over 20 countries, including the US. Second, with the provision of several linguistic models, Basque families have the ability to sort themselves into schools that match their cultural identity. In turn, this explains the relatively stable evolution of the Spanish and Basque identities in the region [Aspachs-Bracons *et al.* (2008)].[2] Third, the two languages display little lexical relatedness, and thus exhibit limited learning complementary. Fourth, the Basque Country is highly homogeneous along the ethnic and religious cleavages. Therefore, the Basque cultural identity is likely to have a strong linguistic basis, especially after the local language was banned during Franco's dictatorship. Altogether, these elements indicate that the study of language model choices is informative towards understanding the role of identity affiliations in schooling decisions.

To answer the question at hand, I develop a structural model to estimate the unobserved preferences for schools governing parental choices. Every year, parents express their preferences by exercising their right to choose with an ordered ranking of schools. In the model, parental preferences depend on the school's linguistic model, the school-home distance, the ownership of the school, its amount of ethnic diversity, and the size of the amenities. Families differ in their preferences for school quality and socioeconomic composition based on their income, and are affected by private-taste shocks. Schools are filled using the Parallel Mechansim (PM) [Chen & Kesten (2017)], an assignment algorithm that is not *strategy-proof*.[3] To account

[2] The authors find that, while the compulsory bilingual policy implemented in Catalonia intensified Catalan identity feelings; the introduction of the current Basque choice policy did not alter the development of individual identity.

[3] In early 2018, the allocation mechanism was modified to the Deferred Acceptance (DA) algorithm.

for the manipulable nature of the PM, households are strategic and choose their payoff-optimal list by considering their admissions probabilities to the different schools.

To understand the determinants of parental schooling choices, I fit my model to rich data from administrative sources. In particular, I use parents' applications to pre-primary schools in Bilbao, the largest city in the region. With these data, I estimate the preference weights that rationalize applicants' observed choices via Simulated Maximum Likelihood. By expressing households' preferences as a linear combination of school characteristics, I assume that parents trade off these attributes against each other. In this manner, I compare the preference parameters attached to the different linguistic models with those from school quality.

Several findings emerge from this analysis. First, parents are, on average, willing to concede a significant amount of quality to attend the Basque linguistic model. In particular, I find a willingness to forego 0.7-1.0σ of mean test scores to avoid the bilingual model. In contrast, my results present a remarkably high analogous trade off (2.0-2.6σ) for the Spanish monolingual model, that is only attended by 4% of the students. The latter finding is consistent with families displaying a strong aversion towards this option [Vega-Bayo & Mariel (2019)], and parents exhibiting a moderate valuation of schools' academic quality. Second, depending on their income, families display heterogeneous preferences for schools' quality and student body composition. Specifically, there is a monotonic association between test scores and the socioeconomic composition with respect to households' income terciles. For instance, a 1 unit increase in mean test scores raises parents' program valuations by 4.1 utils in the lowest income tercile, but it does so by 5.1 and 5.3 utils for households in the second and the highest terciles. This finding is consistent with the previous literature [e.g. Hastings *et al.* (2009), Glazerman & Dotter (2017), Abdulkadiroğlu *et al.* (2020)]. Finally, I find that parents positively but modestly appraise semi-public schools, the ethnic composition, and school surface. Overall, the qualitative nature of the results is robust to the use of a rank-ordered logit specification, which is a suitable model under the assumption of truthful reporting of preferences.

The observed trade off between linguistic choices and academic quality need not solely reflect identity considerations. One important data limitation is the lack of cultural background and the academic ability from applicants. This aspect of heterogeneity is not considered in the model, and thus the results are limited in their capacity to capture matching effects. To test for the presence of an identity channel, I correlate the census units' electoral outcomes and the aggregate patterns of linguistic choices. Results indicate that the share of votes to non-nationalist parties is significantly and negatively associated with the decision of enrolling children in the Basque model. However, it is possible that other mechanisms are confounding the results. On the one hand, I show that significant differences exist in the observed academic results between subjects and language models. The results may therefore reflect heterogeneous preferences in the relevance attached to these subjects. On the other hand, parents might anticipate higher learning difficulties if they enroll their children in a language they are not fluent in. This could limit parents' willingness to attend an extraneous linguistic model, absent identity concerns.

This chapter adds to several strands of research. First, it speaks to the growing literature on cultural transmission and identity economics. Since Akerlof & Kranton (2000) and Bisin & Verdier (2001), an increasing number of papers have looked at the causes and economic consequences of social norms and identity affiliations along several dimensions. Among others, these include articles investigating the economic impact of first names [Fryer Jr & Levitt (2004), Algan *et al.* (2013)], the dynamics of nation-building policies [Almagro & Andrés-Cerezo (2020)], the nature of political discourse [Glaeser (2005)], or the determination of attitudes towards redistribution [Shayo (2009)]. Interestingly, the field of education has also benefited from the thriving interest in identity in areas like student attainment [Schüller (2015)] and effort [Akerlof & Kranton (2002)], indoctrination [Voigtländer & Voth (2015)], or the construction of national identities [Aspachs-Bracons *et al.* (2008), Clots-Figueras & Masella (2013)]. This chapter arguably expands this body of work by studying the intensity of parental preferences for the intergenerational transmission of identity via educational language choices.

Second, it contributes to the empirical literature of parental preferences estimation. Since Hastings *et al.* (2009), there has been a significant increase in the amount of work using preference data from centralized assignment mechanisms. Because of its attractive theoretical properties, researchers have mainly focused on the Deferred Acceptance (DA) mechanism [Gale & Shapley (1962)].[4][5] In contrast, the body of work examining preferences in manipulable

[4] These include, among others, Burgess *et al.* (2015), Harris & Larsen (2015), Glazerman & Dotter (2017), Abdulkadiroğlu *et al.* (2020) and Beuermann *et al.* (2019).

[5] The DA is a *strategy proof* mechanism, i.e. submitting rankings in order of true preferences is a weakly dominant strategy [Abdulkadiroğlu & Sönmez (2003)]. However, Haeringer & Klijn (2009) and Calsamiglia *et al.* (2010) show that, under certain circumstances, truth-telling might not be optimal when parents have truncated lists.

systems is relatively scant and focuses on the Boston Mechanism (BM) [He (2016), Hwang (2016), Agarwal & Somaini (2018), Calsamiglia *et al.* (2020), Kapor *et al.* (2020)]. The emphasis of these articles is on the efficiency and welfare properties of the BM. This essay is different in several ways. On the one hand, existing papers typically explore the role of peer characteristics and school effectiveness in student preferences. Instead, I take a different direction to investigate the role of linguistic choices. On the other hand, I focus on preference estimation under the PM, for which there is to my knowledge no comprehensive study.[6]

The remainder of this chapter is organized as follows. The next section describes the institutional background. Section 1.3 presents the model. Section 1.4 introduces the data sources and the sample restrictions. Section 1.5 summarizes the estimation approach. Section 1.6 presents the main findings. Section 1.7 provides evidence on mechanisms. Section 1.8 discusses the limitations of the chapter and Section 1.9 concludes.

1.2 Institutional Setting

1.2.1 The Basque Cultural and Political Context

The Basque Country is an autonomous community located in the eastern side of northern Spain, bordering France. In the Basque 1978 Statute of Autonomy, the region defines its population as a *nationality* in recognition of its differentiated collective sociocultural and linguistic identity.[7] As such, the area is constituted by a bilingual community, with two officially recognized languages: Basque and Spanish.

The Basque language or *Euskara* is the last remaining Pre-Indo-European language in Western Europe. Linguistically, it is considered a *language isolate* with no demonstrable connections with other languages. Moreover, the origins of Basque still spark debates among scholars given its significant distinctiveness with respect to their neighboring Romance languages. This feature gives rise to several interesting observations. First, the uniqueness of *Euskara* implies that the Basque Country represents a very different case relative to other

[6]The PM, a hybrid between the DA and the BM, is the assignment mechanism employed in Shanghai and other several provinces.

[7]I use the term Basque Country to refer to the Spanish autonomous community *Euskadi*, which is made up of three provinces: *Bizkaia*, *Araba* and *Gipuzkoa*. The Statute of Autonomy collects the basic institutional norms of the autonomous community and thus it constitutes the *foundational* document of the region in its present form.

multilingual regions, like Catalonia or Switzerland. The large linguistic distance between Spanish and Basque suggests that there exists limited complementarity between these two language skills. This aspect reinforces the importance of learning the language through formal education, especially for non-Basque-speaking families. Second, the antiquity of Basque, together with Spain's homogeneity along the religious and racial cleavages, suggests that the Basque identity has a strong linguistic basis [Echeverria (2003), Gardeazabal (2011)]. Interestingly, the term *Euskaldun*, used to refer to a Basque person, literally means "that who has Basque" in *Euskara*.

The Basque language experienced a period of intense repression during the Franco dictatorship (1936-1975), when its use was prohibited and punished. With the introduction of democracy, the demands for the recognition of the local language arouse, also in the education sector. Ever since and shaped by the presence of the terrorist organization ETA (1958-2018), the nationalist issue has been at the forefront of Basque politics. This feature has shaped the coexistence of two broad identity groups along the nationalist ideological cleavage, which has an arguably strong connection with the affiliation to the Basque culture and language. In this regard, a majority of citizens are identified as nationalist, in light of the electoral results favoring these parties throughout the decades.[8]

Over the last years, the use of Basque has increased significantly. For instance, between 1991 and 2016 the proportion of active Basque-speakers increased from 24.1% to 33.9%.[9] However, the region displays substantial heterogeneity. For example, in Bilbao only 18.6% are Basque-speakers, while 60.3% of the population is monolingual Spanish-speaker.

Because of its limited geographical spread and minority use in most parts of the region, the private economic returns to Basque can be described as modest. In the retailing sector, a majority of workers and business-owners (70%) use only Spanish to communicate among them or with their clients and suppliers.[10] Yet, access to employment in some areas of the public sector require formally accredited knowledge of Basque. These include, among

[8]Nationalist voting accounted for 59% and 67% of votes in the 2016 and the 2020 regional elections, respectively. I consider as nationalist parties both PNV (*Partido Nacionalista Vasco*, the center-right christian democrat nationalist party) and EH Bildu (the left-wing Basque independentist party).

[9]Source: 2016 Sociolinguistic Survey. A significant proportion of the population (*passive Basque-speakers*) have Basque knowledge but do not speak it regularly (19.1%).

[10]Source: Basque Government report from 2003, based on a survey to retail businesses. Link to the report: **Here**

others, teachers, police and local administration. Nevertheless, the public employment (14% of working population) has a relatively low weight compared to the national (15%) or the OECD (18%) average. Altogether, this emphasizes the idea that Basque knowledge is mostly associated with the non-monetary returns stemming from the conservation of and identification with the local culture.

1.2.2 The Basque Education System

Spanish regional authorities are granted with significant autonomy in education policy-making. As a consequence, the Basque government has full control over the regulation of its education network, as long as it does not contradict the basic national guidelines. Since 1982, the Basque system is organized around three linguistic models that differ in the emphasis given to Basque and Spanish. First, the model A teaches all subjects, except Basque and English, in Spanish. Second, the bilingual model B uses both Basque and Spanish, with both languages receiving similar weights. In contrast, the model D employs only Basque, except for the Spanish and English subject. Although Basque is not the dominant language, 75% of students attend model D and only 4% are enrolled in model A.[11] For exposition purposes, I will hereby denote model A as Spanish, model B as bilingual, and model D as Basque.

Throughout the last decades, public authorities have been promoting the adoption of Basque to expand its general use. Since 1983, the proportion of students in the Basque model has increased significantly (more than 50 percentage points), and the number of schools using the Spanish-monolingual model has decreased. In their regulatory framework, the Basque Department of Education explicitly states its willingness to adjust the linguistic supply based on demand considerations. This suggests that the expansion of the Basque model has been predominantly driven by evolving parental preferences.

The public education system is composed by two education networks that have similar size, but differ in their management and financing criteria. Public schools are fully financed by the regional governments and are free of charge. By contrast, semi-public or *concerted* schools are both publicly and privately funded, but have financial freedom over the distribution of resources. Public sources account approximately for 60% of the total per pupil cost in

[11] Source: Eustat for academic year 2017/18. It considers students in kindergartens (0-5 years old), primary schools (6-11 years old), and middle school (12-16 years old).

concerted schools. For this reason, semi-public schools are de facto allowed to charge parents for private donations to cover their operating expenses.[12] Despite these regulatory disparities, public and semi-public schools depend on the same admission procedure. Additional details about the rules and design of the centralized assignment mechanism to schools are provided in the next subsection.

1.2.3 School Choice in the Basque Country

Parents typically make their schooling choices when children are two years old, age at which the schooling rate is 93.1%.[13] Although every family has the right to access a publicly-funded school, each specific school has a fixed capacity, and can therefore be over-demanded. Since providing every parent with their preferred choices might not be possible, the public authority establishes a centralized mechanism that regulates the assignment process. I now proceed to describe the timing of the assignment procedure, the legal criteria for prioritizing applications, and the particularities of the assignment mechanism design.

Timing of the assignment procedure. Between January and February, parents that want to enroll their children in the public system need to submit a ranking with their most preferred schools. Families can choose up to three different schools, and in each of them, the linguistic model they prefer, in preference order. I denote the combination of a school and linguistic model as a *program*. Thus, students can submit a ranking of up to nine programs in three different schools (i.e., three programs per school, with a maximum of three schools). After all interested applicants submit their preferences, the regional authority implements a centralized mechanism that places students to schools. The provisional assignment list is made public in late March. Parents can then present claims on the provisional list or withdraw their enrollment request. After such considerations are recognized, the Department of Education publishes the final assignment list in April. Finally, enrollment takes place in September, before the beginning of the academic year.

[12] In 2012, parents devoted approximately 707.6€ per year for basic education services in *concerted* schools.

[13] In the academic year 2016/17, 39.5% of applications to public schooling were for two years old children, followed by those beginning first year of middle school (12 years old, 25.2%) and high school (16 years old, 15.2%).

Prioritization criteria. If schools are over-demanded, applications need to be ordered based on a priority rule established by the regional authority.[14] A description about the points that students receive based on their personal characteristics can be found in Appendix 1.10.1. Without loss of generality, the amount of priority-points sc_{ij} student i obtains if she applies to program j is given by: (i) whether family's residence l_i is inside school j's catchment area, z_j; (ii) the presence of family members in the school (fam_{ij}), and (iii) other characteristics of the family and of the student-school match, g_{ij}. I denote student i's priority-type t_{ij} the realization of personal characteristics $(fam_{ij}, l_i \in z_j, g_{ij})$ relative to program j. Based on her t_{ij}, applicant i gets $sc_{ij} = \Phi(t_{ij})$ points following the official prioritization rule.

Assignment Mechanism. Once parents submit their school ranking, schools are filled using the Parallel Mechanism. This mechanism is described as a hybrid of the Boston Mechanism and the Deferred Acceptance Mechanism by Chen & Kesten (2017). Similar to the BM and DA, the PM gradually fills schools by considering applications in sequential rounds. The process is repeated for R rounds, where R is the maximum number of programs parents can list. The PM is characterized by the use of choice-bands (l). Choice-bands consist of a number of rounds after which assignments are made permanent. The assignment is composed by two choice-bands. Let $l_1 < l_2$ be the number of cumulative rounds at which the first and second choice-bands are completed. In the Basque system, $l_1 = 3$ and $l_2 = 9$.[15] I now proceed to formally describe the PM procedure.

- *Round 0:* A single lottery number is drawn for each student.
- *Round* $r = 1$*:* Each student applies to her most preferred program. Each school considers its applicants. For each program, students who have listed them as their first option are tentatively assigned following the priority criteria, from hight to low, one at a time. This is done until either there are no seats left or there is no student left who has listed them as their top choice. The remaining students are discarded.
- *Round* $r \geq 2$*:* Rejected students then apply to their next favored program. Each school considers its applicants with those students who have been conditionally accepted in the

[14] In 2016, the admission criteria of students to public and semi-public schools for the academic year was regulated by the Article 9 of Decree 35/2008.

[15] In my sample, schools only provide up to two programs. Thus, the effective choice bands are defined at $l_1 = 2$ and $l_2 = 6$.

previous round. Students are again tentatively assigned following the priority criteria until either there are no seats left or there is no student left who has listed them as their r^{th} choice.[16]

- If $r = l_1$ or $r = l_2$, all tentative assignments are made final, and school capacities are shortened by the amount of assigned students.

Note that parents rejected from their top-listed school can only access their second- or third-ranked school if there still are seats available after the first choice-band. Hence, the PM prioritizes students at higher-ranked options. Therefore, parents have incentives towards misrepresenting their true preferences. In particular, similar to the BM, families have two motives for strategic reporting under the PM [Agarwal & Somaini (2018)]. First, because assignments are final after the first choice-band, parents have incentives for skewing their ranking towards schools for which they have higher admission probabilities. Second, since the length of the ranking is truncated, families may avoid listing too many programs for which they have limited entry options. Thus, parents face incentives towards not only considering their true ex-post utility of attending a school, but also their admission risks. Section 1.10.2 in Appendix presents suggestive empirical evidence of household strategic behavior using regression analysis.

1.3 Model

I consider a population of N students applying to public or semi-public schools. In what remains, I use the terms student, household and family interchangeably. There are J programs provided by S schools distributed across the city. Programs are defined as a combination of a school s and an instructional model $m \in \{A, B, D\}$. Households are indexed as $i \in \{1, ..., N\}$, programs as $j \in \{0, ..., J\} = \mathcal{J}$ and schools as $s \in \{1, ..., S\} = \mathcal{S}$. I denote program 0 as being left unassigned by the mechanism. Established exogenously, a positive fraction of schools supply two programs that vary in their linguistic model. The remaining schools grant solely one program. Let $\mathcal{J}_s \subset \mathcal{J}$ be the subset of programs from school s and $\mathcal{S}_2 \subset \mathcal{S}$ denote the subset of schools that supply two programs. Each family is required to submit an application

[16] If $r > l_1$, the amount of points corresponding to criteria associated with the first requested schools are deducted.

that includes a ranking of programs. After the official deadline, a planner implements the assignment mechanism described above, taking into account the population of lists submitted by parents, the priority rules and school fixed capacities.

1.3.1 Household Preferences

Each household i is endowed with personal characteristics c_i and a location l_i. Families have private information about the preferences over the assignment to each program. This information is unobserved by the researcher. Let $v_{ij} \in \mathrm{I\!R}$ be parents i's utility to assignment to program j and $\mathbf{v}_i = \{v_{i0}, ..., v_{iJ}\}$ stand for their vector of random utilities. Denote $d_{ij} = d(l_i, l_j) \in \mathrm{I\!R}_{++}$ the distance between l_i and the location of school s where program j is provided (i.e. l_j).

As a location normalization, I set the ex-ante value of being left unassigned by the mechanism to zero (i.e. $v_{i0} = 0$). Results should therefore be interpreted as being relative to remaining in the waiting list. I assume that parent i's utility from attending j is given by:

$$v_{ij} = \delta fam_{ij} + \mathbf{Z}_j^{'}\alpha + \mathbf{X}_{ij}^{'}\beta - d_{ij} + \epsilon_{ij}, \tag{1.1}$$

where $fam_{ij} = 1$ if applicant i has a family member (i.e. a parent working or a sibling enrolled) in the school of program j. Here, $\mathbf{Z}_j^{'}$ is a vector of program characteristics, that includes dummies on whether the program is Spanish or bilingual, an indicator on whether the school is semi-public, the fraction of foreign-born students, and the size of school amenities.[17] Similarly, $\mathbf{X}_{ij}^{'}$ is a vector of student-school attributes, i.e. interactions between income terciles with, on the one hand, instructional quality of program j, and on the other, its student body composition. Finally, ϵ_{ij} is an i.i.d idiosyncratic private-taste shock of i over program j with $\epsilon_{ij} \sim N(0, \sigma_\epsilon)$.

The primary restriction of the specification is its additive separability form plus the independence of the errors. With the former, I allow parents to compare the above attributes against each other to form a valuation of each program. To identify the scale parameter σ_ϵ, I set the coefficient of d_{ij} equal to -1. This type of utility representation is commonly used in

[17] In the estimation, I add one dummy variable if there is missing data about program's characteristics other than its linguistic model and the school ownership.

the school choice literature [Agarwal & Somaini (2018), Calsamiglia *et al.* (2020), Kapor *et al.* (2020)].

1.3.2 Beliefs over Assignment Probabilities

I assume that agents believe that they are small players and take admission probabilities as given. Families do not know other households' preferences, v_{-i}, or submitted rankings, A^*_{-i}. Instead, they have perfect information about the priority-score cutoff ($s\bar{c}_j$) and the choice band ($\bar{b_j}$) at which program j gets filled. Thus, they perceive no uncertainty about programs' cutoffs. Under these assumptions, admission probabilities can be almost entirely characterized by $s\bar{c}_j$ and $\bar{b_j}$. This feature brings relative computational simplicity to solving the model. I provide the formal definition of beliefs in Section 1.5.

In the model, students have two sources of uncertainty over the assignment probabilities in tie-breaking situations. First, households need to submit their ranking before their lottery number is drawn, and therefore face uncertainty over their tie-breaker. Second, families do not know the distribution of submitted rankings from other students, and thus are not aware of the priority-types they will encounter in tie-breaking circumstances.[18]

1.3.3 Household Problem

To model the strategic behavior of agents, I assume that parents best respond given the beliefs over assignment probabilities, and choose the ranking that maximizes their expected payoff.

Two relevant features are worth discussing based on the assignment mechanism described in Section 1.2.3. First, because the mechanism tries to assign students to their most preferred options, the rankings' payoff depend on the order in which programs are listed. This implies that the solution to the problem corresponds to choosing the best ranking over all possible permutations. Second, certain schools offer two distinct language models. In their rankings, families can list up to three schools and, within each school, sort the programs they wish in preference order. This has relevant implications for how choices can be materialized. Assume, for instance, that household i considers listing school s, which uses the bilingual and Basque models, as their favorite candidate. In her application, student i can rank school s in four

[18] In the current version of this work, I assume that parents believe that priority-types are uniformly distributed. This induces a slight deviation from purely rational expectations.

different ways: (i) with the bilingual model only, (ii) with the bilingual model as first option and the Basque as second option, (iii) with the Basque model as first option and the bilingual as second, or (iv) with the Basque model only. Then, she can proceed to apply up to two additional schools, following the same logic. Hereafter, I denote the alternative ways a school can be listed based on their program supply as *flow-rankings*. Let $a \in \mathcal{A}$ be any plausible flow-ranking, $|a| \in \{1,2\}$ refer to its length and $a(h) \in \mathcal{J}$ constitute the h^{th} element of a, where $h \in \{1,2\}$.[19]

The latter feature of the application process implies that households can construct their rankings by listing up to three flow-rankings. Given the relatively large number of programs and schools available ($J = 68$ and $S = 63$), solving the model is computationally demanding. To deal with this task, I adapt the solution method developed in Calsamiglia *et al.* (2020). This procedure exploits the sequential nature of most allocation systems to deal with the high dimensionality of the problem. In the assignment, the k^{th} listed program is only significant if one has been rejected by the previously ranked $k-1$ programs. Thus, the k^{th} choice needs to be optimal, conditional on reaching that stage in the assignment mechanism. Altogether, this means that the problem can be solved by means of backward induction.

Let $A_i^* = (a_{i1}^*, a_{i2}^*, a_{i3}^*)$ denote the optimal list for household i, where a_{ik}^* is the k^{th} listed flow-ranking. Defined by c_i and l_i, household i has priority-type t_{ij} in program j, and gets $sc_{ij} = \Phi(t_{ij})$ points. Let $p_j^r(t_{ij})$ be the admission probability to program j in application round r for household i. In what follows, I drop the dependence on this object and denote it as p_j^r for notational convenience. Finally, denote as r_k the application round at which the first element of the k^{th} listed flow ranking is considered in the allocation process.[20]

The backward induction method solves the optimal list starting from a_{i3}^*. Intuitively, the procedure checks whether each flow-ranking is optimal given that the student was rejected from all previous programs. The procedure starts from the lowest-ranked option, and constructs the optimal list by sequentially considering higher-ranked alternatives. Given (c_i, l_i, ϵ_i), a_{ik}^* needs to solve the following problem for $k = \{3,2,1\}$:

[19] Formally, $\mathcal{A} = \{\mathcal{J} \cup \left(\cup_{s \in \mathcal{S}_2} \pi_2(\mathcal{J}_s)\right)\}$, where $\pi_2(\mathcal{J}_s)$ is the set of permutations of order 2 over set $\mathcal{J}_s$. Given that $\mathcal{S}_2 \neq \emptyset$, $|a| \in \{1,2\}$.

[20] Because schools only supply up to two programs, $r_1 = 1$, $r_2 = 3$ and $r_3 = 5$.

$$V^k(c_i, l_i, \epsilon_i) = \max_{a \in \mathcal{A}} \{R^k(a, c_i, l_i, \epsilon_i)\}, \tag{1.2}$$

$$s.t.\ \{R^k(a, c_i, l_i, \epsilon_i)\} = \begin{cases} \{p^{r_k}_{a(1)} v_{ia(1)} + (1 - p^{r_k}_{a(1)}) V^{k+1}(c_i, l_i, \epsilon_i)\} & \text{if } |a| = 1, \\ \\ \Big\{p^{r_k}_{a(1)} v_{ia(1)} + (1 - p^{r_k}_{a(1)}) p^{r_k+1}_{a(2)} v_{ia(2)} + \\ (1 - p^{r_k}_{a(1)})(1 - p^{r_k+1}_{a(2)}) V^{k+1}(c_i, l_i, \epsilon_i)\Big\} & \text{if } |a| = 2, \end{cases} \tag{1.3}$$

$$V^4(.) = v_{i0} = 0, \tag{1.4}$$

where $V^{k+1}(c_i, l_i, \epsilon_i)$ is the continuation value and (c_i, l_i, ϵ_i) comprise the state variables. Argument (1.2) returns a^*_{ik} by selecting the flow-ranking that provides the highest expected utility, given the state variables. Condition (1.3) breaks down the computation of the expected payoffs depending on the length of the flow-ranking. Finally, condition (1.4) introduces the normalization for being left unassigned by the mechanism.

One important remark is that several lists may generate the same expected payoff. Consider, for example, an optimal ranking where $p^1_{a^*_{i1}} = 1$. Then, any ranking $A'_i = (a^*_{i1}, a'_{i2}, a'_{i3})$ is payoff-equivalent to A^*_i. I denote as $\Lambda^*(c_i, l_i, \epsilon_i)$ the set of optimal lists, that yield the same payoff.[21]

1.4 Data

1.4.1 Data Sources and Sample Selection

The Department of Education (*Hezkuntza Saila*) provided application data and enrollment records for the academic year 2016/17. The former cover the entire population of applicants (approximately 42,000 students) that participate in the assignment process to public and semi-public schools. They contain, for each student, her submitted ranking, the amount of priority points received according to her personal characteristics in the first option[22], her home

[21]Rankings in this set, including A^*_i, yield the same allocation outcome and are equivalent up to the payoff-relevant part.

[22]The data available only describe the realization of the criteria for the top-ranked school. This implies that I cannot infer the points households would acquire for other schools if they were top-ranked for two criteria that

address, her country of birth, and the result of the assignment algorithm. By contrast, the latter consist of the enrollment records for public and semi-public schools in non-university grades, with available information for more than 380,000 students. These data contain students' demographics and personal characteristics; including gender, eligibility to financial aid or parents' ID, among others.

Additionally, the Basque Institute for Research and Evaluation in Education (ISEI-IVEI) provided data with the results of standardized tests from 9-10 and 13-14 years old students. The ISEI-IVEI is the Basque public agency in charge of evaluating the quality of the education system, and with that intent administers low-stake standardized tests to the student population attending 4^{th} grade of primary and 2^{nd} grade of middle school. It does so every two years since 2011, and its evaluation encompasses several subjects; including Math, Science, and languages. To proxy for the academic quality of each program, I obtain the GPA for each student and average the result across students in the program.[23] Alternatively, I use their Socioeconomic and Cultural Index (hereafter, ISEC) to proxy for the socioeconomic composition of the school. The variables used in the model are described in Table 1.6.

One important restraint of the aforementioned quality measure is that it conflates peer academic quality and value added. Reliably disentangling these two factors requires a significant amount of data that unfortunately is not available. However, previous evidence indicates that while parental choices are shaped by test scores [see, for instance, Black (1999) or Figlio & Lucas (2004)], they are not responsive to school effectiveness [Mizala & Urquiola (2013), Abdulkadiroğlu *et al.* (2020)]. A potential explanation for this finding is that parents use test scores to proxy for school quality, given that value added is typically hard to observe. Thus, my measure of academic quality arguably constitutes a good metric for *perceived* quality. Notwithstanding the above-mentioned limitation, I follow several papers on the literature of

are specific to each school. In particular, these are: (i) being a partner of the school (1 point) and (ii) fulfilling some other discretionary considerations set by each school (up to 2 points). To deal with this limitation, I impute the counterfactual amount of points on these dimensions if the school was top ranked by relying on certain assumptions. The imputation procedure is described in Appendix 1.10.3.

[23]To construct the GPA, I use the scores in Math, Spanish and Basque. I exclude the use of Science and English for two reasons. First, they display a higher incidence of missing values that significantly reduces the sample size. Second, these subjects have a lower weight in determining student progression. In middle school, grade retention is decided when students do not pass a certain number of subjects. This threshold is equal to two when the failed subjects are Basque, Spanish and Math, but it is three subjects otherwise.

preference estimation that use mean test scores as a proxy for academic quality [Burgess *et al.* (2015), Calsamiglia *et al.* (2020)].

To perform the analysis, I complement the preceding administrative databases with two additional sources. First, I use the Atlas of Household Income Distribution (ADRH) from the National Institute of Statistics (INE) to obtain the average household income at the census unit level. This allows me to impute applicants' income based on their residence, using high resolution data. Second, I use geospatial data with the geographic delimitation of the schooling zones from Bilbao Data Lab. With this information, I derive the residence-based priority points applicants get for every program. Figures 1.4 and 1.5 illustrate respectively the geographical distribution of catchment areas and average income across the city of Bilbao.

For estimation, I focus on families with two years old children whose home address is in Bilbao, and are applying to public or semi-public schools located in the city in 2016. I exclude 191 families whose assignment is inconsistent with the official rule, and whose residence location cannot be consistently matched. As a result, the final sample consists of 1,846 applicants.

1.4.2 Descriptive Statistics

Table 1.1 presents the summary statistics of households from the analytical sample. The results are disaggregated by income terciles. Overall, 92.3% of students are assigned to their top ranked school, suggesting that families are strategic players and consider their admission probabilities to the different programs. According to their economic status, households in upper terciles submit, on average, longer rankings (33.5% rank three schools compared to 28.7% in the lowest tercile). In contrast, a higher proportion remain unassigned after the implementation of the mechanism (2.9%).

The characteristics of the top-listed school differ across income groups. As household income increases, parents display a higher preference for semi-public schools. Furthermore, there exists an economic gradient in the selection of linguistic models. In particular, I observe a monotonic increase in the proportion of families that choose the bilingual model as income goes up. The Basque monolingual model is the most demanded (70.4%), while the Spanish monolingual is only chosen by 2.1% of students. On average, families have 8.4 schools that

Table 1.1 Applicant Characteristics

HOUSEHOLD CHARACTERISTICS	Tercile = 1 (1)	Tercile = 2 (2)	Tercile = 3 (3)	Overall (4)
Have older sibling(s)	37.7%	38.3%	38.2%	38.1%
# Schools Listed = 2	48.0%	52.5%	50.9%	50.4%
# Schools Listed = 3	28.7%	33.1%	33.5%	31.7%
Assignment: 1^{st} ranked school	93.2%	92.4%	91.2%	92.3%
Assignment: 2^{nd} ranked school	3.9%	5.4%	4.9%	4.7%
Assignment: 3^{rd} ranked school	1.1%	1.0%	1.0%	1.0%
Assignment: Left Unassigned	1.8%	1.2%	2.9%	2.0%
Top Choice: Semi-public	47.3%	60.9%	73.7%	60.6%
Top Choice: Linguistic Model = Spanish (A)	3.0%	0.5%	2.8%	2.1%
Top Choice: Linguistic Model = Bilingual (B)	20.8%	26.5%	35.3%	27.5%
Top Choice: Linguistic Model = Basque (D)	76.2%	73.0%	61.9%	70.4%
# Schools in-zone	8.316	7.933	9.006	8.420
	(2.536)	(2.631)	(2.840)	(2.707)
Average home-school distance (meters)	2,238.8	2,218.0	1,952.6	2,136.5
	(480.5)	(565.2)	(414.1)	(506.9)
Distance to top-listed school (meters)	607.4	638.0	668.6	637.9
	(525.3)	(598.0)	(600.1)	(575.6)
N	621	608	617	1,846

Notes: Standard deviations in parenthesis.

grant them the maximum amount of priority points based on their residence (i.e. *in-zone* schools). In terms of the average home-school distance, wealthier families have a higher convenience to access city schools. This is not surprising given that they are typically located closer to the city center. However, the distance to the top listed school is larger for this group. This suggests that parents with higher socioeconomic status display a larger willingness to travel for their children's education [Gortazar *et al.* (2020), Glazerman & Dotter (2017)].

Table 1.2 summarizes the program characteristics of the estimation sample. The linguistic model supply differs between the public and semi-public network. In particular, public schools

Table 1.2 Program Characteristics

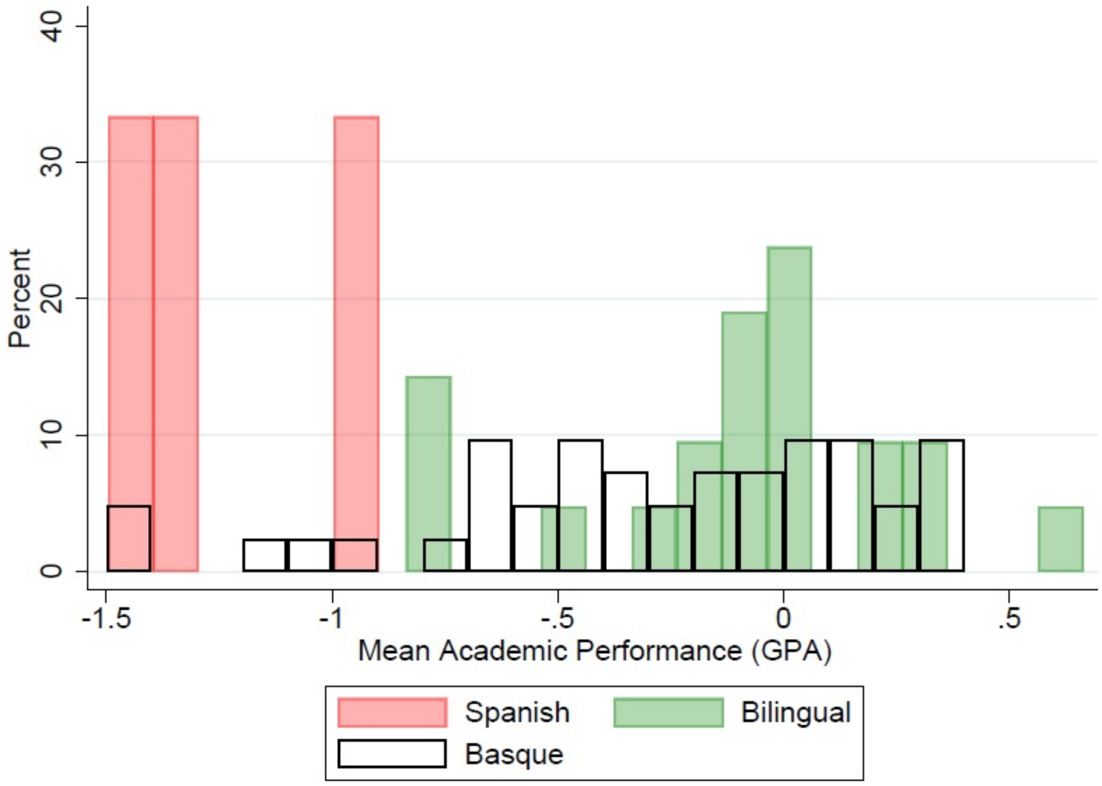
PROGRAM CHARACTERISTICS	Public (1)	Semi-Public (2)	Overall (3)
Linguistic Model = Spanish (A)	8.6%	3.03%	5.88%
Linguistic Model = Bilingual (B)	0%	66.7%	32.4%
Linguistic Model = Basque (D)	91.4%	30.3%	61.7%
% of foreign-born students[a]	6.7%	5.2%	6.0%
	(4.39)	(5.78)	(5.09)
Academic Quality: Average GPA z-score[b]	-0.447	-0.0649	-0.271
	(0.527)	(0.367)	(0.405)
Peer composition: ISEC Index[a]	-0.777	0.0291	-0.405
	(0.747)	(0.665)	(0.813)
Surface[a]	6.076	11.26	8.469
	(4.219)	(12.15)	(9.118)
N	35	33	68

Notes: Standard deviations in parenthesis. Academic quality is measured with the students' GPA from Math, Basque and Spanish using standardized test scores in primary education. It is expressed with a program-specific average from students' z-score (computed using the regional mean and standard error). The ISEC index stands for the average Socio-Economic and Cultural Index of the school in primary education (provided by ISEI-IVEI). 3 programs from semi-public schools have missing information with regards to variables flagged with [a] and [b]. [a]: Variables defined at the school level. [b]: Variable defined at the program level.

specialize in both monolingual models. Conversely, 66.7% of the programs from semi-public schools are bilingual. On average, public schools have a slightly higher proportion of foreign-born students. Additionally, the average quality and ISEC index of semi-public schools are higher. However, public schools display a higher dispersion in both of these measures. Finally, the average size of the amenities from semi-public schools is larger than that from public entities.

Figure 1.1 describes the program quality distributions by language models. The bars show the fraction of programs with certain mean standardized test scores, conditional on the linguistic option. First, consider the Spanish-monolingual alternative. These programs exhibit significantly worse academic performance (mean $= -1.27$) than their bilingual and Basque counterparts. Furthermore, the highly skewed score distribution displays no common support with programs from the bilingual model, and it only overlaps with the bottom tail

Figure 1.1 Program Quality Distribution by Language Models



Notes: This figure plots the histograms of the program quality measures, separately for each language model. The academic quality measure uses the program-specific average of students' GPA scores from Math, Basque and Spanish in primary education. **Sample Size by Language Models:** Spanish-monolingual = 3, Bilingual = 21, Basque-Monolingual = 42. Three programs from semi-public schools have missing information of their performance; one in each language model.

of the Basque-monolingual option. In contrast, the bilingual and Basque alternatives display significant heterogeneity in academic performance, and show substantial overlap in their distribution supports. Nevertheless, the average performance of bilingual programs (mean $= -.10$) is slightly higher than that from Basque-monolingual programs (mean $= -.28$).

The above metric depicts a unidimensional representation of academic performance. However, programs might display significant across-subject heterogeneity. To examine this matter, Figure 1.6 shows the analogous subject-specific histograms. The qualitative nature of the results remain with the exception of the Basque subject. Here, the Basque-monolingual programs are the highest performing alternatives.

1.5 Estimation

To estimate the parameters of interest $\theta = (\delta, \alpha, \beta, \sigma_\epsilon)$, I use a two-step estimation process. First, I calculate the assignment probabilities from the universe of applications by computing the programs' clearing cutoffs and by using a *resampling* approach. Second, I estimate the preference parameters θ using the simulated probabilities in the first step via Simulated Maximum Likelihood. I now turn to discuss both of these estimation phases.

1.5.1 First Step: Assignment Probabilities

To estimate the assignment probabilities, I use application and assignment data from all households that participated in the 2016 assignment process. With this information, I calculate the market-clearing conditions, which are characterized by the specific choice-band ($\bar{b}_j$) and the priority points cutoff ($s\bar{c}_j$) at which every program j is filled.

Following Calsamiglia *et al.* (2020), I define the admission probabilities as follows:

$$p_j^r(t) = \begin{cases} 1 & \text{if } r < l_{\bar{b}_j - 1} \text{ or } (l_{\bar{b}_j - 1} + 1 \leq r \leq l_{\bar{b}_j} \text{ and } \Phi(t) > s\bar{c}_j), \\ \hat{p}_j^r(t) & \text{if } l_{\bar{b}_j - 1} + 1 \leq r \leq l_{\bar{b}_j} \text{ and } \Phi(t) = s\bar{c}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.5}$$

where $\hat{p}_j^r(t)$ is the simulated probability of assignment to program j in round r if $\Phi(t) = s\bar{c}_j$, i.e. in case of ties at the priority points cutoff. The need for simulating these probabilities comes from the tie-breaking method at hand. Unlike several mechanisms that only use lottery numbers, the Basque Country first prioritizes applicants' personal characteristics. Specifically, applications are sorted by comparing several criteria, one by one, in the order presented in Appendix 1.10.1. The random lottery number is solely used in case families display the same characteristics. This feature prevents the use of a closed-form approximation to these probabilities. To deal with this limitation, I use a resampling approach that is based on Hortaçsu

& McAdams (2010) to compute these probabilities.[24] Section 1.10.3 in Appendix provides additional details about the simulation algorithm.

1.5.2 Second Step: Estimation of Preference Parameters

By taking the beliefs over admission probabilities as given, I proceed to estimate the preference parameters via Simulated Maximum Likelihood. The model needs to maximize the probability of households' observed application rankings, given applicants' characteristics, the attributes of the schools and the assignment probabilities. Let $\tilde{A}_i$ be the observed ranking submitted by household i. The contribution of student i to the likelihood is given by:

$$L_i(\theta) = \int I\Big(\tilde{A}_i \in \Lambda^*(c_i, l_i, \epsilon_i; \theta)\Big) dF_\epsilon(\epsilon; \sigma_\epsilon), \tag{1.6}$$

where $\Lambda^*(c_i, l_i, \epsilon_i; \theta)$ is the model predicted set of optimal rankings. The log-likelihood of the estimation sample is: $L(\theta) = \sum_i \log L_i(\theta)$.

It is well known that probit probabilities need to be numerically approximated given that they do not have a closed-form representation. Here, I use the Accept-Reject (A-R) simulator for this task. The method was originally proposed by Lerman & Manski (1981), and it is carefully described in Train (2009). To optimize the likelihood, I employ the Nelder-Mead algorithm, a commonly used direct-search gradient-free simplex method for multidimensional optimization problem [Nelder & Mead (1965)]. Further details about the estimation procedure are presented in 1.10.3.

1.6 Results

Table 1.3 summarizes the estimated household preference parameters. To provide a comparative baseline, column 1 displays the results from a standard exploded logit model. Under truthful reporting, this type of specification constitutes a computationally convenient method to estimate preferences from ranked data. However, it does not consider households' admission risks to the

[24] This method has been adapted by Agarwal & Somaini (2018) and Calsamiglia *et al.* (2020) to the school choice context. My approach differs to theirs in that I only use this method to compute the assignment probabilities for priority-types that might be subject to ties. Instead, they simulate the whole assignment procedure multiple times and record individual assignments in each simulation.

different programs, and therefore misses a relevant component of their strategic behavior. The specification from column 2 expands the model by adding students' probability of admission to each program in the first choice-band as an additional covariate [Beuermann *et al.* (2019)]. Finally, column 3 presents the results from the structural model. Columns 1 and 2 do not reliably capture parents' incentives for strategic misreporting, and thus they should not be interpreted directly. Yet, the qualitative nature of the results remains fairly stable across the three specifications.

The first row reports the estimated linear distance cost. Consistent with previous research, the results from columns 1 and 2 suggest that parents dislike more remote schools. This indicates that the scale normalization of the distance cost is set with the appropriate sign, which takes value -1 in column 3. The next row shows that, all things equal, households disfavor public schools compared to semi-public schools. Despite public schools are free of charge, a preference for the larger management autonomy and supply of extracurricular activities from semi-public schools rationalizes this finding. Unfortunately, I do not have school fees data to disentangle the sole impact of tuition costs.

The fourth and fifth row present, respectively, the preference parameters for the Spanish-monolingual and the bilingual models. The results suggest that, on average, parents find the Basque-monolingual D model substantially more preferable. The utility cost is about 840 meters for the Spanish model and 315 meters for the bilingual model. The aversion to these language models is deemed as economically large, especially for the former. Two features support this statement. First, in a medium-sized city like Bilbao, schools typically do not provide transport services. Therefore, most parents need to walk with their children to the school. Second, home-school distances are measured in the Euclidean space. Thus, the observed associations imply considerably larger effective walking distances. Altogether, these entail relatively long commuting times for a two years old child, and constitute significant daily time losses for working parents. The finding that language models comprise a highly relevant choice consideration is also found in Vega-Bayo & Mariel (2019). Using a Discrete Choice Experiment, they find that the main language of instruction is the most relevant school characteristic defining parental preferences in the Basque Country, and that parents display a strong aversion towards the Spanish monolingual option.

Table 1.3 Exploded Logit and Structural Model Results

	Exploded Logit		Structural Model
VARIABLES	(1)	(2)	(3)
Distance (100m)	-0.259***	-0.244***	{-1}
	(0.004)	(0.004)	
Public school	-0.152***	-0.225***	-0.395
	(0.054)	(0.057)	
Semi-public school	{0}	{0}	{0}
Ling. Model = Spanish (A)	-1.916***	-2.843***	-8.396
	(0.166)	(0.172)	
Ling. Model = Bilingual (B)	-0.433***	-1.095***	-3.152
	(0.049)	(0.055)	
Ling. Model = Basque (D)	{0}	{0}	{0}
Tercile 1×Academic Quality	0.918***	1.514***	4.056
	(0.170)	(0.172)	
Tercile 2×Academic Quality	0.710***	1.488***	5.123
	(0.164)	(0.165)	
Tercile 3×Academic Quality	0.346**	0.956***	5.251
	(0.167)	(0.165)	
Tercile 1×ISEC index	-0.180	-0.122	2.038
	(0.112)	(0.113)	
Tercile 2×ISEC index	0.263**	0.262**	2.867
	(0.110)	(0.108)	
Tercile 3×ISEC index	0.584***	0.659***	3.617
	(0.112)	(0.111)	
% foreign-born students	-0.065***	-0.060***	0.026
	(0.008)	(0.008)	
Surface	0.007***	0.013***	0.05
	(0.003)	(0.002)	
Probability of admission (if top-listed)		1.490***	
		(0.056)	
Family member in school			104.47
Taste-shock dispersion (σ_ϵ)			10.094
N (applicants)	1,846	1,846	1,846

Notes: {0} and {-1} imply that the parameters are constrained to 0 and -1. Robust standard errors in parenthesis for columns 1 and 2. Bootstrapping standard errors for column 3 to be estimated. The estimation includes a dummy variable if information of program quality, ISEC index, immigrant composition, or size of amenities is missing, but the coefficient is suppressed here.
$*p < .1$,$**p < .05$,$***p < .01$.

Rows 6 through 8 exhibit the distance-quality trade off for each income group. The impact of program quality is modelled with a tercile-specific linear function. Contrary to the logit results, the structural parameters from column 3 indicate a positive monotonic association between income and preferences for school quality. A one unit increase in the GPA z-score is associated with utility gains ranging from 406 and 525 meters, depending on the income group.

In light of the above findings, one can conclude that parents face a quantitatively significant trade off between linguistic choices and school quality. The estimates suggest that families are willing to concede around 1.96-2.55σ and 0.74-0.96σ of mean GPA test-scores to avoid the Spanish and the bilingual models, respectively.[25] Two factors drive the sizable nature of this trade off. On the one hand, households display an intense inclination towards the Basque monolingual model and an acute reluctance of the Spanish model. On the other, the estimated impact of school quality is fairly modest in quantitative terms.

Next, rows 9 through 11 show the impact of peer composition by income groups. The three columns display a positive and monotonic association of income with the schools' average ISEC. The finding that parents differ in their regard for the student body composition is consistent with the literature [e.g. Hastings *et al.* (2009), Glazerman & Dotter (2017), Abdulkadiroğlu *et al.* (2020)]. The following two rows explore the impact of schools' ethnic composition and the size of amenities. For the former, it is interesting to note that the sign of the effect changes between the logit and the structural model. However, given the modest dispersion in the share of foreign-born students between schools, the economic significance of this variable is deemed almost zero. With regards to school surface, I find that, holding everything else equal, households value larger schools positively. Yet the linear term is very small, and thus not deemed as economically significant.

The presence of a sibling studying or a parent working in the school is an attribute that parents find specially appealing. This is not surprising given that this feature seems particularly convenient for parents to minimize commuting times. Finally, the last row displays the dispersion of private taste-shocks (σ_ϵ), which takes a rather small value.

[25]These effects are computed by dividing the ratio of the language models' coefficients to those of academic quality with the sample standard deviation of the quality measure.

1.7 Discussion of Potential Mechanisms

So far, I have examined the average preference for different school attributes, including language model choices and instructional quality. However, what drives the presence of heterogeneous linguistic choices? Because of missing data about applicants' cultural orientation and language skills, the model abstracts from relevant sources of heterogeneity that are left unexplained. To shed light on this matter, I empirically investigate three mechanisms through which the diverse linguistic decisions can be interpreted.

1.7.1 Identity Affiliations and Language Choices

The first possibility is that language model choices reflect parents' regard for the intergenerational transmission of their own identity. This hypothesis is consistent with evidence that families in the Basque Country sort their children into classrooms based on cultural affiliations [Aspachs-Bracons *et al.* (2008)]. In their seminal paper, Bisin & Verdier (2001) introduce the notion of *imperfect empathy*, a behavioral friction that rationalizes this type of parental decisions. Intuitively, families judge choices for their offspring based on their own preferences, and thus they favor the cultural transmission of their own identity trait. This type of bias is sufficient for explaining opposing language decisions by emphasizing the existence of non-monetary returns to matching children education to one's identity.

To empirically test for the presence of an identity channel, I use data from the 2016 Basque Regional Elections. In particular, I correlate the aggregate patterns of education linguistic choices with the electoral results at the census unit level. Ideally, I would study the association between family-specific attachments to the Basque identity and their language model decisions. Unfortunately, these data are not available. To avoid this limitation, I instead utilize the vote share to non-nationalist parties to proxy for less intense attachment feelings to the Basque culture in a certain area.[26] Because there exists a tight connection between cultural sentiments and party affiliations, this exercise is informative for the presence of an identity channel.

[26] I consider as Non-Nationalist the following political parties: PP (*Partido Popular*), PSE-EE (*Partido Socialista Euskadi-Euskadiko Ezkerra*) and C's (*Ciudadanos*). I have excluded *Podemos* (*Podemos Ahal Dugu-IU*) because of their mixed stance with respect to the nationalist issue. Although they belong to a State-wide party with federalist vocation, the Basque branch of the party demanded *" a new territorial status"* for the Basque Country and defended the region's *"right to self-determination"*.

Table 1.4 Electoral Results and Linguistic Choices (aggregates at census unit level)

DEPENDENT VARIABLE	% applicants rank Model A or B		% applicants top-rank Model A or B	
	(1)	(2)	(3)	(4)
% Vote Non-Nationlist	1.290***	0.876***	1.434***	1.202***
	(0.234)	(0.241)	(0.201)	(0.201)
N	265	265	265	265
Income tercile	✗	✓	✗	✓
R^2	0.083	0.175	0.152	0.190

Notes: Robust standard errors are in parentheses. Electoral outcomes and income terciles data come from the *National Institute of Statistics* (INE). Income tercile 2 is the omitted category. The constant term and the two tercile coefficients are not reported in the interest of saving space. The proportions of applicants ranking Model A or B are obtained by aggregating the estimation sample onto census units. The 2018 definition of census units is used. Due to some minor mismatching, 26 applicants are left out of the regression sample ($N = 1,829$).
$*p < .1$,$**p < .05$,$***p < .01$.

Regressions use census tracks as the unit of analysis, which imply relatively accurate results given the narrow population size of these geographical units (around 1,000-2,500 inhabitants).

Table 1.4 summarizes the regressions results. The findings from column 1 indicate that there exists a significant and positive association between the proportion of non-nationalist voters and the share of households listing a Spanish-monolingual or bilingual program in their ranking. Column 3 suggests a similar link between non-nationalist voting and the proportion of applicants ranking models A or B as their top-choice. Columns 2 and 4 show that the results are robust to the inclusion of income tercile dummies from the census unit.[27] Altogether, it seems that a higher presence of non-nationalistic voting is associated with a higher propensity for avoiding the Basque monolingual model. These findings are consistent with the view that identity considerations significantly shape students' sorting into the different linguistic classes.

1.7.2 Skill Formation and Language Models

A second hypothesis is that language models specialize in different skill sets for which families have heterogeneous preferences. If parents regard certain abilities, they are likely to sort into the model that better develops these skills. Hence, disparate outcomes between language options may result in heterogeneous choices, even in the absence of identity considerations.

[27] I replicated the analysis using individual applicant choices and including additional controls. The results do not significantly change. Additionally, the findings are robust to the inclusion of *Podemos* as a non-nationalist party. Results are displayed in Table 1.7.

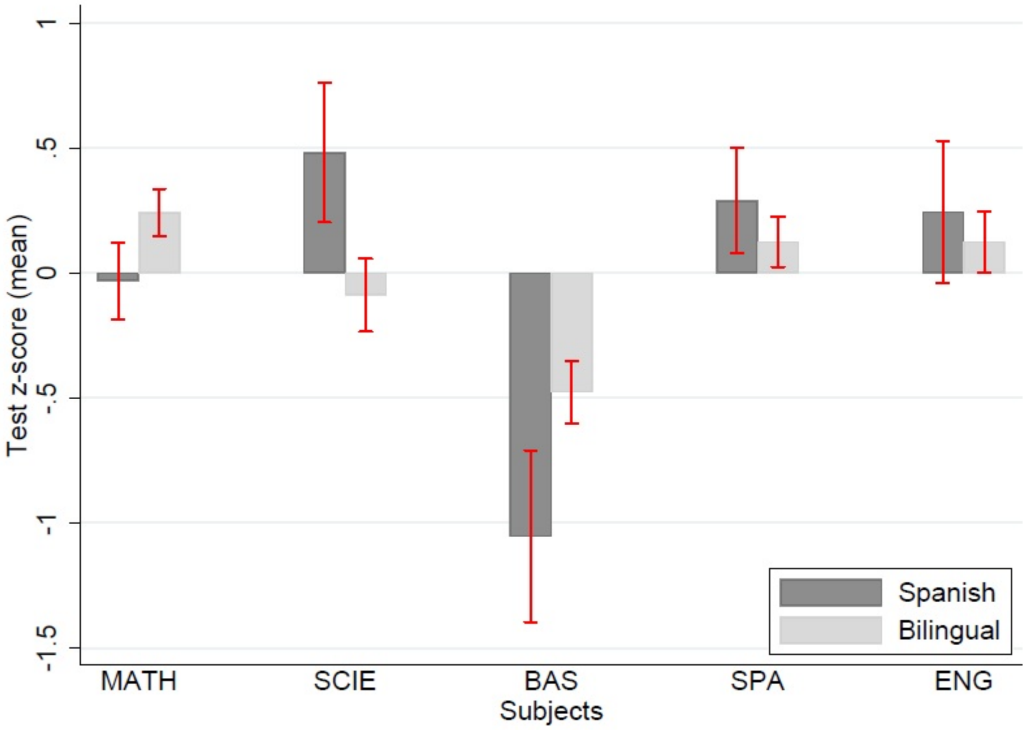
To empirically investigate this channel, I use the test records from the aforementioned ISEI-IVEI database. These assessments produce measures of cognitive abilities about Math, Science, Basque, Spanish and English for 9-10 and 13-14 years old students. Focusing on the 2016 cohort, I explore the observed differences between linguistic models by subject. For this purpose, I employ individual-level regressions linking the different test scores with the model students attend. To control for nonrandom sorting, I use the student's GPA in other skills (cubic specification), month of birth, gender, origin, income, home-speaking language, special-needs and grade retention indicators as covariates. I also add cubics of the school socioeconomic composition and average test score from other subjects to control for peer effects. With the introduction of variables conditioning linguistic selection, this analysis provides of a more accurate comparison of the skills formation between language models than that discussed in Section 1.4.

Figure 1.2 reports the results for 4^{th} grade primary students from Bilbao. The results reflect the differences with respect to the Basque monolingual model, that is the omitted category. As a robustness check, I replicated the analysis using the entire sample from the Basque Country. The qualitative nature of the findings prevail. Interestingly, I find that students from the Spanish model better perform in Science (0.52σ, $p = .000$), Spanish (0.29σ, $p = .009$) and English (0.24σ, $p = .095$). In contrast, the bilingual model is associated with higher test scores in Math (0.24σ, $p = .000$).[28] Finally, pupils from the Basque-monolingual model are clearly dominant in Basque language skills. The differences are 1.05σ and 0.47σ relative to the Spanish and bilingual options (both with $p = .000$). These differences remain constant at higher education levels, given the similar results found for middle school students.

The analysis abstracts from relevant unobserved dimensions students are sorted along. Therefore, the findings should not be interpreted causally but as merely indicative of whether meaningful disparities exist between language models. Altogether, the above evidence suggests that this is the case, and that overall, the Spanish and bilingual models display a comparative advantage in skills other than Basque. Consequently, some student sorting along this dimension is justified if parents have heterogeneous preferences for distinct abilities. In light of the large differences in Basque dexterity, one might argue that foregoing the observed extent of Basque

[28] Compared with the Basque model, the bilingual model displays higher scores in Spanish (0.12σ, $p = .017$) and English (0.12σ, $p = .051$).

Figure 1.2 Differences in Subject-Specific Sills, relative to the Basque-monolingual Model



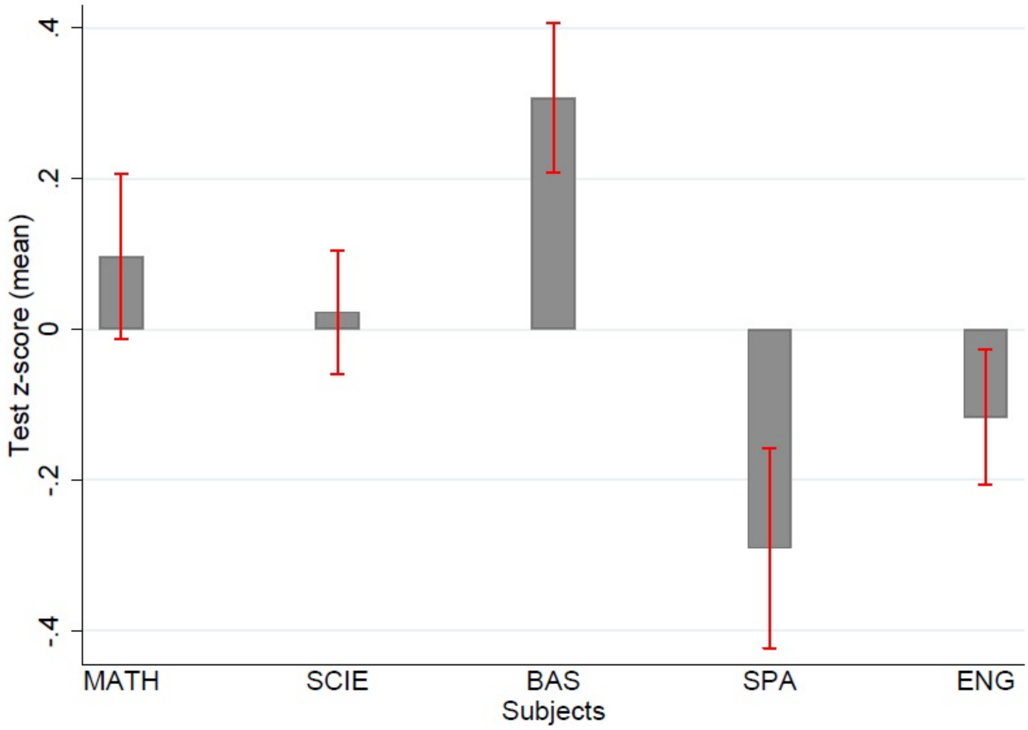
Notes: OLS estimates (and confidence intervals) for the association between Model A and B and the z-score in ISEI-IVEI assessments, by subject. Results are relative to Model D (omitted category). **Sample:** Students of 9-10 years old living in Bilbao ($N = 2,211$). **Sample sizes by language models:** Spanish - 77, Bilingual - 889 and Basque - 1,245. **Individual Controls:** cubic specification of GPA of other skills (e.g. if Math regression, I consider the GPA of Science, Basque, Spanish and English), dummy for female, set of indicators for whether the student is first- or second-generation immigrant (or from unknown origin), month of birth (normalized value, 0 if December and 1 if January), income-groups that define scholarship eligibility, dummy for whether the student is Basque home-speaker, special-needs indicator and grade retention dummy. **School level controls:** Cubic specification on the school ISEC index and on the school-level average of GPA test score in other skills. **Robust standard errors. Clustered at the school level.**

skill formation by choosing other language models is likely to be associated with identity considerations. Therefore, this mechanism can be interpreted as complementary, and not substitute, to the identity channel.

1.7.3 Learning Differences and the Language Spoken at Home

One reason some families may find the Basque option less appealing is that they do not dominate the language. Anticipating difficulties to assist their children throughout the learning process, parents are possibly attracted to other options that find more convenient. If students that do not speak Basque at home perform systematically worse, sorting along language models may reflect learning differences based on the maternal language.

Figure 1.3 Basque-Speaking Premium in the Basque-monolingual model



Notes: OLS estimates (and confidence intervals) for the association between being a Basque home-speaker and the z-score in ISEI-IVEI assessments, by subject. Results are relative to not speaking Basque at home (omitted category). **Sample:** Students of 9-10 years old living in Bilbao and studying in the Basque model $(N = 1{,}245)$. **Basque Home Speakers:** $N = 207$. **Individual Controls:** cubic specification of GPA of other skills (e.g. if Math regression, I consider the GPA of Science, Basque, Spanish and English), dummy for female, set of indicators for whether the student is first- or second-generation immigrant (or from unknown origin), month of birth (normalized value, 0 if December and 1 if January), income-groups that define scholarship eligibility, special-needs indicator and grade retention dummy. **School level controls:** Cubic specification on the school ISEC index and on the school-level average of GPA test score in other skills. **Robust standard errors. Clustered at the school level.**

In Figure 1.3, I test whether Basque home-speakers obtain better outcomes than other students from the Basque model. The specification replicates that employed in Figure 1.2. Here, I focus on 4^{th} grade primary students attending the Basque option. Rather unsurprisingly, children display a comparative advantage in the language spoken at home. In particular, I find that Basque speakers have higher scores in Basque language tests (0.31σ, $p = .000$), but perform significantly worse in Spanish (0.29σ, $p = .000$). With regards to the remaining skills, they obtain better results in Math (0.09σ, $p = .088$), but these are lower in English (0.11σ, $p = .015$). The difference in Science is almost zero (0.03σ, $p = .582$). The findings suggest that there are mixed differences based on the language being spoken at home. Replicating the analysis for middle school students does not substantially alter the results. Overall, one cannot safely conclude that this constitutes a significant channel affecting student sorting in Bilbao.

Two relevant features suggest that the above findings need a cautious interpretation. First, the allocation of students may reflect optimal choices. It is possible that Spanish-speaking students from model D are precisely those who are less affected by studying in the Basque option. However, this does not mean that other pupils would not suffer from larger difficulties. Thus, the results may underestimate the size of the Basque-speaking premium. Second, the observed differences increase in effect size and statistical significance when the analysis exploits the whole Basque sample. In conclusion, it is possible that the language spoken at home is a significant determinant in shaping the observed preferences for the Basque-monolingual model in other areas of the region.

1.8 Limitations and Caveats

The findings of this study present some limitations. First, because schooling zone maps are only available for institutions located in Bilbao, the sample selection and households' choice sets suffer from some constraints. The availability of the catchment area boundaries is an important ingredient to infer the residence-based points that determine applicants' admissions. Unfortunately, these data are not readily available for every Basque school. This feature imply two relevant limitations. On the one hand, I need to exclude families that apply to schools located outside Bilbao from the analytical sample. As a consequence, the results focus on households that specially value school proximity. Therefore, my findings possibly understate the relevance of linguistic choices and academic quality for the general population. On the other hand, parents are assumed to construct rankings by solely considering schools from Bilbao. However, there are no geographical restrictions to the schools households can apply to. Thus, the feasible school choice set that families contemplate might be larger that the one being considered.

A second limitation is that parents might not use the ISEI-IVEI evaluations to measure program quality. One convenient aspect of these tests is their standardized nature, which alleviates concerns of biased results from distinct assessment criteria between schools. However, these data are not available to the general public, and they are not easily observable. Another related drawback is that I have no data available about students' own ability. Hence, the findings abstract from potential ability matching concerns. Yet, Abdulkadiroğlu *et al.* (2020)

find little evidence in support of school selection based on matching, and thus they probably do not constitute a highly relevant consideration. Fourth, residential locations are treated as exogenous. Nevertheless, parents might adopt residential choices in response to schooling decisions. The exogeneity assumption is a widely spread assumption in the literature. In this sense, the inclusion of residential sorting in the determination of preferences is a worthwhile avenue for future research.

A fifth limitation is that the model assumes that every household acts strategically. However, there may exist families that simply submit rankings in order of their true preferences. Agarwal & Somaini (2018) suggest that the proportion of sincere applicants is negatively associated with the share of applicants that obtain their first option. Given that 92.3% of households are admitted to their top-ranked alternative, the fraction of truth-telling parents is probably small.[29] Hence, this is not likely to be a significant concern. Nevertheless, by means of a lab experiment, Chen & Kesten (2017) show that the proportion of truth-telling agents is larger in the PM than in the BM. Thus, extending the model to allow for heterogeneous agent sophistication may be a direction worth pursuing. Recent empirical papers have developed models to account for this feature [Agarwal & Somaini (2018), Calsamiglia *et al.* (2020)].

Tightly linked with the previous point, there exists a sixth and final limitation. Implicit in the i.i.d structure of household private-taste shocks lies the Independence of Irrelevant Alternatives (IIA) assumption. Consequently, parental tastes for schools display constrained substitution patterns. As previously mentioned, a more flexible model that allows for distinct agent sophistication is considered for the future. Nonetheless, the incorporation of this component in the model precisely requires the i.i.d assumption of errors for identification [Calsamiglia *et al.* (2020)].

1.9 Concluding Remarks

Does identity affect schooling choices? Focusing on the Basque Country, this chapter studies how much academic quality families trade off for different language models. This is an

[29] According to Calsamiglia *et al.* (2020), 93% of students are assigned the first option in Barcelona. They estimate that over 96% of parents are strategic.

interesting question given that the choice of a vehicular language is tightly associated with the coexistence of distinct cultural identities in multilingual communities.

Here, I propose the use of a structural model to investigate parental preferences using school application data from Bilbao, the largest city in the region. To control for the manipulable nature of the PM assignment algorithm, I model household behavior as reflecting optimal choices from admission probabilities. The empirical results indicate that, on average, parents are willing to concede a significant amount of instructional quality to attend the Basque monolingual option. The estimated preferences suggest an increasing preference for schools' quality and socioeconomic composition with higher income.

One important limitation is that there are no data available about the identity affiliations from applicants. Therefore, the results do not capture identity-matching effects and need to be interpreted as an *average preference* in the sample. To study the mechanisms driving the observed heterogeneous choices, I complement the structural estimation with regression analysis. Several findings emerge from this examination. First, I show that a significant and robust association exists between nationalistic voting patterns and language choices at the census unit level. Given the connection between Basque nationalism and the regard for the local culture, this suggests that schooling decisions are affected by identity associations. Second, I find that there are significant differences in the academic results between linguistic models. Therefore, the presence of heterogeneous preferences for different subjects might also affect student sorting. Third, I observe that children display some learning differences based on the language spoken at home. Overall, this underlines the notion that the results cannot be solely attributed to identity considerations.

There are several worthwhile directions for insightful research in this topic. Among others, it seems particularly interesting to allow heterogeneous preferences based on applicants' own identity. In this sense, combining survey and administrative data would be particularly valuable.[30]

Understanding school demand is crucially relevant for equitable education policy making. The reason is that what parents value significantly shapes the incentives that school face under choice-based competition. However, the literature thus far has not emphasized the role of

[30] Kapor *et al.* (2020) is a remarkable example of research that combines survey and administrative data. By using surveys to school applicants, they introduce subjective beliefs in the estimation of a school choice model.

parents' regard for identity in schooling decisions. In light of the central role of identity in the current political arena, further work focusing on the role of cultural considerations in school choice provides numerous open avenues for fruitful research.

1.10 Appendix

1.10.1 Priority Criteria

As regulated by the Decree 35/2008 from the Department of Education, together with the Order from December 10 2015, the following points are awarded to applicants based on their characteristics:

- Annual income of the family unit (cumulative items up to the maximum score). **Maximum score of this section: 3 points.**

  - If the general tax base of the Income Statement for 2014 does not exceed 42,000 euros: **1.5 points.**

  - If the general tax base of the 2014 Income Statement exceeds this amount: **0 points.**

  - In addition, in the above cases, **0.25 points** will be added for each son or daughter under age other than that of the applicant.

- Proximity of address (not cumulative concepts). **Maximum score of this section: 5 points.**

  - Address of the student in the area of influence of the requested center: **5 points.**

  - Address of the student in the area of influence bordering on that of the requested center: **2 points.**

  - Address of the student in the municipality where the center is located, but outside the areas of influence and bordering: **1 point.**

  - Workplace of the father, mother, legal guardian of the student in the area of influence of the requested center: **2 points** (this score is incompatible with that given by the fact that the father, mother, guardian or legal guardian of the student work in the requested center).

- Existence of family members who study or work in the center (not cumulative concepts). **Maximum score of this section: 9 points.**

- One or more sisters or brothers enrolled in the requested center or in an attached center: **9 points.**

- The father, mother, guardian or legal guardian works in the requested center: **7 points** (this score is incompatible with that granted by the fact that the father, mother, guardian or legal guardian of the student work in the area of influence of the requested center).

- Other criteria (cumulative concepts).

  - For belonging to a large family of general category: **1 point.**

  - For belonging to a large family of special category: **1.5 points.**

  - Due to disability (maximum score for this concept: 2 points): For the applicant **2 points**, for the father, mother, legal tutor or any sibling **1 point**

- By condition of cooperative member or partner of the requested center of any of the members of the family unit: **1 point.**

- Criteria freely included by the School Board or Maximum Representation Body from the center, which may also be some or some of those expressed above, established in accordance with public, objective and non-discriminatory criteria by reason of birth, race, sex, religion, opinion or any other personal or social condition or circumstance: **up to 2 points** (the criteria required to be made public and communicated to the corresponding Delegate or Territorial Delegate of Education before the beginning.

- In case of a tie, it will be resolved by attending the highest score obtained in the previous criteria, comparing them one by one and in the order indicated below:

  - Higher score obtained in the sisters section or siblings enrolled in the center or parent, mother, guardian or legal guardian working on it.

  - Higher score obtained in the proximity section of the student's address or of the position of work of the applicant himself or his father, mother, legal guardian.

- Higher score obtained in the subsection of concurrence of a disability in the student or student or in his father, mother, legal guardian or guardian or in some brother or sister.

- Higher score obtained in the subsection of Large family.

- Higher score obtained in the section of the Annual income of the family unit.

- Higher score obtained in the subsection of status of cooperative partner of the Center.

- Random tie-breaker.

1.10.2 Empirical Evidence of Strategic Behavior

Here, I present some suggestive evidence of household strategic behavior via regression analysis. To do so, I rely on a similar test to that used by Calsamiglia *et al.* (2020). In particular, using household-program pair observations, I run the following OLS specification:

$$y_{ij}^{r} = \beta I(l_i \in z_j) + \delta d_{ij} + \delta_2 d_{ij}^2 + \delta_3 d_{ij}^3 + \gamma fam_{ij} + \Theta X_i + \mu_j + \epsilon_{ij}, \tag{1.7}$$

where y_{ij}^{r} is applicant i's top-listed program from r^{th} ranked school, d_{ij} is the distance to the school of program j, fam_{ij} is a dummy for whether student i has a parent working or a sibling studying in the school of program j, X_i is a vector of household characteristics, μ_j are program fixed effects and ϵ_{ij} is an error term. Standard errors are clustered at the applicant level. The coefficient of interest is β. The variable $I(l_i \in z_j) = 1$ if family i lives inside the catchment area of program j.

Having the residence located inside this zone awards 5 points in the application to the top-ranked school. Typically, this is sufficient for granting admission to the program. Catchment areas are defined to favor allocations based on proximity. Intuitively, households prefer schools that are closer. However, one would not expect the ranking behavior to change discontinuously if families do not act strategically and preferences are continuous in distance.

Table 1.5 summarizes the findings. To study the sensitivity of the results, I experiment with the inclusion of applicants' characteristics and program fixed effects. Columns 1 through 4 validate the robustness of the results. I observe an increase of 5.15 pp in the top choice probability if the program is inside the catchment zone ($p = .000$). I further replicate the analysis by focusing on applicants with varying ranking lengths. Columns 5 and 6 show that the association remains in size and statistical significance. In contrast, columns 7 through 9 show a substantial lower jump in the probability that a program is ranked as second or third option. This is reasonable given that residence-based points are deducted in the allocation to second and third choices. The size of this associations are similar to those observed in Agarwal & Somaini (2018).[31]

[31] Further, I perform two additional robustness checks not reported here. First, the results do not change when we consider a smaller set of programs for each household (25 selected at random). Second, the qualitative nature of findings remain when we leave out the higher ranked options in the probability jumps for the first and second choices.

Table 1.5 Strategic Behavior: Regression Analysis

DEPENDENT	Top-Ranked						Ranked 2^{nd}		Ranked 3^{rd}
VARIABLE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-zone program	0.0510***	0.0516***	0.0509***	0.0515***	0.0591***	0.0595***	0.0329***	0.0344***	0.0303***
	(0.00160)	(0.00165)	(0.00160)	(0.00165)	(0.00224)	(0.00285)	(0.00275)	(0.00347)	(0.00339)
Distance	-0.131***	-0.132***	-0.131***	-0.132***	-0.129***	-0.119***	-0.110***	-0.103***	-0.0758***
	(0.00415)	(0.00403)	(0.00416)	(0.00404)	(0.00563)	(0.00699)	(0.00545)	(0.00686)	(0.00620)
Distance^2	0.0450***	0.0449***	0.0450***	0.0449***	0.0441***	0.0405***	0.0352***	0.0330***	0.0226***
	(0.00152)	(0.00148)	(0.00152)	(0.00148)	(0.00208)	(0.00255)	(0.00197)	(0.00246)	(0.00222)
Distance^3	-0.00458***	-0.00457***	-0.00458***	-0.00457***	-0.00452***	-0.00414***	-0.00347***	-0.00324***	-0.00213***
	(0.000164)	(0.000161)	(0.000164)	(0.000161)	(0.000227)	(0.000275)	(0.000210)	(0.000261)	(0.000234)
N	125,528	125,528	125,528	125,528	63,308	39,848	63,308	39,848	39,848
R^2	0.082	0.093	0.082	0.093	0.107	0.107	0.066	0.065	0.048
Rank length	All	All	All	All	{2,3}	3	{2,3}	3	3
Applicant Charact.	✗	✗	✓	✓	✓	✓	✓	✓	✓
Program FE	✗	✓	✗	✓	✓	✓	✓	✓	✓

Notes: Reports from OLS regressions using household-program pair observations. Robust standard errors. Clustered at the applicant level. Distance is measured in kilometers. Controls include: presence of family member in the school, dummy for large family, special needs and income tercile. I do not report the controls' coefficients in the interest of saving space. Rank length describes the number of schools listed in the application ranking.
*$p < .1$,**$p < .05$,***$p < .01$.

1.10.3 Empirical Methods

Imputation of Priority Points

Applicants get some priority points based on whether they are partners of the school (1 point) and on whether they fulfill some discretionary considerations that each school sets (up to 2 points). The data available only describe the realization of these criteria for the top-ranked school. To deal with this limitation, I rely on certain assumptions to impute the points that households would acquire for other schools if they were top ranked.

- For the partnership dimension, I assume that parents belong to at most one school and that they list it as their top-ranked option. This seems a reasonable assumption. Being a school partner is associated with having a shared history with that institution. Thus, parents are unlikely to belong to several schools simultaneously and they probably rank their partner school first. Additionally, the weight of this criterion is small (1 point compared to the residence-based 5 points) and its incidence is limited in the sample (only 6.4% of households get membership points).

- For the discretionary criterion, I impute the mode of the distribution (0 points) to families' applications for programs other than their top-ranked option. This simplifying assumption, although imperfect, seems like a natural decision. However, 55% of household get some positive amount of points in this dimension. To account for this distortion, it would be interesting to evaluate the robustness of the results by using some other more complex imputation rules.

Pseudo-Code for Admission Probabilities Simulation

Since assignments to over-demanded programs depend on applicant characteristics and lottery draws in case of ties, I simulate admission probabilities to every program j as follows:

1. I compute the amount of seats assigned using the tie-breaking rule in each program ($seats_j$) and the number of applicants that applied to j in choice-band $\bar{b_j}$ with $\bar{s}_j$. I denote the latter as tie_j.

2. I fix a t such that $\Phi(t) = \bar{s}_j$ and assign a lottery number.

3. I create 1,000 copies of every priority-type with $\Phi(t) = \bar{s}_j$ and assign to each copy a random lottery number.

4. I take $(tie_j - 1)$ from these copies at random.
5. Using the tie-breaking rule, I record $seats_j$ assignments among the $(tie_j - 1)$ selected copies and the fixed priority type.
6. I repeat the process 500 times and integrate over the simulations.
7. I repeat steps (2) through (5) for every t with $\Phi(t) = \bar{s}_j$ and every program j.

Pseudo-Code for Simulated Maximum Likelihood

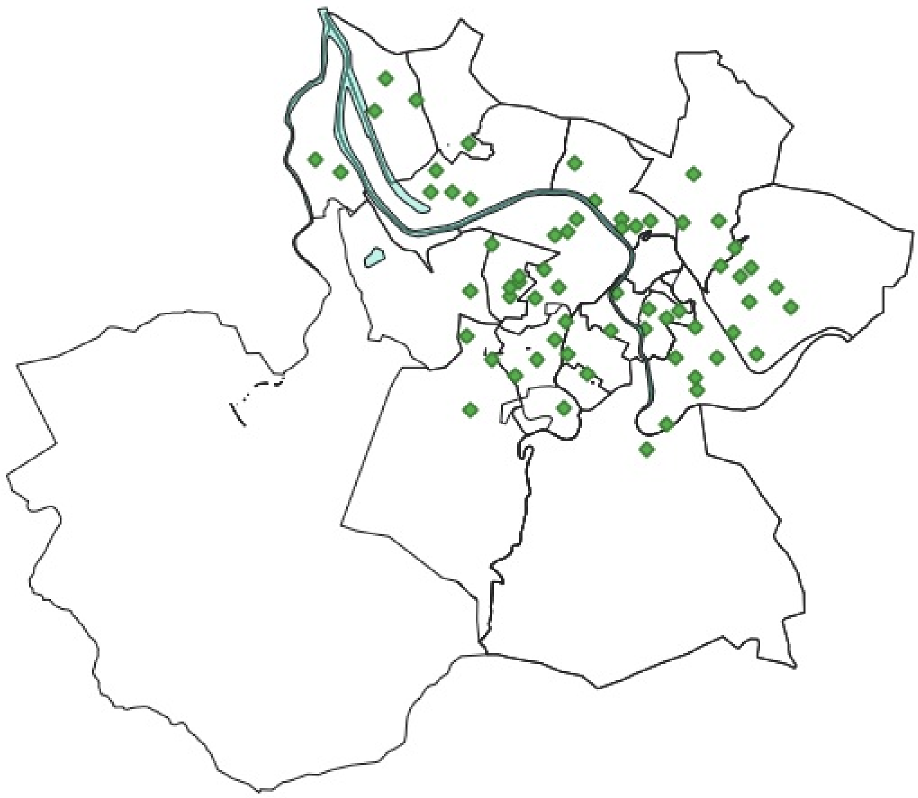
The simulated maximum likelihood is solved by applying the Nelder-Mead (N-M) algorithm. In particular, I employ the particular algorithm version proposed by Press *et al.* (1997). To numerically approximate (1.6), I adapt the A-R simulator proposed by Train (2009). Specifically, the A-R proceeds as follows:

1. For each household i, draw a J dimensional vector of errors $\epsilon_i = (\epsilon_{i1}, ... \epsilon_{iJ})$ from a $\epsilon_{ij} \sim N(0, 1)$. These are left unchanged for every iteration of the N-M.
2. Using these errors and parameter values θ, construct A_i^* via backward induction.
3. Determine whether $\tilde{A}_i \in \Lambda^*(c_i, l_i, \epsilon_i; \theta)$. If so, record auxiliary variable $I^q = 1$ and 0 otherwise.
4. Repeat the previous steps Q times, where $Q = 500$.
5. The simulation of the likelihood is given by $\hat{L}_i(\theta) = \frac{1}{Q} \sum_{q=1}^{Q} I^q$.

One drawback of the A-R is the non-smoothness of the likelihood with respect to the parameters. This creates situations under which the optimization algorithm gets stuck in a false maximum. To ensure that the N-M converges to the desired solution, I re-run the algorithm by modifying and restarting the resulting simplex until no further improvement of the log-likelihood is achieved.

1.10.4 Data Appendix

Figure 1.4 School Catchment Areas (Source: Bilbao Data Lab)



Notes: Green dots represent the different schools. The catchment zone boundaries are in superposition with each other. Individual maps for each school area available at Bilbao Data Lab's Git Hub **- (Click Here)**.

Figure 1.5 Average 2016 Household Income, by Census Unit (Source: Atlas of Household Income Distribution, from INE)

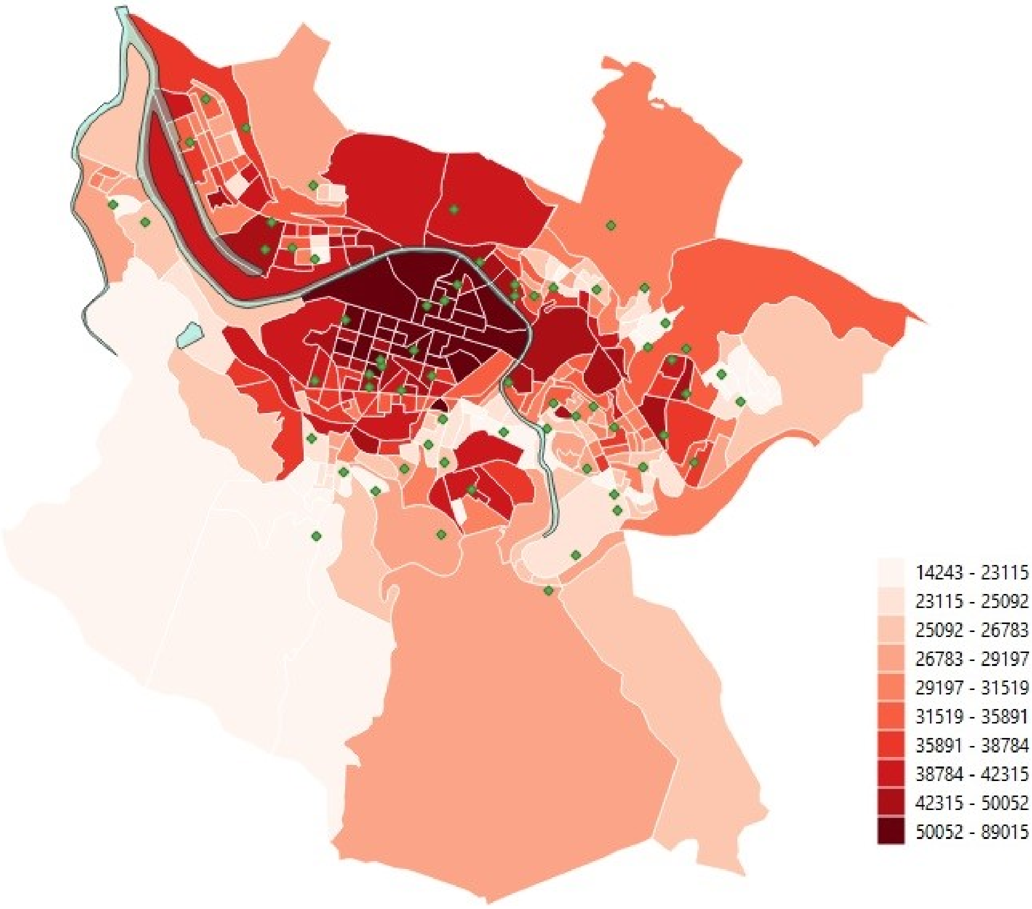
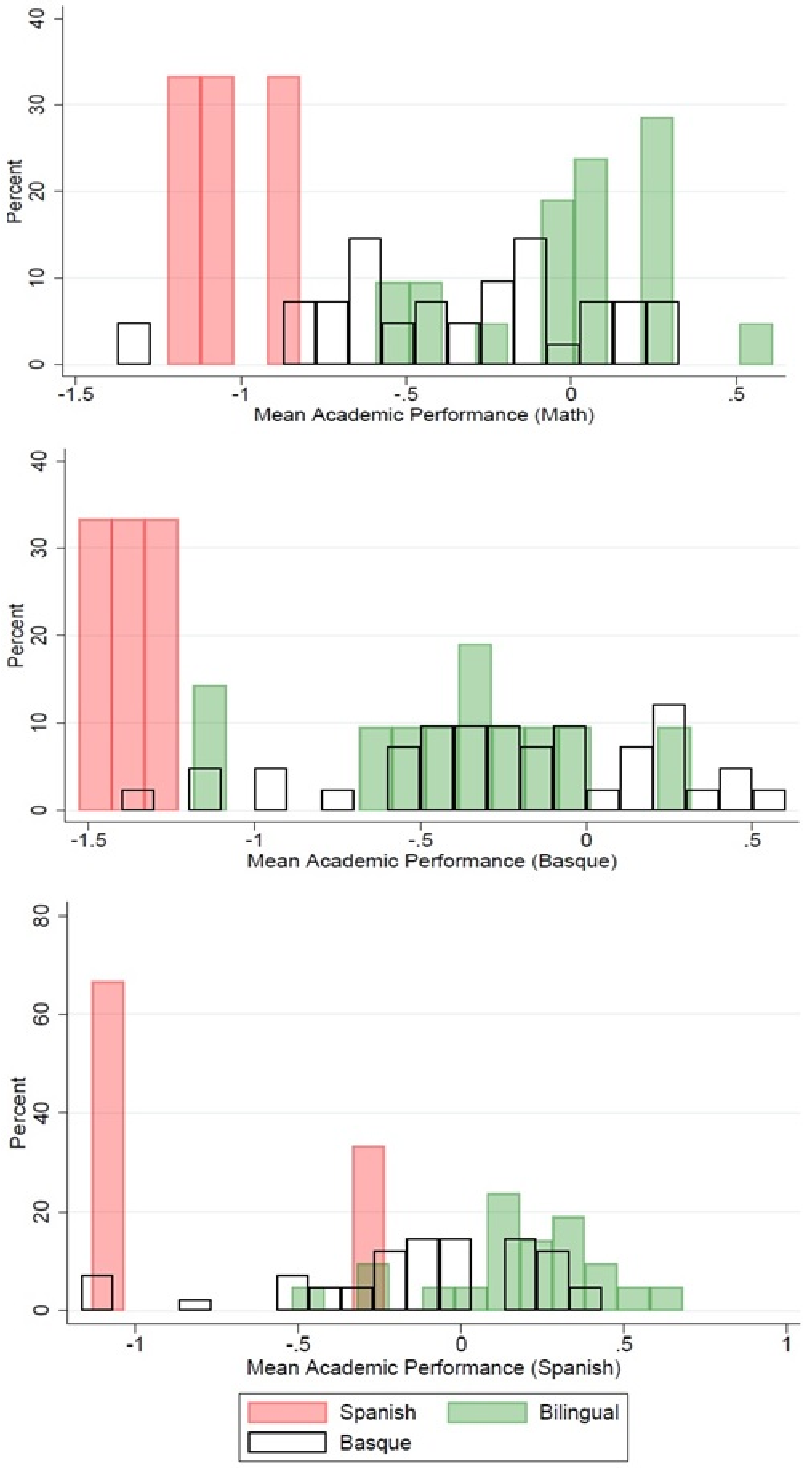


Table 1.6 Data Description

VARIABLE	DESCRIPTION
Distance	Euclidean distance from student i to program j, calculated with QGIS.
Academic Quality	Average of the 2011-2017 student-level standardized score of ISEI-IVEI exams from 4th primary grade students from program j. GPA of Math, Basque and Spanish. Standardization using the regional mean and standard deviation for students in their academic year.
ISEC index	Average Socio Economic and Cultural index of the school. It is calculated based on a student questionnaire by ISEI-IVEI. It includes the professional and maximun level of studies from parents, number of books at home and other possession of other cultural goods (e.g. PC and Internet).
Income Tercile	Terciles computed using the 2016 Atlas of Household Income Distribution from the National Institute of Statistics. It employs estimated household average income at the census unit level.
% foreign-born students	Percentage of foreign-born students in primary education in school j in 2016, using administrative data.
School Surface	Size of school amenities as provided by administrative data. Measured in 1,000 square meters.
In-zone program	An indicator if student i lives in school j's catchment area.

Figure 1.6 Subject-specific Academic Quality Distributions, by Language Models



Notes: This figure plots the histograms of the program quality measures, separately for each subject. This metric uses the program-specific students' GPA scores in Math, Basque and Spanish in primary education. **Sample Size by Language Models:** Spanish-monolingual = 3, Bilingual = 21, Basque-Monolingual = 42. Three programs from semi-public schools have missing information of their performance; one in each language model.

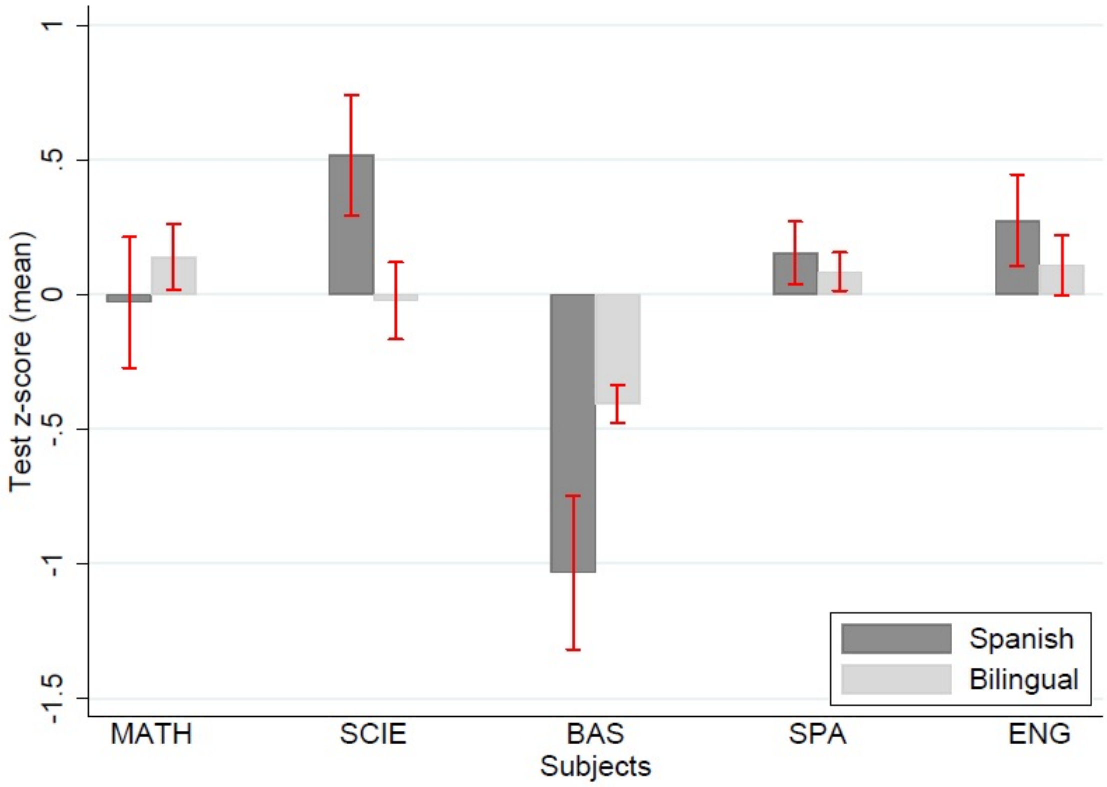
Table 1.7 Electoral Results and Linguistic Choices (individual-level regressions)

DEPENDENT	=1 applicants rank Model A or B				=1 if top-rank Model A or B			
VARIABLE	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
% Vote Share Non-Nationlist	0.0146***	0.0108***			0.0154***	0.0135***		
	(0.00235)	(0.00244)			(0.00194)	(0.00192)		
% Vote Share Non-Nationlist (+ Podemos)			0.00445	0.00724***			0.00872***	0.0115***
			(0.00309)	(0.00267)			(0.00241)	(0.00218)
N	1,829	1,829	1,829	1,829	1,829	1,829	1,829	1,829
Applicant Characteristics	✗	✓	✗	✓	✗	✓	✗	✓
R^2	0.039	0.090	0.002	0.077	0.050	0.080	0.010	0.063

Notes: Robust standard errors in parenthesis. Clustered at the census unit level. Controls include: Income terciles, dummuy for presence of a family member in a school, indicator for large family and special needs, minimum distance to a school that supplies a program under Model A or B and number of schools that supply model A or B in a radius of 750 meters. The coefficients on these covariates are not reported in the interest of saving space. The 2018 definition of census units is used. Due to some minor mismatching, 26 applicants are left out of the regression sample ($N = 1,829$).

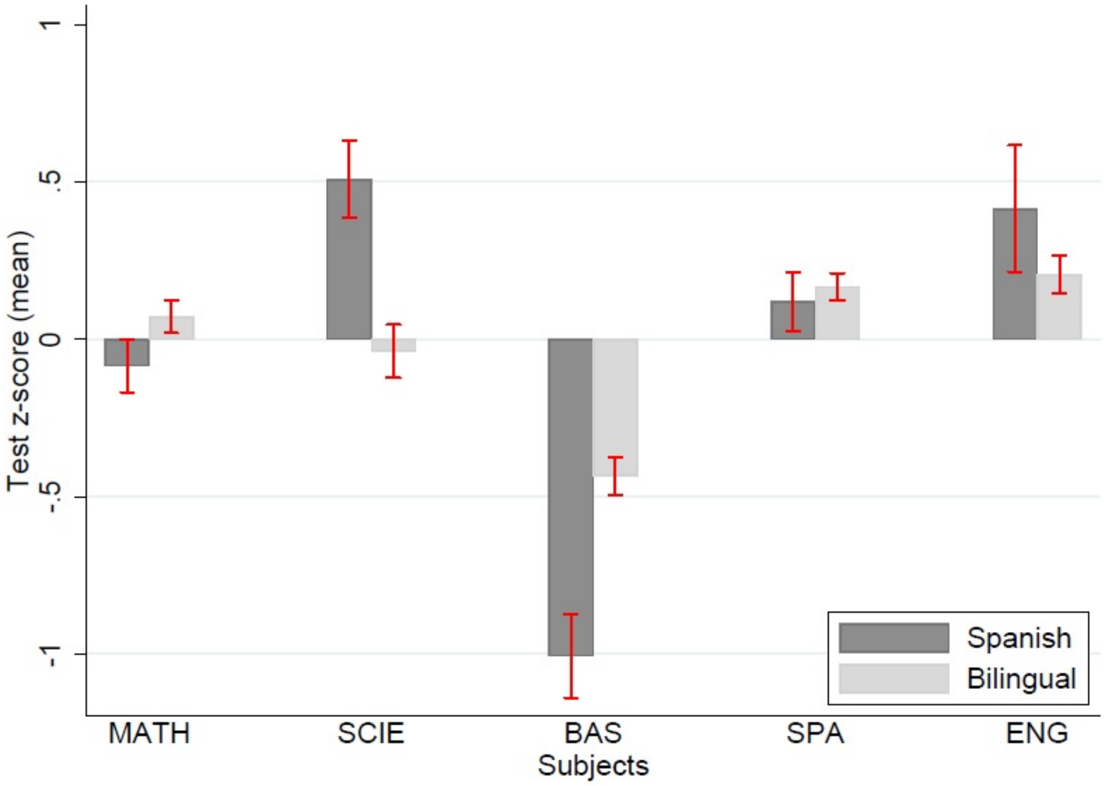
*$p < .1$,**$p < .05$,***$p < .01$.

Figure 1.7 Differences in Subject-Specific Sills, relative to the Basque-monolingual Model (Middle School)



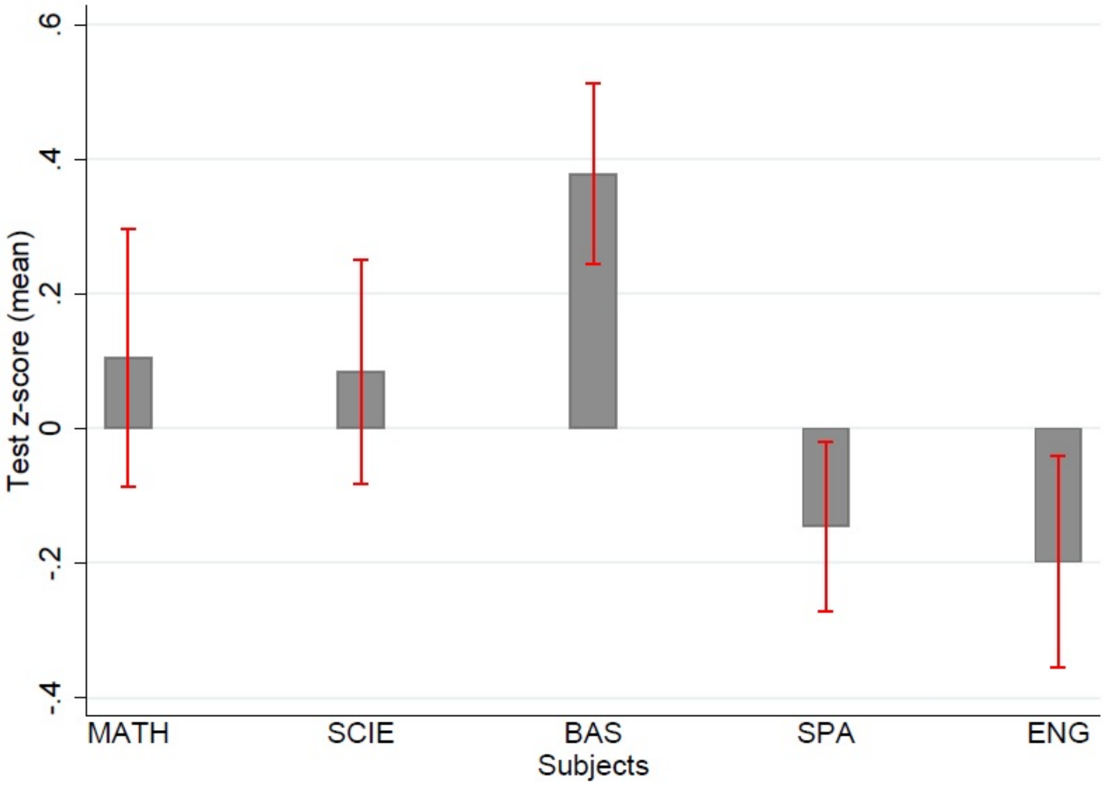
Notes: OLS estimates (and confidence intervals) for the association between Model A and B and the z-score in ISEI-IVEI assessments, by subject. Results are relative to the Basque-monolingual model (omitted category). **Sample:** Students of 13-14 years old living in Bilbao ($N = 1,995$). **Sample sizes by language models:** Spanish - 138, Bilingual - 868 and Basque - 989. **Individual Controls:** cubic specification of GPA of other skills (e.g. if Math regression, I consider the GPA of Science, Basque, Spanish and English), dummy for female, set of indicators for whether the student is first- or second-generation immigrant (or from unknown origin), month of birth (normalized value, 0 if December and 1 if January), income-groups that define scholarship eligibility, dummy for whether the student is Basque home-speaker, special-needs indicator and grade retention dummy. **School level controls:** Cubic specification on the school ISEC index and on the school-level average of GPA test score in other skills. **Robust standard errors. Clustered at the school level.**

Figure 1.8 Differences in Subject-Specific Sills, relative to the Basque-monolingual Model (4^{th} grade of primary, Full Regional Sample)



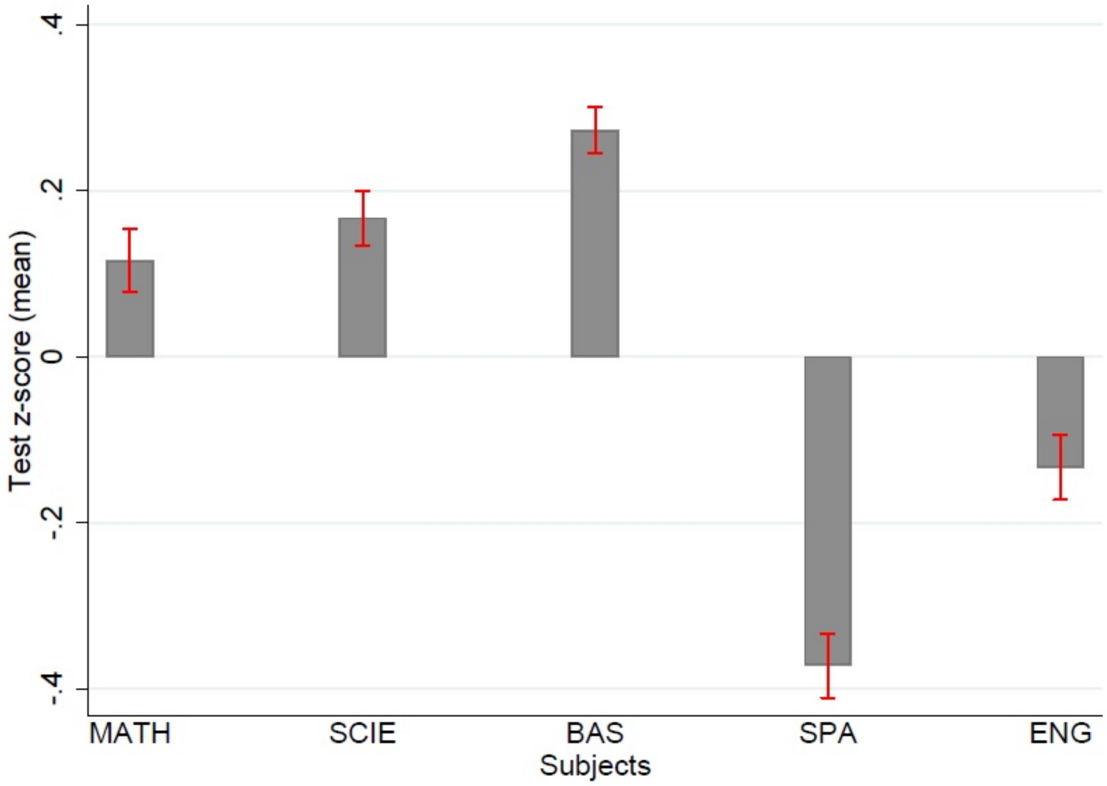
Notes: OLS estimates (and confidence intervals) for the association between Model A and B and the z-score in ISEI-IVEI assessments, by subject. Results are relative to the Basque-monolingual model (omitted category). **Sample:** Students of 9-10 years old living $(N = 15,802)$. **Sample sizes by language models:** Spanish - 655, Bilingual - 3,594 and Basque - 11,553. **Individual Controls:** cubic specification of GPA of other skills (e.g. if Math regression, I consider the GPA of Science, Basque, Spanish and English), dummy for female, set of indicators for whether the student is first- or second-generation immigrant (or from unknown origin), month of birth (normalized value, 0 if December and 1 if January), income-groups that define scholarship eligibility, dummy for whether the student is Basque home-speaker, special-needs indicator and grade retention dummy. **School level controls:** Cubic specification on the school ISEC index and on the school-level average of GPA test score in other skills. **Robust standard errors. Clustered at the school level.**

Figure 1.9 Basque-Speaking Premium in the Basque-monolingual model (Middle School)



Notes: OLS estimates (and confidence intervals) for the association between being a Basque home-speaker and the z-score in ISEI-IVEI assessments, by subject. Results are relative to not speaking Basque at home (omitted category). **Sample:** Students of 13-14 years old living in Bilbao and and studying in the Basque model ($N = 989$). **Basque Home Speakers:** $N = 95$. **Individual Controls:** cubic specification of GPA of other skills (e.g. if Math regression, I consider the GPA of Science, Basque, Spanish and English), dummy for female, set of indicators for whether the student is first- or second-generation immigrant (or from unknown origin), month of birth (normalized value, 0 if December and 1 if January), income-groups that define scholarship eligibility, special-needs indicator and grade retention dummy. **School level controls:** Cubic specification on the school ISEC index and on the school-level average of GPA test score in other skills. **Robust standard errors. Clustered at the school level.**

Figure 1.10 Basque-Speaking Premium in the Basque-monolingual model (4^{th} grade of primary, Full Regional Sample)



Notes: OLS estimates (and confidence intervals) for the association between being a Basque home-speaker and the z-score in ISEI-IVEI assessments, by subject. Results are relative to not speaking Basque at home (omitted category). **Sample:** Students of 9-10 years old studying in the Basque model ($N = 11,553$). **Basque Home Speakers:** $N = 4,087$. **Individual Controls:** cubic specification of GPA of other skills (e.g. if Math regression, I consider the GPA of Science, Basque, Spanish and English), dummy for female, set of indicators for whether the student is first- or second-generation immigrant (or from unknown origin), month of birth (normalized value, 0 if December and 1 if January), income-groups that define scholarship eligibility, special-needs indicator and grade retention dummy. **School level controls:** Cubic specification on the school ISEC index and on the school-level average of GPA test score in other skills. **Robust standard errors. Clustered at the school level.**

2 Cultural Assimilation and Ethnic Discrimination: An Audit Study with Schools

2.1 Introduction

Over the last decade, discussions over immigrant assimilation have become increasingly heated in political arenas across the Western world. On the one hand, nativist politicians argue that rising multiculturalism poses a threat to the preservation of autochthonous culture and hence ensuring immigrant assimilation is essential. On the other hand, some mainstream democratic governments, such as that of Denmark, have made assimilation mandatory to grant immigrants full access to social benefits. While most liberal democracies agree to concede rights and protection to immigrants, whether assimilation should constitute a duty for accessing public services is an increasingly debated question. Yet little is known about the role that assimilation plays in immigrants' access to important services, like health or education. Do discriminatory attitudes of natives affect immigrants' access to these services? Can newcomers mitigate entry barriers to welfare services by signalling assimilation effort? What incentives drive such biases?

Using a field experiment, this chapter addresses these questions by focusing on access to early- and compulsory-education. Education is arguably the most crucial means of easing the socioeconomic integration of first- and second-generation immigrants. First, it constitutes a fundamental engine of human capital formation and upward mobility (Chetty *et al.* (2011, 2017), Card *et al.* (2018)). Second, it acts as a powerful tool to inculcate shared civic values and to enhance social cohesion (Billings *et al.* (2014)). In this study, I assess the presence of discriminatory attitudes towards immigrant families in the acquisition of school-related information, and evaluate the potential of cultural assimilation to mitigate such barriers. Before the school registration period takes place, families engage in an active process of gathering information in order to decide upon a school for their children. By selectively providing information, schools can influence their applicant pool and “cream skim" students by discouraging certain incoming families from enrolling their children. Although declining to provide information and denying access to a given service is not equivalent, previous evidence suggests that the former is causally linked to participation decisions (Hastings & Weinstein (2008), Hoxby & Turner (2013)). Ultimately, this behavior can result in worsened educational outcomes for minority students through limited access to high-quality schools, thus harming their later labor market success. Hence, investigating unequal treatment in the provision of information helps to reveal underlying discriminatory attitudes from schools that influence immigrant students’ long-term labor market outcomes (Acemoglu & Angrist (2001)).

To quantify the size of these discriminatory frictions and study the effect of immigrants’ cultural assimilation efforts on discrimination, I emailed more than 2,500 kindergartens and compulsory schools located in the Community of Madrid in the period leading up to the 2018 student registration. In these messages, three types of fictitious couples, one Spanish and two Romanian, requested to visit the school.[1] The respective Romanian couples signalled different levels of cultural assimilation effort in the origin of their child’s name. Furthermore, similar to other studies, I signed the emails using common Spanish- or immigrant-sounding names to evoke the different ethnic origins of the couples. Whether cultural assimilation significantly mitigates discriminatory attitudes is evaluated through study of the presence of distinctive response patterns from the schools. To enhance the comparability across family profiles, I fix a number of socioeconomic characteristics that drive the selection into assimilation and affect discrimination. These characteristics include paternal occupation and working sector, the language proficiency of the couple, and the *age-adequacy* of the child.

The Spanish context provides a particularly suitable context for analyzing the aforementioned questions. From 1998-2008, the foreign-born share of the population increased by 10 percentage points. This constituted one of the most remarkable immigration periods experienced in recent history by an OECD country and had considerable social consequences. With specific regard to the education system, immigration led to significant native flight towards private schools and increased student-teacher ratios in public schools (Farré *et al.* (2015)).

[1] Romanians constitute the most numerous immigrant group in Madrid (21.7% of the foreign-born population) and are culturally relatively similar to Spaniards compared to other immigrant groups.

Moreover, this occurred after public education had been transferred from the central to regional governments, the size of the private school network had increased, and parents' freedom of choice favored (Arellano & Zamarro (2007)).

A focus on immigrant families' name choices as a proxy of assimilation effort provides a number of advantages relative to other widespread measures. First, studying the implications of cultural assimilation by looking at *actual* assimilation outcomes, like intermarriages or labor-market assimilation, can produce misleading conclusions. The reason being that such measures are equilibrium outcomes that depend on the interplay of both natives' and immigrants' actions and are therefore constrained by the stance of the native population (Fouka *et al.* (2018)). Second, other proxies of assimilation effort, such as language acquisition or naturalization, might be affected by financial restrictions, labor market prospects, the composition of immigrants' social networks, etc. They hence can fail to disentangle assimilation effort from other factors associated with socioeconomic success since they may be disturbed, ultimately, by discriminatory attitudes. Name-choices, on the other hand, are unconstrained by such barriers, act as crucial markers of individual identity, and carry significant cultural content (Algan *et al.* (2013)). At the social level, names are relevant descriptors of a family's cultural background and trigger preliminary judgments about ethnic group affiliation. Moreover, names reflect a personal choice that embodies a migrant family's trade-off between transmitting their own cultural identity and adapting to the traditions of their new environment.

Several findings emerge from this analysis. First, I find that immigrants who signal low efforts towards assimilation face important discriminatory barriers in the acquisition of information (12% lower response rate relative to a baseline of a 77% response rate for natives). Second, I show that child's name origin significantly reduces discrimination by 50%. The emails from the Romanian couple who chose a Spanish name for their child are significantly more likely to get a response than those from the couple who selected a Romanian name (73% as opposed to 68% response rate). I perform a number of checks (i.e., considering the qualitative tone of the response and employing a duration analysis of the delay in response) that support the robustness of my findings.

To uncover the main mechanisms driving the results, I assess whether a diverse set of predictors explain the differences in response patterns. In particular, I consider both the characteristics of the schools and their surrounding geographical areas. At the school level, I examine attributes that are associated with higher tuition expenditures or with learning difficulties for immigrants. This set of characteristics includes the type of school (i.e., whether it is public, charter, or private), the age of the child, the instruction language, and the number of extra services provided by the school. At the community level, I account for political preferences, the incidence of Romanian immigration, and the size and housing prices of the neighborhood or municipality where the school is located. From a statistical standpoint, I find no significant heterogeneous effects in any of the considered dimensions. In terms of magnitude, however, I find that response patterns display substantial variability based on these characteristics. The lack of significant differential responses suggests that the findings may suffer from limited statistical power to detect heterogeneity. The evidence is however consistent with school level incentives (e.g., the age of the child being served) and community characteristics (e.g., relative presence of Romanian migrants) mediating observed discrimination.

In this particular setting, the case for statistical discrimination cannot be entirely ruled out, in spite of being weaker than in audit studies focusing on the labor market. Thus, results may not be entirely attributed to prejudice or taste-based discrimination. In absence of direct evidence for the role of taste-based *vis-à-vis* statistical discrimination, some patterns suggest that the differential responses can be, at least partially, associated with the former. For instance, the observation that there is a larger gap between assimilated and non-assimilated families in kindergartens is consistent with perceived place of birth not being the main driver of the results. However, I cannot exclude the possibility that statistical discrimination plays a significant role in explaining the gap between assimilated and non-assimilated families. Despite this limitation, the findings are nevertheless relevant from a regulatory perspective. Even if school staff use names as proxies for parental background, the differential treatment based on ethnically distinct names is likely to be seen by courts as a pretext for discrimination (Kline & Walters (2020)).[2]

This chapter is related to several strands of the literature. First, it speaks to the ample body of work on immigrant assimilation. An increasing number of studies have leveraged the informativeness of migrant parents' name choices as a proxy of cultural assimilation effort. However, most focus on understanding the labor market penalty associated with foreign-sounding names using observational data (for instance, Arai & Skogman Thoursie (2009),

[2]According to the national law, publicly-funded schools need to provide families with equally favorable treatment, no matter their race, religion, place of birth, opinion or other personal circumstance.

Goldstein & Stecklov (2016) and Abramitzky *et al.* (2016)). My study differs in two ways. First, it evaluates the influence of assimilation efforts on discriminatory attitudes in access to education services. Second, it concentrates on the implications of assimilation at early-life stages as opposed to evaluating outcomes observed in adulthood.

My analysis also contributes to the literature on school choice. Existing scholarship has provided ambivalent evidence regarding the impact of the latter on student stratification (Böhlmark *et al.* (2016), Söderström & Uusitalo (2010), Zimmer & Guarino (2013), Gortazar *et al.* (2020)). The presence of heterogeneous preferences based on family socioeconomic characteristics has been a widely studied motive that rationalizes these findings (Burgess *et al.* (2015), Hastings *et al.* (2009), Beuermann *et al.* (2019)). Yet little research has been conducted on the role of school "cream skimming" through the generating of informational frictions. This is largely due to the fact that observational studies do not allow to disentangle the two. In response to such limitations, a few studies have adopted an experimental approach to understand the active role of schools in dissuading minority students. These include, among others, Diaz-Serrano & Meix-Llop (2016) for homosexual parents in Catalonia (Spain), Pfaff *et al.* (2018) for religious minorities, and Bergman & McFarlin Jr (2020) for children with poor behavior, learning disabilities, and low achievement in the US.[3] To my knowledge, however, there has been no comprehensive study that quantifies specifically the role of stereotypical attitudes towards immigrants and the impact of assimilation on schools' attitudes towards migrants.

The remainder of the chapter is organized as follows. I begin in Section 2.2 by introducing the institutional setting. Section 2.3 presents the experimental design and the data. Section 2.4 shows the main results. Section 2.5 and 2.6 provide the heterogeneity analysis and the robustness checks. Section 2.7 discusses the limitations of the experiment and the interpretation of results. Finally, Section 2.8 concludes.

[3] Giulietti *et al.* (2019) study the existence of racial discrimination in information provision by school districts in the US. They also evaluate access to other services like public libraries or sheriff offices.

2.2 Institutional Setting

The final assignment of children to a given Spanish school can be understood as the outcome of a three-stage process in which families, public authorities, and schools participate. First, parents evaluate the performance of schools by acquiring information about their characteristics. Second, they identify the strategy that optimizes the probability of being admitted to their preferred school in light of the institutional rules. Finally, children are allocated as a result of the strategies implemented by families, the preferences of the school population, and the legal framework that regulates the admission processes. To uncover the presence of discrimination and the effects of assimilation efforts, I focus on the first phase of the process by examining schools' relative eagerness to provide information to parents.

Most Spanish schools offer the possibility of visits through the scheduling of personal appointments or the organization of open doors. Such occasions grant families an excellent opportunity to understand the school's educational project, assess the quality of the facilities, or interact first-hand with teachers and school coordinators. Hence, neglecting this moment could crucially affect parents' ultimate decision to solicit enrollment in a particular school. By rejecting visit requests, schools can signal nonconformity with parent choice and affect their likelihood of attempting to join the establishment. Therefore, evaluating the presence of distinctive response rates to such requests can effectively measure discriminatory attitudes towards immigrants.

My sample covers public, charter, and private schools, which are subject to different frameworks in their admission and funding procedures. This has relevant implications for using responses to visit requests as a measure of discrimination, since these regulatory constraints arguably affect schools' ability and incentives to cream skim.

With regards to admissions, access to private schools depends on discretionary criteria established by the school and is independent of public regulation. In contrast, publicly-funded charter and public schools are subject to a centralized assignment mechanism. In particular, children are allocated to publicly-funded schools based on the Boston Mechanism (BM) and a series of legally imposed priority criterion that depend on student and family characteristics. This system implies two relevant consequences for the experimental setting.

First, it leaves these schools with very limited margins to decide who is being admitted.[4] Overall, this means that publicly-funded schools have less freedom to cream skim compared to private schools. Hence, denials of visit requests comprise one of their few strategies to shape student composition, once tuition fees are set. Second, the BM is vulnerable to manipulation and thus significantly penalizes parents that are not well informed about the process (Abdulkadiroğlu *et al.* (2006)). Therefore, informational frictions are likely to result in more inequitable assignments in comparison to settings that employ truthful *strategy-proof* mechanisms (Abdulkadiroğlu & Sönmez (2003)).

With regards to their funding sources; public, charter and private schools display some relevant institutional differences. While private schools are entirely financed through private contributions, public and charter schools are government-funded to a varying degree. In contrast to public schools, that depend solely on public funds; government resources only cover a fraction of charter schools' expenses (i.e., staff salaries and other specific indirect costs, like non-ICT equipment and academic support programs). As a result, charter schools call upon households for quasi-compulsory private donations since subsidies only account for 60% of total per pupil cost (Calsamiglia *et al.* (2020)).

In sum, private schools display conflicting motivations to more intensely discriminate against immigrants. On the one hand, they have a wider variety of screening opportunities and mechanisms to limit certain students' access at their disposal. On the other hand, they need to finance their operating expenses, including support programs (that are disproportionately attended by immigrants), through the exclusive use of private sources. This arrangement may motivate them to respond differently. Thus, the inclusion of private schools can provide interesting comparative observations.

One important point of contention is the impact of immigrant enrollment on perceived school quality. On average, immigrant students display higher drop out rates and attain lower grades. Hence, one might argue that schools may try to deter immigrant admission to protect their school performance measurements. Until 2016, Madrid was the only region making the average results of schools' external standardized assessments available to the public and thus

[4] Schools have the possibility of conceding a discretionary extra point to students that fulfill conditions deemed relevant. An application can submitted with up to 22.5 points. Thus, the weight of this decentralized point is small (4.44% of the maximum total points an application can have). In 2017/18, 93.4% of families were granted their first schooling option, according to the Office of Communication of the Community of Madrid.

allowing comparisons of performance levels among schools (Anghel *et al.* (2015)).[5] Similarly, the average performance indicators from the University Entrance Exam (*Prueba de Acceso a la Universidad*, PAU) up until 2017-18 were publicly available. This information is however no longer accessible to the general public. As a result, schools did not have clear incentives to differently respond to an inquiry from immigrant families based on the monitoring of students' skills through their perceived observable impact on performance measures and school rankings at the time of the experiment.

In terms of composition, the school segregation by socioeconomic level in the Community of Madrid is the highest among all Spanish autonomous regions and almost all countries in the entire EU (Murillo & Martínez-Garrido (2018)).[6] In contrast, student stratification by national origin is low compared to the national and the EU average (Murillo *et al.* (2017)). Altogether, this implies that unequal access to more favorable schools by minority students may be mostly driven primarily by their social, economic and cultural capital differences rather than by their ethnic origin.

2.3 Experimental Design and Data

2.3.1 Experimental Design

The experimental sample consists of schools that were listed in the Madrid School Search Engine (hereafter, MSSE)[7] and that offered either pre-primary or compulsory education, not differentiated by gender. I eliminated hospital classrooms, private kindergartens for government or company employees, and schools that have an arts- or sports-specific curriculum. Schools

[5] These tests were introduced in 2004/05 under the name *Prueba de Conocimientos y Destrezas Habituales* or "Indispensable Knowledge and Skills Tests". These exams had a compulsory nature were compulsory for every primary and middle school. These low-stakes tests initially assessed students at the end of primary. However, since 2007/08 they also evaluated students from the 3^{rd} grade of middle school.

[6] Using data from PISA-2015, they find that the segregation level in Madrid is only surpassed by Hungary within the European Union in secondary education.

[7] MSSE is a database created by the Department of Education of Madrid. It employs administrative data and is available to the general public. Its primary aim is to provide school-level information to parents who are in the process of choosing a school for their children.

whose email addresses were not available were also discarded.[8] For schools sharing the same email address, I randomly selected one to participate in the experiment.

As a result of the above criteria, the sample consists of 2,584 schools. Some messages were not delivered because the associated email inboxes were full, while others replied that the education grade associated with the age of the child was not supplied in that school.[9] After excluding these schools from the analysis, the final sample consists of 2,551 schools.

I designed three fictitious family profiles: two Romanian immigrant couples and a control profile of Spanish natives. To evoke ethnic minority status, I selected names and surnames based on name-frequency data available from the National Institute of Statistics (INE). More specifically, I randomly selected the identities of the family members using the list of most popular gender-specific names and surnames held by persons of Spanish and Romanian nationality living in Spain. To avoid the proliferation of different treatments, I only chose one gender for the child, which was randomly selected to be male.[10] This decision, together with that of focusing on a single nationality, was motivated by statistical power concerns. In what follows, I denote a family with Spanish names as *natives*, a family with Romanian names as *non-assimilated immigrants* and a family whose parents' names are Romanian but their child's name is Spanish as *assimilated immigrants*, for the sake of brevity.

I focus on Romanian immigrants for three reasons. First, they constitute the largest fraction of immigrants in the region by a substantial margin. As of January 2018, Romanians represented 21.7% of immigrants, followed by Moroccans (8.9%) and Chinese (6.8%). Second, my experimental design poses some limitations for the study of discriminatory behavior against Latinos since Latin Americans and Spaniards use similar names. Finally, Romanians show relatively higher integration levels than other ethnic groups.[11] From an economic standpoint, Romanian newcomers display education levels that are roughly at Spanish levels (de la Rica

[8]I obtained email addresses from *Educateca*, a private initiative offering information about the education sector in Spain since 1999. A preliminary analysis suggested that the email addresses available from *Educateca* tend to better match those on the schools' websites compared to the email addresses available through the MSSE.

[9]The latter responses correspond to six schools that have an unconventional educational supply in that they do not provide instruction for an entire education cycle. Their inclusion in the analysis is inconsequential for the results.

[10]For each family profile, I choose two male names (associated with the father and the son), one female name (associated with the mother) and one surname (pertaining to the father) at random. The name of the son of the assimilated immigrant family was identical to that of native family's son.

[11]From an anecdotal perspective, Romanian immigration has been praised as an example of integration in Spain and Madrid by government authorities (see here) and in media articles (here).

& Ortega (2012)), and have favorable legal status since the 2007 EU enlargement, which facilitates their labour market integration relative to other non-EU citizens. Furthermore, they use a Latin-based language and are a Christian (Eastern Orthodox) population. Thus, Romanians display higher cultural similarity along genetic, religious and ethnolinguistic lines than other large ethnic groups traditionally present in Madrid (with the exception of Latinos and some other European nationals).[12]

These distinct features have implications for the representativeness of the experimental results for other minority groups. Previous work indicates that natives display heterogeneous acceptance across cultural-ethnic groups based on the perceived cultural distance *vis-à-vis* the native population, with more distant groups being increasingly penalized (Fouka *et al.* (2018), Bisin & Tura (2019), Adda *et al.* (2020)). At the same time, efforts towards adaptation might provide greater gains for more culturally distant migrant groups since they lower the perceived cultural gap to a greater extent. Thus, the results of this study arguably provide a lower-bound estimate of discrimination and the impact of assimilation for more culturally dissimilar non-EU immigrants (e.g., Moroccan or Chinese)

Distinctive name origins might capture differences in attributes other than ethnicity, such as the socioeconomic status of the family. Disparities in relevant unobservables can lead to forms of statistical discrimination that worsen the identification of taste-based discrimination. To mitigate concerns about biased results from socioeconomic discrimination, I sent all inquiries from three family-specific email accounts in a "$malename.surname@reformas-surname.net$" format.[13] This choice aimed at signalling that inquiries were sent from the fictitious fathers using their professional email accounts and eliciting that paternal occupations were identical across the treatment groups, without rendering the email suspicious. In particular, the email accounts suggested that each father was the owner of a small-to-medium construction firm. Hence, the socioeconomic status of the family was one of the first pieces of information that schools could infer together with the family's ethnicity.

[12] Appendix Figure 2.1 summarizes the cultural distance of several cultural-ethnic groups relative to Spain. For this purpose, I use genetic, linguistic and religious distance measures from Spolaore & Wacziarg (2016). Overall, it shows that, among the largest immigrant groups present in Madrid, Romanians are culturally the closest, only surpassed by Latinos, Italians ad Polish.

[13] *"Reformas"* is the Spanish word for renovations. The Spanish Labor Force Survey (EPA) indicates that a significant proportion of both male Romanian immigrants and natives work in the construction sector. In addition, anecdotal evidence suggests that it is common for SMEs to include the owner's surname in the name of the company.

I sent one single email to each school. This decision allowed to minimize both the risk of increasing suspicion about the credibility of the emails and the associated time losses that the experiment could entail, without undermining the validity of the experimental design. In the spirit of Ewens *et al.* (2014), each email text contained (1) an introductory *hello* statement, (2) a *presentation* statement in which the names of the three family members and the age of the child were presented, (3) a *statement of interest* that showed the couple's interest in the school, (4) an *inquiry statement* asking for the school's availability to meet with the couple and (5) a *closing* statement that thanked the email recipient and was followed first by the name of the father and then that of the mother. Appendix Table 2.5 provides a sample email template and the list of identities for the different family members.

I set the age of the child to be appropriate for the grade needed to begin the next academic year. This allowed to alleviate discrimination from grade retention or late entry rates across ethnic profiles that could bias the results. I targeted the first grade of the lowest education cycle of each school.[14] Competition among parents for seats in these grades is likely to be higher since applications are disproportionately concentrated at these stages.[15] It could consequently be argued that they constitute the set of grades where there is a higher probability of encountering discrimination.

The large bulk of mailing required for this experiment implies a high probability of being identified as a spammer. The triggering of spam filters is determined, among others, by the sender's prior mailing frequency and volume. At the same time, service providers impose superior daily sending limits to users with higher past activity. Thus, to avoid potential delivery issues, I gradually increased the number of emails sent to other accounts that were created for these preparatory purposes in the weeks prior to the experiment. In all cases, the messages were successfully delivered.

[14] I treat kindergartens and other schools (hereafter: non-kindergartens) differently. I define the former as schools that solely provide education to children between the ages of 0-3. The experimental sample has a high number of such schools (37.5%). I exclude the possibility of other schools that have superior levels receiving an email mentioning a 1 year-old child. This allows for a better balance in the amount of observations between kindergartens and other schools, as well as grants higher statistical power in the heterogeneity analysis when this characteristic is studied.

[15] Parents typically enroll their children at the beginning of each educational stage. The majority of applications thus concentrate at the beginning of the second-cycle of pre-primary, followed by the first grade of primary and of middle school (Gortazar *et al.* (2020)).

For the implementation of the experiment, I automatized the mail dispatch with a bulk-emailing software from February 26 to March 2, 2018. I waited for schools to respond until the start of the registration period, or the date when responses are arguably no longer useful to parents. I did not follow up in the case of a non-response and tried to politely decline invitations.

2.3.2 Measuring Responses

I look at whether schools systematically showed different response patterns before the pre-registration period began. For the empirical results, I mainly focus on a response indicator as the outcome variable, which identifies whether or not the school answered the email.

To evaluate the robustness of the results, I further categorized responses into positive or negative by considering the qualitative tone of the answer. Simple differences in the probability of receiving a response may result in misleading conclusions if one treatment group is more likely to receive explicit negative or less cordial answers. Thus, negative responses include (1) non-answered emails, (2) emails that declined to grant an invitation and (3) emails from schools that did not explicitly decline a visit but lacked direct contact information or availability of dates.[16] Finally, to consider the intensive margin of the replies I measure the number of days a school took to answer the email and evaluate the existence of discrimination through a duration analysis.

2.3.3 Additional Covariates

I use administrative data from several sources in order to show that covariates are balanced across treatment groups and evaluate the presence of heterogeneous treatment effects. Data on school characteristics is collected by *webscrapping* the MSSE.[17] I also assembled information on the extra services provided by each school (i.e., lunch, transport, or *extended day* services), the type of school (i.e. public, charter or private), and whether it offers a bilingual education.

[16]In their answers some schools implied that they did not have seats available. To deal with the ambiguity of these answers, I categorized them into being positive if they additionally grant an invitation in an explicit manner. Emails missing contact information likely imply that the respondent is discouraging the visit by sounding uninterested.

[17]I requested additional administrative data on the socioeconomic composition and average performance of schools on standardized tests from the Education Department of the Regional Government of Madrid. Due to privacy considerations, however, they kindly declined to share this information.

In addition, I gathered data on the schedule upon which the school operates (i.e., continuous or split shift)[18] and the education levels at which they teach (i.e., first or second cycle of pre-primary, primary, or middle school).

To study the relevance of community characteristics, I compiled information on electoral outcomes from the 2016 general elections, the size of the municipality, and the housing prices and the share immigrant population of the area where the school is located. For establishments located in the city of Madrid, I use electoral and real estate data at the neighborhood level. For those outside the capital, I used municipality-level data. The use of neighborhood level information for schools located in the city aims at increasing the statistical power and reliability of the heterogeneity analysis.

I matched each school located in Madrid with its corresponding neighborhood by using GIS data.[19] Neighborhood-level data on electoral and real-estate outcomes come from the Madrid City Council. Covariates on the size of the municipality and electoral outcomes for municipalities different than Madrid city come from Madrid's Municipal and Zonal Data Bank, commonly known as *Almudena*. Data on the presence of immigrants with Romanian citizenship are from the Municipal Registry (*Padrón Municipal*). Additional details on the variables can be found in Appendix Table 2.6.

Table 2.1 reports the baseline mean statistics of school characteristics across treatment groups. All groups were similar relative to the observed school characteristics under the assignment, indicating that randomization was successful. There are no statistically significant differences across the three treatment groups in the proportions of the child's age signalled in the emails, the types, or the bilingual nature of the schools, their provision of extra-services, the kind of shift, the size of the municipality, the housing prices, the political preferences of the school area, and the majoritarian presence of Romanians (relative to other migrants) in the area. There is a marginally significant difference ($p < 0.1$) in the share of Romanian migrants of the surrounding area between the native and the assimilated family-profiles. However, this difference, which can be attributed to random sampling error, is quantitatively small (0.3 p.p.).

[18]Schools that function under a continuous shift have more concentrated class hours such that children can leave earlier. On the contrary, schools that use a split shift have longer breaks and continue their classes after lunch, finishing later in the afternoon.

[19]I used *Open Cage Geocoding* for retrieving GPS coordinates from school addresses. Then, I matched coordinates with neighborhoods using virtual cartographic maps from the Madrid City Council.

Table 2.1 Balancing Test

SCHOOL CHARACTERISTICS	Nat	AI	NI	t-test (p-value)		
	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(2)-(3)
Age signalled = 1	0.379	0.372	0.376	0.765	0.906	0.857
	(0.485)	(0.484)	(0.485)			
Age signalled = 2	0.488	0.498	0.508	0.665	0.399	0.682
	(0.500)	(0.500)	(0.500)			
Age signalled = 5	0.012	0.010	0.007	0.818	0.314	0.436
	(0.107)	(0.102)	(0.083)			
Age signalled = 11	0.122	0.120	0.109	0.883	0.402	0.490
	(0.327)	(0.325)	(0.312)			
Public school	0.573	0.575	0.555	0.922	0.450	0.394
	(0.495)	(0.495)	(0.497)			
Charter school	0.158	0.177	0.177	0.302	0.278	0.959
	(0.365)	(0.381)	(0.382)			
Private school	0.269	0.249	0.268	0.322	0.945	0.357
	(0.444)	(0.432)	(0.443)			
Bilingual school	0.252	0.271	0.270	0.380	0.388	0.988
	(0.434)	(0.445)	(0.444)			
Continuous shift	0.423	0.413	0.398	0.696	0.295	0.511
	(0.494)	(0.493)	(0.490)			
Number of extra-services	1.702	1.715	1.706	0.735	0.903	0.827
	(0.853)	(0.858)	(0.845)			
Very Small Muni.	0.105	0.099	0.099	0.690	0.684	0.994
	(0.306)	(0.298)	(0.298)			
Small Muni.	0.135	0.137	0.146	0.888	0.494	0.587
	(0.342)	(0.344)	(0.353)			
Medium Muni.	0.128	0.125	0.122	0.885	0.709	0.819
	(0.334)	(0.331)	(0.327)			
Big Muni. (Madrid city excluded)	0.243	0.244	0.235	0.955	0.725	0.683
	(0.429)	(0.430)	(0.425)			
Madrid city	0.390	0.395	0.398	0.844	0.745	0.898
	(0.488)	(0.489)	(0.490)			
Left-wing area	0.278	0.266	0.263	0.588	0.506	0.902
	(0.448)	(0.442)	(0.441)			
Housing: 1,000€ per sq meter	2.816	2.818	2.799	0.978	0.842	0.818
	(1.127)	(1.108)	(1.085)			
Housing: Tertile = 1	0.399	0.388	0.391	0.779	0.829	0.948
	(0.490)	(0.488)	(0.489)			
Housing: Tertile = 2	0.315	0.324	0.329	0.823	0.698	0.869
	(0.465)	(0.469)	(0.471)			
Housing: Tertile = 3	0.286	0.288	0.280	0.942	0.866	0.809
	(0.452)	(0.454)	(0.450)			
Immigration: % Romanian population	3.248	2.977	3.117	0.089*	0.432	0.376
	(3.450)	(3.127)	(3.417)			
Immigration: Romanian majoritarian	0.730	0.719	0.708	0.622	0.325	0.625
	(0.444)	(0.450)	(0.455)			
N	861	861	862			

Notes: Nat, AI and NI stand, respectively, for natives, *assimilated*, and *non-assimilated* immigrants. Standard deviations in parentheses. p-value refers to the t-test for the difference in means for each pairwise comparison across treatment groups: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Information on housing prices is only available for schools located in Madrid city ($N_{Nat} = 336$, $N_{AI} = 340$ and $N_{NI} = 343$). Information on immigrant shares exclude schools (24) from three neighborhoods due to data incompatibility of neighborhood definitions across sources. Data sources of these variables are presented in Table 2.6 in Appendix.

Differences in attrition across groups would raise concerns about the validity of the results. If the non-delivery of emails to certain accounts affected the balance on schools characteristics, distinct response behaviors across treatment groups might be driven by differences in average characteristics. Appendix Table 2.8 shows that observables are balanced after the small attrition of certain schools.

2.4 Main Results

Overall, 72.6% of the inquiries were answered before the pre-registration period began. This large response rate reinforces the validity of the findings in three ways. First, it confirms that these requests are relevant to schools and thus constitute an informative metric for evaluating the presence of discrimination. Second, it indicates that the experiment was credible. Finally, it shows that the delivery of emails was generally successful and that they were not systematically identified as spams.

Next, I examine the existence of differences in response rates and positive response rates to assimilated and non-assimilated immigrants in a regression framework. In particular, I estimate the following linear probability model:

$$R_i = \alpha + \beta_1 AI_i + \beta_2 NI_i + X_i'\delta + d + g + \epsilon_i, \tag{2.1}$$

where R_i is the outcome for school *i* (a dummy variable indicating a response or a positive response to the inquiry email). AI_i and NI_i are binary variables on whether the email was sent, respectively, from the assimilated or non-assimilated immigrant profile. Here, X_i is a vector of characteristics associated with the school and its surroundings. This set of dummy variables include the school type (i.e., whether it is a charter or a private school), the supplied extra services, the age of the child mentioned in the email, an indicator on whether the school is located in a left-wing area, the type of shift, and the municipality size. Finally, d are dummies on the calendar days the emails were sent and g represents fixed effects for the geographical areas where the schools are located.

Table 2.2 reports the main results. Panel A and Panel B summarize, respectively, the difference in response rates and positive response rates. In columns (1) and (6), I show the

Table 2.2 Difference in Response Rates and Positive Response Rate

	Panel A: Response					*Panel B:* Positive Response				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Assimilated Immigrant	-0.044**	-0.044**	-0.044**	-0.047**	-0.047**	-0.049**	-0.049**	-0.049**	-0.048**	-0.047**
	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.021)	(0.023)	(0.022)
Non-Assimilated Immigrant	-0.090***	-0.089***	-0.091***	-0.093***	-0.094***	-0.092***	-0.091***	-0.093***	-0.096***	-0.097***
	(0.022)	(0.021)	(0.021)	(0.023)	(0.022)	(0.022)	(0.022)	(0.021)	(0.023)	(0.023)
N	2,551	2,551	2,551	2,551	2,551	2,551	2,551	2,551	2,551	2,551
Native group mean	0.771	0.771	0.771	0.771	0.771	0.750	0.750	0.750	0.750	0.750
Test on $\beta_1 = \beta_2$ (p reported)	0.039	0.042	0.029	0.046	0.038	0.056	0.061	0.045	0.040	0.033
R^2	0.007	0.012	0.067	0.123	0.175	0.007	0.011	0.064	0.134	0.184
Delivery day F.E.	✗	✓	✗	✗	✓	✗	✓	✗	✗	✓
Applicant/School Characteristics	✗	✗	✓	✗	✓	✗	✗	✓	✗	✓
Geographic Area F.E.	✗	✗	✗	✓	✓	✗	✗	✗	✓	✓

Notes: The outcome variables of the regressions are: (i) an indicator on whether the school responded to the inquiry email (Panel A), and (ii) an indicator on whether a positive response to the email was provided (Panel B). Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Applicant/School Characteristics include: indicators on the school-type (i.e., charter and private), a set of dummies on the extra services provided, a set of dummies on the age mentioned in the email, and a dummy for continuous shift. In the interest of saving space, I do not report the coefficients on the control variables.

results of the basic model, which only contains two dummy variables on whether the school received the email from the non-assimilated or assimilated immigrant family as regressors. Columns (2) and (7) expand the model by adding delivery day fixed effects. Columns (3) and (8) instead include the aforementioned control variables. Alternatively, in columns (4) and (9), I add geographical area fixed effects to control for unobserved heterogeneity in neighborhoods and municipalities. Finally, columns (5) and (10) contain at once both sets of fixed effects and the covariate vector. Rather unsurprisingly, the inclusion of additional controls does not significantly change the size of the estimated gaps due to random assignment. All the models considered here use robust standard errors. This decision follows the findings from Abadie *et al.* (2017). First, the sample essentially includes the whole population of schools in Madrid. Second, the experiment follows a complete random assignment design. Thus, there is no clustering in the sampling and in the treatment assignment that justifies the use of clustered standard errors.

Panel A reveals that visit requests from Spanish names have a 77% probability of being answered. Analogous emails with Romanian names have a 68% possibility of getting a response. This represents a 9 percentage point gap in response rates, which translates into non-assimilated immigrants being answered 12% less of the time. The difference is statistically significant ($p < 0.01$) and robust to every specification. This finding suggests that schools show, on average, systematic differences in tastes and beliefs about Romanian immigrants.

In contrast, inquiries from assimilated immigrants have a 73% chance of receiving an answer. This result allows to draw two interesting conclusions. First, it shows that children's name-origin choices on the part of immigrant families are valued by schools and significantly affect discriminatory attitudes in the provision of school information. I find a 5 percentage point response gap between assimilated and non-assimilated immigrant families. The associated 7% difference is statistically significant ($p < 0.05$) and robust. Second, it implies that name-choices are not, however, sufficient to entirely eliminate ethnic discrimination. Assimilated immigrants are 4 percentage points (or equivalently, 6%) less likely to get a response compared to natives. The observed gap is also statistically significant ($p < 0.05$). Taken together, this suggests that name-choice mitigate discrimination by 50%.

Considering a response as equivalent to an invitation may result in inaccurate conclusions about the presence of discrimination. Because schools might display distinct propensities for explicitly declining visit requests based on the treatment, Panel B summarizes the differences in the quality of the answers by examining positive responses. Comparing the baseline response rate and positive response rate for natives in columns (1) and (6) suggests that only 2.1% of their responses are negative. Similarly, the observed gaps barely adjust when contemplating the qualitative tone of the answers, for both assimilated (4.4 vs 4.9 p.p.) and non-assimilated immigrants (9 vs 9.2 p.p.). Altogether, these small differences suggest that schools do not typically reply with a rejection, but instead articulate discriminatory attitudes through their decisions to respond. It thus seems that schools rely on subtle passive forms of discrimination instead of using active explicit declines to visit requests. With the exception of a small decline of significance in the difference between assimilated and non-assimilated immigrants, the nature and robustness of the results from positive responses are notably similar to those from response rates.[20]

2.5 Heterogeneity Analysis

Thus far, the above results, which focus only on the average differences across family profiles, may hide some level of heterogeneity. In particular, the extent of response gaps may vary depending on the features of the school or the characteristics of the area in which they are located. Thus, assessing the presence of heterogeneous responses across these features can help to better understand the main mechanisms.

To explore this heterogeneity, I rely on two different approaches. First, I test whether schools with distinct characteristics answer at different rates using various Difference-in-Differences (DiD). Table 2.3 show the results for response rates (Panel A) and positive responses (Panel B). More specifically, entries come from separate LPM regressions of a response indicator on an assimilated and a non-assimilated immigrant dummy, the characteristic listed, and the two interactions of that given characteristic with the immigrant dummies. Each entry of columns (1) and (3) represents the marginal effect of a particular school or community

[20]I perform two additional checks concerning the functional form specification and the quality of responses. First, I use a probit model instead of a LPM (see Appendix Table 2.9). Results are similar to the estimates from the baseline specification. Second, I employ an alternative response-quality measure by coding the predisposition for personal visits among respondents, excluding invitations to open doors (see Appendix Table 2.10). Results are marginally significant ($p < 0.1$) after the inclusion of controls. All point estimates are negative for assimilated and non-assimilated profiles. The coefficient for the former is marginally superior to that for non-assimilated families.

characteristic in the assimilated immigrant versus native response gap. Column (2) and (4) describe the analogous marginal effects for non-assimilated immigrants' response gap. The second approach towards exploring heterogeneity is to independently estimate response gaps by running separate regressions on each of the subgroups in the sample. Results are presented in Appendix Table 2.11.

In what follows, I turn to discuss the results, separately for each school characteristic, by making reference to Table 2.3. However, the discussion also incorporates the insights from Appendix Table 2.11. Given the qualitatively similar nature between the two outcome variables, I focus mainly on discussing the findings from response rates.

Heterogeneity based on the age of the child.

Name-origin choices from immigrant families may trigger distinct judgements about the child's country of birth (i.e, on whether he is a first- or second-generation immigrant). Therefore, response gaps may reflect differences in school staff's expectations between Spanish- and Romanian-named migrants that do not involve a significant appraisal of assimilation efforts. If differences in perceived place of birth were the primary driver of the results, one would expect the gap between assimilated and non-assimilated to increase with child's age by reducing the probability that he is born in Spain. The first row of Table 2.3 shows that kindergartens display a lower response rate for non-assimilated families (-5.5 p.p.), but a higher one for assimilated families (1.5 p.p.) compared to primary and middle schools. Accordingly, the gap between Romanian-named and Spanish-named immigrant families is large and economically significant for younger children (7 p.p.). This suggestive pattern tentatively supports the interpretation that results reflect taste-based discrimination. The results are qualitatively similar for positive response. The differences between kindergartens and other schools are statistically non-significant at conventional levels. This suggests that the results lack statistical power to detect significant differences based on the age of the child.

Heterogeneity by type of school.

Peer composition is used as a key informative proxy of school performance (Rothstein (2006), Mizala & Urquiola (2013), Abdulkadiroğlu *et al.* (2020)). Thus, in order to protect their reputation and appeal, more expensive schools may have stronger incentives to discriminate

Table 2.3 Marginal Effects of School and Community Characteristics

CHARACTERISTICS	*Panel A:* Response			*Panel B:* Positive Response			Omitted category
	AI	NI	t-test (p-value)	AI	NI	t-test (p-value)	
	(1)	(2)	(1)-(2)	(3)	(4)	(3)-(4)	
Kindergarten	0.015	-0.055	0.132	0.032	-0.027	0.218	Non-Kindergarten
	(0.045)	(0.046)		(0.046)	(0.046)		
Charter School	0.054	0.011	0.862	0.071	0.022	0.872	Public Schools
	(0.051)	(0.051)		(0.052)	(0.052)		
Private School	0.024	0.034	0.410	0.056	0.046	0.362	
	(0.060)	(0.061)		(0.062)	(0.063)		
Bilingual School	0.002	0.071	0.176	-0.030	0.050	0.136	Non-Bilingual Schools
	(0.048)	(0.049)		(0.051)	(0.051)		
Left-wing area	0.025	0.009	0.747	0.037	0.023	0.790	Right-wing area
	(0.047)	(0.049)		(0.049)	(0.050)		
Small municipality	0.080	0.124	0.633	0.079	0.121	0.656	Very Small municipality
	(0.093)	(0.092)		(0.095)	(0.093)		
Medium municipality	0.040	0.004	0.721	0.010	-0.019	0.777	
	(0.095)	(0.096)		(0.096)	(0.097)		
Big municipality	0.036	0.009	0.764	0.030	-0.024	0.545	
	(0.085)	(0.086)		(0.086)	(0.087)		
Madrid city	0.070	-0.006	0.373	0.045	-0.040	0.324	
	(0.082)	(0.082)		(0.083)	(0.083)		
Housing Prices: Tercile 2	-0.066	0.054	0.153	-0.111	0.019	0.129	Housing Prices: Tercile 1
	(0.078)	(0.081)		(0.080)	(0.082)		
Housing Prices: Tercile 3	0.035	0.099	0.465	0.035	0.068	0.716	
	(0.082)	(0.087)		(0.086)	(0.090)		
Romanian majoritarian	0.057	0.065	0.866	0.056	0.077	0.676	Romanian immigrants non-majoritarian
	(0.048)	(0.049)		(0.049)	(0.050)		

Notes: The outcome variables of the regressions are: (i) an indicator on whether the school responded to the inquiry email (Panel A), and (ii) an indicator on whether a positive response to the email was provided (Panel B). Standard errors in parentheses. $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Entries represent the interaction terms for separate DiD regressions of response on AI, NI, the listed characteristic, and the two interaction terms. Division lines separate the different regressions. p-values report the result of the t-test for both marginal effects. Regression on bilingual schools exclude kindergartens ($N = 1,595$). Regression on housing prices exclude schools outside Madrid ($N = 1,019$). Additional details on the variables can be found in Appendix Table 2.6.

minority students. An indirect way to evaluate whether schools that charge higher tuition behave differently is to look at the type of school. Public, charter, and private schools show systematic differences in their required payments. Public and charter schools, which are publicly-funded by taxpayers, should in theory be free of charge. However, in practice, charter schools charge quasi-compulsory payments that impose *de facto* entry barriers to poorer families.[21] These payments are nevertheless substantially lower than those demanded by private schools.

Rows (2) and (3) indicate that charter and private schools show a larger propensity to respond to immigrant families. In particular, charter schools display an economically significant difference in answering assimilated migrants (5.4 p.p.). Alternatively, private schools show a more favorable response pattern for non-assimilated migrants (3.4 p.p.). This observation is in line with private schools having the opportunity to screen applicants and cream skim more freely at later stages. Thus, they would not need to resort to declining visit requests as stringently. The finding of less discrimination in privately-funded, although surprising, is not statistically significant. The results do not substantially change when considering the qualitative tone of responses.[22]

The language of instruction: bilingual schools.

The government of Madrid implemented a rapidly-expanding bilingual program in publicly-funded schools in 2004/05. Schools operating under this plan must use English as the instruction language at least one-third of the time and teach English as a language 5 sessions per week. Since it could be argued that a bilingual educational approach increases the difficulty of learning content for immigrant children, bilingual schools might display different levels of discrimination. In this direction, previous evidence suggests that this expansion has led to student sorting based on socioeconomic conditions (Anghel *et al.* (2016)).

[21] Such disbursements represented in 2012, on average, 501€ per year. In comparison, the analogous costs for public schools were, on average, 17€ (Farré *et al.* (2015)).

[22] I also test for heterogeneous effects based on the amount of extra services provided by schools (i.e., the sum of lunch, transport, and *extended day* offerings). If these services are costly and schools have some mechanism to enforce their payment, the provision of different services could also serve as an effective entry barrier for families with less economic resources and affect the student socioeconomic composition. The marginal effects (interaction terms) for assimilated (b=0.025, s.e.=0.024) and non-assimilated migrants (b=0.09, s.e.=0.049) are quantitatively small and statistically non-significant.

Row (4) indicates that bilingual schools are more likely to respond to foreign family profiles than non-bilingual schools. In particular, bilingual schools respond substantially more (7 p.p.) to non-assimilated immigrants. Additionally, Romanian-named families receive a slightly larger proportions of response than their Spanish-named immigrant analogs (80.1% vs 82.2%). These results are not statistically significant at any of the conventional levels.

Heterogeneity by electoral outcomes of neighboring area.

Due to parental preferences for proximity, schools typically reflect neighborhood or municipality characteristics. Thus, they may show heterogeneous attitudes towards minorities depending on the political preferences, wealth, or similar features of their associated communities. From a political standpoint, given the conventional stance of right-leaning parties on immigration, conservative areas might have more negative attitudes towards newcomers and a stronger regard for assimilation efforts. Therefore, I explore the role of political preferences of the municipality or neighborhood in which the school is located.[23] For this purpose, I split the sample into geographical areas defined by whether they mostly voted to left- or right-leaning parties. Row (5) of Table 2.3 shows that response gaps for assimilated and non-assimilated migrants do not systematically vary between these subgroups. Treatment effects are not qualitatively different based on the political orientation of the school's neighboring area.

Heterogeneity by municipality size.

Because of distinct attitudes and composition of rural and urban communities, schools may display distinct response patterns based on the size of the municipality they are located in. Rows (6) through (9) from Table 2.3 display the marginal effects by municipality size. The omitted baseline category are very small municipalities. Overall, the results do not imply a systematic relationship between population size and response gaps. Although there is substantial variability in response patterns across the different subgroups, the amount of discrimination does not clearly increase or decrease with population size. Consistent with differential discrimination based on name-origin choices, schools more often respond to assimilated families than to their assimilated analogs, with the exception of those located in

[23]The analysis of political-preferences uses neighborhood (municipal) data if the school is (not) located in the city of Madrid. The study of heterogeneous effects based on housing prices includes only those schools found in Madrid and uses neighborhood-level data.

small municipalities (see Appendix Table 2.11).

Heterogeneity by housing prices.

An alternative way to evaluate the influence of schools' socioeconomic composition is to test for the presence of heterogeneous effects based on the real-estate prices of the area surrounding the school. To this end, I categorize schools according to the tercile of housing prices of their associated neighborhood. Rows (10) and (11) in Table 2.3 display the results, by focusing on schools in the city of Madrid.[24] Overall, I find that schools from costlier neighborhoods more likely respond to both assimilated and non-assimilated migrants, though the differences are not statistically significant. This suggestive evidence is consistent with schools in less wealthy areas being more reluctant to interacting with immigrant families, which may be surprising. Ultimately, the qualitative pattern of non-assimilated families receiving responses less frequently remains. At the same time, assimilated families obtain more favorable outcomes than their non-assimilated counterparts generally, no matter the real estate prices of the area (see Appendix Table 2.11).

Heterogeneity by size of Romanian immigration.

Contact hypothesis states that social interactions between natives and foreign out-groups can reduce prejudice (Allport (1954)). By facilitating learning about minorities, inter-group exposure may be an important element shaping discriminatory attitudes of the majority (Pettigrew & Tropp (2006), Schindler & Westcott (2020)).

To test this hypothesis, I categorize schools depending on whether they are located in areas where Romanians are the most numerous immigrant group and estimate heterogeneous effects. The last row of Table 2.3 displays the results. Schools in areas where Romanians constitute the majoritarian immigrant are more likely to respond to both assimilated (5.7 p.p.) and non-assimilated migrants (6.5 p.p).[25] The findings, although quantitatively large, are not

[24]This part focuses solely on schools located in the city of Madrid since real-estate data was not readily available for other municipalities.

[25]In unreported analysis, I estimate marginal effects using an alternative measure of exposure to Romanian immigration. For this purpose, I split the sample for schools located in areas with above and below median presence of Romanian population. I find more muted results using this alternative measure. Using responses as the outcome variable, the estimated interaction coefficients are 0.001 (s.e.=0.042) for assimilated and 0.021 (s.e.=0.043) for non-assimilated families.

significant at conventional levels, but are robust to considering the qualitative tone of response (see last row of Panel B). Altogether, these suggestive results are consistent with inter-group contact being a relevant moderating factor influencing the presence of discrimination. However, I find that non-assimilated migrants have a significantly lower response rate than natives in both types of areas. At the same time, results indicate that schools are more likely to answer to assimilated families than to their non-assimilated analogs, indistinct of the area. This means that, although to a varying degree, name-origin choices might comprise an important source for triggering heterogeneous attitudes towards foreign-born families.

As mentioned earlier, the findings indicate relatively sizable variability in the response patterns according to the school subgroups. The marginal effects however do not differ significantly at conventional statistical levels. One possibility is that the observed non-significance is driven by limited sample sizes, leading to reduced statistical power to detect significant heterogeneous effects. To evaluate the robustness of results, I pool schools that received emails from assimilated and non-assimilated immigrant profiles and re-run the analysis. The marginal effects of both dependent variables remain statistically non-significant. Ultimately, however, I cannot exclude the possibility that the absence of statistically significant heterogeneous effects along these various dimensions are not driven by statistical power limitations. Notably, Table 2.11 shows that assimilated families generally obtain better outcomes than their non-assimilated analogs, with the exception of schools in small municipalities and bilingual schools.

2.6 Robustness Check: Duration Analysis

Thus far, the analyzed outcomes include the extensive margin of school answers and the qualitative tone of these replies. I now turn to an evaluation of the size of discrimination by assessing the intensive margin of responses, measured by the number of days schools took to reply.

So as to exploit the whole sample size, the analysis also includes non-responses by treating them as right-censored observations. The imputed duration of these spells is the number of days between the date the inquiry email was sent and the date the pre-registration period

Table 2.4 Duration Analysis: Cox Proportional Model

	Panel A: Regression Coefficients					*Panel B:* Hazard Ratios				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Assimilated Immigrant	-0.0864	-0.0858	-0.0896	-0.0922	-0.0938*	0.917	0.918	0.914	0.912	0.910*
	(0.0561)	(0.0561)	(0.0562)	(0.0564)	(0.0565)	(0.0515)	(0.0515)	(0.0514)	(0.0515)	(0.0515)
Non-Assimilated Immigrant	-0.184***	-0.181***	-0.192***	-0.191***	-0.194***	0.832***	0.834***	0.825***	0.826***	0.824***
	(0.0570)	(0.0571)	(0.0571)	(0.0574)	(0.0574)	(0.0474)	(0.0476)	(0.0471)	(0.0474)	(0.0473)
N	2,551	2,551	2,551	2,527	2,527	2,551	2,551	2,551	2,527	2,527
Test on $\beta_1 = \beta_2$ (p reported)	0.091	0.099	0.078	0.088	0.087	0.091	0.099	0.078	0.088	0.087
Delivery day F.E.	✗	✓	✗	✗	✓	✗	✓	✗	✗	✓
Applicant/School Characteristics	✗	✗	✓	✗	✓	✗	✗	✓	✗	✓
Geographic Area Controls	✗	✗	✗	✓	✓	✗	✗	✗	✓	✓

Notes: The dependent variable of the regressions is the number of days each school took to respond (Cox-Proportional Model). Non-responses are treated as right-censored observations. Imputed duration for these observations is the number of days between the date the email was sent and the date the pre-registration period began. $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In order to include responses provided on the same day, I add a constant c =0.5 to response times. Applicant/School Characteristics include: indicators on the school-type (i.e. charter and private), set of dummies on the extra services provided, set of dummies on the age mentioned in the email, and a dummy for continuous shift. Due to the vast number of geographical areas fixed effects, I instead use the following geographical area controls: an indicator on whether the school is located in a left-wing area, and a set of indicators capturing municipality size. In the interest of saving space, I do not report the coefficients on the control variables.

began. To assess the suitability of the duration analysis, I perform a log-rank test of survivor function equality across family profiles, with the null hypothesis being rejected ($p = 0.0001$). I then look at the Kaplan-Meier estimates of the sample hazard rates across the three family profiles, which are displayed in Appendix Figure 2.2. I find that the response risk of each treatment group decreases abruptly after the first day and then smoothly declines. In addition, the Kaplan-Meier hazard estimates are parallel across the three family profiles. Thus, I adopt a continuous time Cox proportional hazard specification with the following form:

$$\theta_i(t) = \theta_0(t) exp(\beta_1 AI_i + \beta_2 NI_i + X_i' \delta + d), \tag{2.2}$$

where $\theta_0(t)$ is the baseline hazard rate. Again, *AI* and *NI* are indicator variables that capture whether the email was sent from the assimilated or non-assimilated family profile. X is the same set of explanatory variables described in Section 2.4 plus additional controls about the school's area, and *d* stands for the calendar day the email was sent.

Table 2.4 displays the results. The models considered are analogous to the specifications described in Section 2.4. While Panel A reports the coefficients of the Cox regression, Panel B reports their corresponding hazard-ratios.[26] Altogether, due to randomization, the estimates' sizes barely change after the inclusion of additional controls. Because both hazard ratios are strictly less than 1, Panel B reveals that being an immigrant has an associated negative effect on the hazard of obtaining a response. In particular, non-assimilated immigrants have a 82.4-83.2% hazard of receiving a response compared to natives. Alternatively, assimilated immigrants' hazard is 91-91.7% of that from natives, depending on the specification. While the significance at $p = 0.01$ for non-assimilated immigrants is not affected by the inclusion of additional covariates and fixed effects, the effect on the answer rate for assimilated immigrants becomes marginally significant ($p < 0.1$) only after the inclusion of additional controls. This finding suggests a lower statistical difference between natives and assimilated immigrants in the intensive rather than in the extensive margin of responses.

Similarly, the hazard-ratio is lower for non-assimilated families compared to assimilated immigrants, though the difference is only significant at $p < 0.1$. This result indicates that differences in response behavior between assimilated and non-assimilated migrants emerge more significantly through the extensive margin rather than through delay in response. In

[26]Hazard-ratios in Cox-Proportional models, that define the relative risk of getting a response relative to natives at any time t, give the effect size of treatments by exponentiating the regression coefficients.

quantitative terms, by comparing the associated risk reductions between these two groups, one can conclude that choosing a Spanish-name increases the hazard-ratio by approximately 10-11%, which implies a 49.4-51.1% reduction in the gap with respect to natives.[27] The size of this reduction is similar to the observed decline in response and positive response rates. Overall, the higher the effort signalled by immigrants through name-choices, the higher the probability of receiving a response at any moment in time.

2.7 Discussion

2.7.1 Interpretation of findings

The results thus far show a significantly distinctive response pattern between immigrants and natives. Moreover, I find that, among emails signed by immigrant-sounding names, those from assimilated families display more favorable responses than messages from non-assimilated parents. This pattern, as shown in Table 2.11, is generally found in schools with different characteristics. In the context of correspondence studies, these gaps are referred to a notion of discrimination since they imply a differential treatment based on ethnic origin that influences access to certain services or markets. However, the extent to which the findings can be attributed to taste-based or some form of statistical discussion remains a relevant discussion.

On the one hand, results may reflect some type of differential taste or prejudice towards migrants, whereby schools are more reluctant to interact with these type of families based on their perceived assimilation efforts. On the other hand, it is possible that name-origins imply the signalling of attributes other than the intended by the experiment. For instance, name-origin choices might be correlated with other unobserved factors affecting admission desirability (e.g. probability of being a second-generation immigrant). Thus, the observed response patterns between assimilated and non-assimilated migrants can be linked to forms of statistical discrimination that blur the attribution of my results to perceived cultural assimilation efforts.

[27] To calculate the gain of assimilation efforts, I use the ratio of the distances to the natives' implied hazard rate. Formally, $\frac{\theta(t|Nat,X,d)-\theta(t|AI,X,d)}{\theta(t|Nat,X,d)-\theta(t|NI,X,d)}$.

To limit the influence of statistical discrimination, certain characteristics, including the language proficiency and the paternal occupation, were kept similar across messages.[28] Despite efforts to fix these signals, schools may yet differently treat family profiles as a result of imperfect information. Although desirable from an experimental design perspective, adapting the emails' content to ensure strict comparability across treatment groups was deemed quite challenging in this context. The reason being that details like the country of birth of these children are not solicited in school visit requests and they can therefore render email treatments suspicious, putting the experiment at risk.

Overall, in contrast to audit studies in the labor market, the case for statistical discrimination is weaker in this setting. Profit-seeking employers may use ethnic profiling as a means for inferring expected productivity of job applicants for cost-efficient hiring. However, this paper is characterized by simple inquiries to schools that do not entail clear first-order financial or reputational consequences. First, from a financial perspective, publicly-funded schools typically rely on inputs-based funding, instead of employing a formula-based funding (REDE (2020)). Thus, considerations like the additional first- as opposed to second-generation immigrant enrollment imply limited direct impact on their resources. Second, as mentioned earlier, schools' academic results are no longer available to the general public. Hence, these moderating factors do not imply clear incentives to treat immigrants and native children differently.

Yet, it cannot be ruled out the possibility that school staff perceive as significant these or other non-monetary costs to serving these type of families. In absence of direct evidence to disentangle the influence of taste-based *vis-à-vis* statistical discrimination, some patterns suggest that the latter is indeed present and that the observed differences can be attributed, at least partially, to perceived assimilation efforts. The results indicate a wider gap between assimilated and non-assimilated profiles in kindergartens relative to other schools. This evidence, although statistically non significant, is consistent with perceived differences in children's birthplace not being the primary driver of results. In line with previous evidence, the geographical variation in treatment effects according to the incidence of Romanian migration

[28]The paternal occupation should act as a signal that families belong to the middle class. By focusing on this subset of families, this choice potentially limits the representativeness of discrimination in the general population. In this sense, given the high socioeconomic segregation in the region's schools, one could expect the response gap to be larger between the average native and Romanian families.

is compatible with inter-group exposure mediating intensity of prejudice. However, these exploratory patterns are only suggestive.

2.7.2 Limitations of the experiment

The results and design of this experiment are subject to a number of caveats. First, it could be debated whether the selected names capture the origin and cultural orientation of the fictitious families. To effectively evaluate the impact of being a migrant, school-staff must associate newcomers' names and surnames with being (Romanian) immigrants. These names (i.e., Ion, Ioan, Mihaela, Ioana, and Nicolae) are not listed among the 100 most common names of Spanish-born citizens and are rarely used by other immigrant groups. Moreover, the two selected immigrant-origin surnames (i.e. Dumitru and Mihai) are, according to the INE, significantly associated only with Romanian-born immigrants. This fact, together with the salience of Romanian immigrants in the Community of Madrid, is likely to induce the desired association and alleviate this concern.

Second, email inquiries may not provide the recipients with precise and clear information to account for all the differences in relevant socioeconomic characteristics that affect discrimination across treatment groups. For instance, the signal of parental occupation, transmitted through the email address, might go relatively unnoticed by message recipients. Estimates on the size of ethnic discrimination may thus be upward biased if socioeconomic discrimination exists. I considered the possibility of explicitly accounting for additional dimensions by providing more information in the email text. However, the inclusion of further information could hamper the credibility of the emails, and put the validity of the experiment at risk by rendering the messages suspicious to administrative staff.

Third, strong correlations between the identity of email-receivers and differences in the propensity of interacting with ethnically-distinct families could raise concern. Assume, for instance, that women are more likely to respond with equal probability to natives and immigrants. If this was true, the estimates may be a result of distinct gender compositions of email-receivers across treatment groups. Although not directly verifiable since email recipients' identities are typically unobserved, randomization should ensure that these characteristics are also balanced across treatment groups, thus alleviating this concern. To provide a suggestive test on whether different recipient groups show disparate responding behaviors across family profiles, I check whether imbalances in the distribution of respondents' gender and occupations exist. I find significant disparities in some of these characteristics (shown in Appendix Table 2.13). However, the differences are modest in quantitative terms, which suggests that the statistical significance is primarily driven by sampling bias.

Finally, the results may capture the (un)conscious negative prejudice biases of individual respondents rather than schools' institutional behavior. If this was the primary mechanism, the results would suggest that the presence of non-coordinated individual discriminatory biases could result in generalized discrimination towards immigrants within an entire education system. Unfortunately, the data did not allow to shed light on the validity of this hypothesis.

2.8 Concluding Remarks

Using a field experiment, this paper presents evidence that schools are more or less reluctant to provide relevant information to immigrant parents based on the name-origin of children. Generally, my results reveal that there are significant signs of discrimination against Romanian immigrants. This is a particularly interesting finding given the relative cultural similarity of Romanian immigrants to Spanish natives, as well as the well-established nature of their community in the region of Madrid. There is thus reason to believe that other, more culturally distant, immigrant groups may suffer to a greater degree from prejudiced attitudes.

I also find that immigrants who choose a Spanish name for their son, which is arguably a subtle but meaningful signal of cultural assimilation, are 50% less discriminated. This observation is in line with schools taking into account efforts towards cultural adaptation on the part of immigrant families. All in all, this result provides suggestive evidence that cultural considerations, like name-origins of children, can play a significant role in shaping the extent of discrimination in the access to education services.

The results indicate substantial variation in the size of treatment effects based on the characteristics of the school. This heterogeneity is however statistically non-significant, suggesting that results might suffer from statistical power limitations. Notably, I find that discrimination varies substantially according to the age of child, indicating that the perceived country of birth of children is not the main driver of results. At the same time, schools in areas with a majoritarian presence of Romanian immigrants display smaller response gaps for both assimilated and non-assimilated families relative to natives. This pattern is consistent with inter-group contact being a relevant determinant of discriminatory attitudes. Additionally, there is evidence that schools largely display discriminatory attitudes towards immigrants, with Spanish-named children being more favourably treated, regardless of their characteristics.

The analysis presents several limitations. First, my findings do not necessarily generalize to female children. The student's gender might affect the discrimination due to average differences in classroom behavior. While this dimension of study is clearly interesting, I did not include it in the paper due to reduced of statistical power from increasing the number of treatment arms. Second, the signal of parental occupation is quite subtle and can go relatively unnoticed by email recipients. Hence, my results might capture some aspects of socioeconomic discrimination. A third related limitation is that the experimental design is limited in its capacity to determine the relative role of statistical and taste-based discrimination in explaining the results.

The findings of this article have important policy implications. First, they underscore the relevance of ensuring equitable access to information in order to protect the integrity of the school assignment procedure. To this end, public authorities need to more closely monitor the information acquisition process. Possible interventions include promoting the use of audits by school officials or establishing communication channels that hide applicants' identity traits. Second, the analysis suggests that subtle cultural expressions imply significant costs for immigrant families. This form of discrimination results in economic incentives deterring foreign cultural transmission and promoting assimilation efforts (Algan *et al.* (2013), Bisin & Tura (2019), Fouka (2019)). Thus, resorting to forced assimilation policies might be unwarranted in settings where the economic penalty for out-group cultural transmission is large. The reason being that assimilation policies might result in an identity backlash (Fouka (2020)) and offset the initial incentives for assimilation.[29] Finally, this paper highlights the importance of tackling ethnic discrimination in public services. In spite of their illegality, the persistence of these practices indicates that they can hardly be eliminated by legislative fiat. In

[29]Whether attaining cultural homogenization is a desirable outcome depends on one's social preferences. The analysis of this paper is only positive. It is not my intent to provide any normative statement about the desirability of enforcing immigrants' cultural assimilation.

this sense, promoting ethnic diversity among school employees is a possible policy to mitigate this type of discriminatory habits (Giulietti *et al.* (2019)).

2.9 Appendix

2.9.1 Data Sources and Variable Description

Table 2.5 Family identities and email text

FAMILY PROFILE	fathername	mothername	sonname	surname
Native	Manuel	Carmen	Javier	Romero
Assimilated Immigrant	Ioan	Mihaela	Javier	Mihai
Non-Assimilated Immigrant	Ion	Ioana	Nicolae	Dumitru

Email text

****

From: <fathername>.<surname>@reformas<surname>.net

****

Hello,

Our names are <fathername> and <mothername>. We are looking for a school for our son <sonname>, who is 5 years old. He will start the first grade of primary next September. We are considering this school as an option. Would it be possible to arrange a meeting to visit the school?

Best,

<fathername> and <mothername>

Table 2.6 Data Sources

Variable	Description	Observational Unit	Year	Source
Public school	Takes value 1 if the school is public, 0 otherwise	School	Nov, 2017	MSSE
Charter school	Takes value 1 if the school is charter, 0 otherwise	School	Nov, 2017	MSSE
Private school	Takes value 1 if the school is private, 0 otherwise	School	Nov, 2017	MSSE
Non-public schools	Takes value 1 if the school is charter or private, 0 otherwise	School	Nov, 2017	MSSE
Kindergarten	Takes value 1 if the school only provides first-cycle of pre-primary education, 0 otherwise	School	Nov, 2017	MSSE
Bilingual School	Takes value 1 if school operates under Bilingual program , 0 otherwise	School	Nov, 2017	MSSE
Lunch services	Takes value 1 if school provides lunch service to students, 0 otherwise	School	Nov, 2017	MSSE
Transport service	Takes value 1 if school provides transport service to students, 0 otherwise	School	Nov, 2017	MSSE
Extended day service	Takes value 1 if school provides 1 daily hour of extra-teaching to students, 0 otherwise	School	Nov, 2017	MSSE
Continuous shift	Takes value 1 if school operates under an intensive continuous shift, 0 otherwise	School	Nov, 2017	MSSE
Number of services	Sum of transport, lunch and extended day services	School	Nov, 2017	MSSE
Very small muni.	Takes value 1 if the municipality where school is located <10,000, 0 otherwise	Municipality	2017	*Almudena*
Small muni.	Takes value 1 if the municipality where school is located is 10,000-50,000, 0 otherwise	Municipality	2017	*Almudena*

Table 2.7 Data Sources (cont'd)

Variable	Description	Observational Unit	Year	Source
Medium muni.	Takes value 1 if the municipality where school is located is 50,000-100,000, 0 otherwise	Municipality	2017	*Almudena*
Big muni. (Madrid excluded)	Takes value 1 if the municipality where school is located is >100,000 but is not Madrid city, 0 otherwise	Municipality	2017	*Almudena*
Madrid	Takes value 1 if the school is located in Madrid city, 0 otherwise	Municipality	2017	*Almudena*
Left-wing area	Takes value 1 if combined vote shares of (PSOE+UP)$>$ combined vote shares of (PP+C's) in the 2016 General Elections at the municipality/neighborhood	Neighborhood (if Madrid=1)	2016	Madrid City Council
		Municipality (if Madrid=0)	2016	*Almudena*
Housing: 1,000€ per sq meter	Average sq meter price of second hand apartments at the neighborhood where school is located	Neighborhood (available only if Madrid=1)	2018	*Madrid City Council*
Housing: Tercile X	Takes value 1 if the school is located in the neighborhood with X^{st} tercile in price per sq meter, 0 otherwise	Neighborhood (available only if Madrid=1)	2018	Madrid City Council
Immigration: % Romanian population	Share of population in the municipality or neighborhood (if Madrid==1) where the school is located that have Romanian nationality	Municipality and Neighborhood (if Madrid=1)	2015	*Padrón Municipal* (through Madrid's Institute of Statistics)
Immigration: Romanian majoritarian	Takes value 1 if school is located in a municipality or neighborhood (if Madrid==1) where Romanians constitute the largest foreign-nationality group	Neighborhood (available only if Madrid=1)	2015	*Padrón Municipal* (through Madrid's Institute of Statistics)

Notes: PSOE (*Spanish Socialist Worker's Party*) is the mainstream center-left social democratic party. UP (*United We Can*) is the republican far-left to left-wing party. PP (*People's Party*) is the mainstream center-right to right-wing conservative and Catholic party. C's (*Citizens*) is the center-right economically liberal party.

Table 2.8 Balancing Test (After Attrition)

SCHOOL CHARACTERISTICS	Nat	AI	NI	t-test (p-value)		
	(1)	(2)	(3)	(1)-(2)	(1)-(3)	(2)-(3)
Age signalled = 1	0.377	0.370	0.377	0.768	0.985	0.783
	(0.485)	(0.483)	(0.485)			
Age signalled = 2	0.488	0.499	0.509	0.646	0.370	0.663
	(0.500)	(0.500)	(0.500)			
Age signalled = 5	0.012	0.011	0.007	0.824	0.314	0.431
	(0.108)	(0.103)	(0.084)			
Age signalled = 11	0.123	0.120	0.107	0.845	0.284	0.381
	(0.329)	(0.325)	(0.309)			
Public school	0.577	0.580	0.556	0.893	0.391	0.321
	(0.494)	(0.494)	(0.497)			
Charter school	0.154	0.175	0.175	0.252	0.244	0.985
	(0.361)	(0.380)	(0.380)			
Private school	0.269	0.245	0.269	0.262	0.988	0.268
	(0.444)	(0.431)	(0.444)			
Bilingual school	0.254	0.272	0.270	0.385	0.449	0.910
	(0.435)	(0.445)	(0.444)			
Continuous shift	0.428	0.416	0.400	0.633	0.250	0.501
	(0.495)	(0.493)	(0.490)			
Number of extra-services	1.700	1.717	1.710	0.688	0.812	0.867
	(0.850)	(0.856)	(0.843)			
Very Small Muni.	0.106	0.099	0.099	0.649	0.626	0.974
	(0.308)	(0.299)	(0.298)			
Small Muni.	0.133	0.138	0.148	0.755	0.370	0.559
	(0.340)	(0.345)	(0.355)			
Medium Muni.	0.129	0.124	0.123	0.736	0.709	0.971
	(0.336)	(0.330)	(0.329)			
Big Muni. (Madrid city excluded)	0.242	0.244	0.235	0.922	0.723	0.651
	(0.429)	(0.430)	(0.424)			
Madrid city	0.390	0.395	0.396	0.836	0.819	0.983
	(0.488)	(0.489)	(0.489)			
Left-wing area	0.275	0.267	0.265	0.695	0.652	0.953
	(0.447)	(0.442)	(0.442)			
Housing: 1,000€ per sq meter	2.824	2.805	2.791	0.833	0.701	0.862
	(1.131)	(1.101)	(1.090)			
Housing: Tercile = 1	0.398	0.391	0.398	0.863	0.999	0.862
	(0.490)	(0.489)	(0.490)			
Housing: Tercile = 2	0.313	0.325	0.326	0.738	0.716	0.977
	(0.465)	(0.469)	(0.470)			
Housing: Tercile = 3	0.289	0.284	0.276	0.874	0.705	0.826
	(0.454)	(0.451)	(0.448)			
Immigration: % Romanian population	3.256	2.999	3.129	0.111	0.450	0.415
	(3.467)	(3.142)	(3.424)			
Immigration: Romanian majoritarian	0.730	0.726	0.711	0.831	0.378	0.505
	(0.444)	(0.447)	(0.454)			
N	851	848	852			

Notes: Nat, AI and NI stand, respectively, for natives, *assimilated*, and *non-assimilated* immigrants. Standard deviations in parentheses. p-value refers to the *t*-test for the difference in means for each pairwise comparison across treatment groups: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Information on housing prices is only available for schools located in Madrid city ($N_{Nat} = 332$, $N_{AI} = 335$ and $N_{NI} = 337$). Information on immigrant shares exclude schools (24) from three neighborhoods due to data incompatibility of neighborhood definitions across sources. Data sources of these variables are presented in Table 2.6 in Appendix.

2.9.2 Additional Tables and Figures

Table 2.9 Difference in Response Rates and Positive Response Rate (Probit Analysis)

	Panel A: Response					*Panel B:* Positive Response				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Assimilated Immigrant	-0.046**	-0.046**	-0.046**	-0.048**	-0.046**	-0.051**	-0.051**	-0.052**	-0.052**	-0.051**
	(0.022)	(0.022)	(0.021)	(0.022)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Non-Assimilated Immigrant	-0.090***	-0.089***	-0.090***	-0.092***	-0.090***	-0.092***	-0.091***	-0.093***	-0.094***	-0.093***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.022)	(0.021)	(0.022)	(0.021)
N	2,551	2,551	2,551	2,527	2,527	2,551	2,551	2,551	2,527	2,527
Native group mean	0.771	0.771	0.771	0.771	0.771	0.750	0.750	0.750	0.750	0.750
Test on $\beta_1 = \beta_2$ (p reported)	0.039	0.043	0.033	0.038	0.033	0.056	0.061	0.049	0.055	0.047
Delivery day F.E.	✗	✓	✗	✗	✓	✗	✓	✗	✗	✓
Applicant/School Characteristics	✗	✗	✓	✗	✓	✗	✗	✓	✗	✓
Geographic Area Controls	✗	✗	✗	✓	✓	✗	✗	✗	✓	✓

Notes: The outcome variables of the regressions are: (i) an indicator on whether the school responded to the inquiry email (Panel A), and (ii) an indicator on whether a positive response to the email was provided (Panel B). Robust standard errors in parentheses: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Marginal effects reported. Applicant/School Characteristics include: indicators on the school-type (i.e. charter and private), set of dummies on the extra services provided, set of dummies on the age mentioned in the email, and a dummy for continuous shift. Due to the vast number of geographical areas fixed effects, I instead use the following geographical area controls: an indicator on whether the school is located in a left-wing area, and a set of indicators capturing municipality size. In the interest of saving space, I do not report the coefficients on the control variables.

Table 2.10 Difference in Invitations to Personal Visits (Non-Responses Excluded)

	(1)	(2)	(3)	(4)	(5)
Assimilated Inmigrant	-0.042	-0.042	-0.042*	-0.055*	-0.039
	(0.028)	(0.028)	(0.025)	(0.030)	(0.027)
Non-Assimilated Inmigrant	-0.045	-0.046	-0.045*	-0.059*	-0.052*
	(0.028)	(0.028)	(0.025)	(0.031)	(0.027)
N	1,854	1,854	1,854	1,854	1,854
Native group mean	0.562	0.562	0.562	0.562	0.562
Test on on $\beta_1 = \beta_2$ (p reported)	0.914	0.900	0.892	0.900	0.642
R^2	0.002	0.002	0.232	0.168	0.370
Delivery day F.E.	✗	✓	✗	✗	✓
Applicant/School Characteristics	✗	✗	✓	✗	✓
Geographic Area F.E.	✗	✗	✗	✓	✓

Notes: The outcome variable of the regressions is an indicator that takes value 1 if the school invited families for a personal visit and 0 otherwise, including in the latter category the invitations to open doors. Non-responses are excluded from the analysis. Robust standard errors in parentheses: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Applicant/School Characteristics include: indicators on the school-type (i.e., charter and private), a set of dummies on the extra services provided, a set of dummies on the age mentioned in the email, and a dummy for continuous shift. In the interest of saving space, I do not report the coefficients on the control variables.

Table 2.11 Differences in Response Rates and Positive Response Rate. Separate LPM regressions by subgroup

SUBGROUP	Age of Instruction		School Type			Language of Instruction		Political Orientation	
	Non-Kindergarten	Kindergarten	Public	Charter	Private	Non-Bilingual	Bilingual	Left-Wing	Right-Wing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Response Rates									
Assimilated Immigrant	-0.051**	-0.036	-0.063**	-0.039	-0.008	-0.053	-0.051	-0.026	-0.051**
	(0.024)	(0.038)	(0.027)	(0.053)	(0.043)	(0.033)	(0.036)	(0.041)	(0.025)
Non-Assimilated Immigrant	-0.069***	-0.125***	-0.098***	-0.064	-0.087**	-0.101***	-0.030	-0.084**	-0.093***
	(0.025)	(0.038)	(0.028)	(0.054)	(0.043)	(0.034)	(0.035)	(0.042)	(0.025)
Intercept (Native)	0.834***	0.667***	0.796***	0.748***	0.729***	0.822***	0.852***	0.761***	0.775***
	(0.016)	(0.026)	(0.018)	(0.038)	(0.029)	(0.022)	(0.024)	(0.028)	(0.017)
N	1,595	956	1,457	428	666	918	677	686	1,865
Test $\beta_1 = \beta_2$ (p reported)	0.479	0.023	0.223	0.642	0.075	0.176	0.568	0.180	0.110
R^2	0.005	0.012	0.009	0.003	0.007	0.010	0.003	0.006	0.007
Panel B: Positive Response Rates									
Assimilated Immigrant	-0.062**	-0.030	-0.077***	-0.021	-0.006	-0.050	-0.080**	-0.022	-0.059**
	(0.025)	(0.039)	(0.028)	(0.055)	(0.044)	(0.034)	(0.038)	(0.041)	(0.025)
Non-Assimilated Immigrant	-0.083***	-0.109***	-0.105***	-0.059	-0.083*	-0.105***	-0.055	-0.075*	-0.099***
	(0.026)	(0.039)	(0.029)	(0.056)	(0.044)	(0.035)	(0.037)	(0.042)	(0.026)
Intercept (Native)	0.817***	0.639***	0.780***	0.710***	0.707***	0.803***	0.838***	0.744***	0.752***
	(0.017)	(0.027)	(0.019)	(0.040)	(0.030)	(0.023)	(0.025)	(0.029)	(0.017)
N	1,595	956	1,457	428	666	918	677	686	1,865
Test $\beta_1 = \beta_2$ (p reported)	0.450	0.045	0.345	0.486	0.087	0.132	0.524	0.221	0.136
R^2	0.007	0.009	0.010	0.003	0.007	0.010	0.007	0.005	0.008

Notes: The outcome variables of the regressions are: (i) an indicator on whether the school responded to the inquiry email (Panel A), and (ii) an indicator on whether a positive response to the email was provided (Panel B). Robust standard errors in parentheses: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. p-values report the result of the t-test on equality of marginal effects. Regression on bilingual schools does not include kindergartens. Additional details on the variables can be found in Appendix Table 2.6.

Table 2.12 Differences in Response Rates and Positive Response Rate. Separate LPM regressions by subgroup (cont'd)

	Size of Municipality					Housing Prices (if Madrid=1)			Romanian majority area	
SUBGROUP	V.Small	Small	Medium	Big	Madrid	Tercile 1	Tercile 2	Tercile 3	Non-Romanian	Romanian
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
Panel A: Response Rates										
Assimilated Immigrant	-0.098	-0.018	-0.058	-0.062	-0.028	-0.017	-0.082	0.018	-0.085**	-0.028
	(0.075)	(0.056)	(0.059)	(0.040)	(0.033)	(0.051)	(0.059)	(0.064)	(0.041)	(0.025)
Non-Assimilated Immigrant	-0.110	0.015	-0.106*	-0.101**	-0.115***	-0.161***	-0.107*	-0.062	-0.137***	-0.072***
	(0.075)	(0.053)	(0.061)	(0.042)	(0.035)	(0.055)	(0.059)	(0.067)	(0.041)	(0.025)
Intercept (Native)	0.633***	0.779***	0.782***	0.816***	0.774***	0.788***	0.798***	0.729***	0.776***	0.768***
	(0.051)	(0.039)	(0.040)	(0.027)	(0.023)	(0.036)	(0.040)	(0.046)	(0.028)	(0.017)
N	258	356	320	613	1,004	397	323	284	702	1,825
Test $\beta_1 = \beta_2$ (p reported)	0.878	0.539	0.454	0.379	0.013	0.010	0.691	0.226	0.231	0.093
R^2	0.010	0.001	0.010	0.009	0.012	0.027	0.011	0.006	0.015	0.004
Panel B: Positive Response Rates										
Assimilated Immigrant	-0.087	-0.009	-0.077	-0.057	-0.043	-0.017	-0.128**	0.018	-0.089**	-0.033
	(0.075)	(0.058)	(0.060)	(0.042)	(0.034)	(0.053)	(0.060)	(0.068)	(0.042)	(0.025)
Non-Assimilated Immigrant	-0.087	0.034	-0.106*	-0.111**	-0.127***	-0.153***	-0.134**	-0.086	-0.148***	-0.070***
	(0.075)	(0.055)	(0.061)	(0.043)	(0.036)	(0.056)	(0.060)	(0.070)	(0.042)	(0.026)
Intercept (Native)	0.611***	0.752***	0.773***	0.796***	0.750***	0.765***	0.798***	0.677***	0.754***	0.747***
	(0.052)	(0.041)	(0.040)	(0.028)	(0.024)	(0.037)	(0.040)	(0.048)	(0.029)	(0.018)
N	258	356	320	613	1,004	397	323	284	702	1,825
Test $\beta_1 = \beta_2$ (p reported)	1.000	0.442	0.659	0.229	0.020	0.017	0.924	0.140	0.185	0.165
R^2	0.007	0.002	0.010	0.011	0.013	0.023	0.018	0.009	0.017	0.004

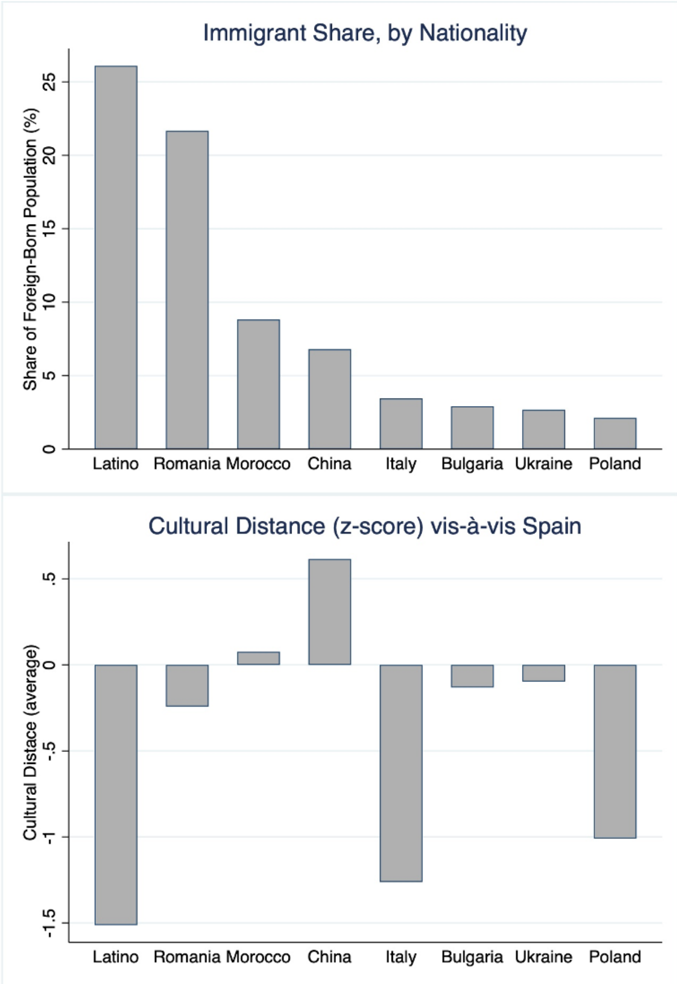
Notes: The outcome variables of the regressions are: (i) an indicator on whether the school responded to the inquiry email (Panel A), and (ii) an indicator on whether a positive response to the email was provided (Panel B). Robust standard errors in parentheses: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. p-values report the result of the t-test on equality of marginal effects. Additional details on the variables can be found in Appendix Table 2.6.

Table 2.13 Balancing Test of Gender and Occupation of Respondents

RESPONDENT CHARACTERISTICS	Nat (1)	AI (2)	NI (3)	t-test (p-value) (1) vs (2)	(1) vs (3)	(2) vs (3)
Director	0.465	0.435	0.440	0.285	0.373	0.873
	(0.499)	(0.496)	(0.497)			
Secretary	0.104	0.122	0.140	0.308	0.052*	0.359
	(0.305)	(0.327)	(0.347)			
Other positions	0.061	0.062	0.041	0.958	0.121	0.114
	(0.239)	(0.241)	(0.199)			
Unknown position	0.174	0.148	0.167	0.207	0.761	0.355
	(0.379)	(0.355)	(0.374)			
Male	0.169	0.157	0.133	0.572	0.075*	0.226
	(0.375)	(0.365)	(0.340)			
Female	0.511	0.456	0.488	0.052*	0.425	0.272
	(0.500)	(0.498)	(0.500)			
Both genders	0.002	0.000	0.000	0.333	0.347	.
	(0.039)	(0.000)	(0.000)			
Unknown gender	0.319	0.386	0.379	0.011**	0.025**	0.802
	(0.466)	(0.487)	(0.486)			
N	656	616	580			

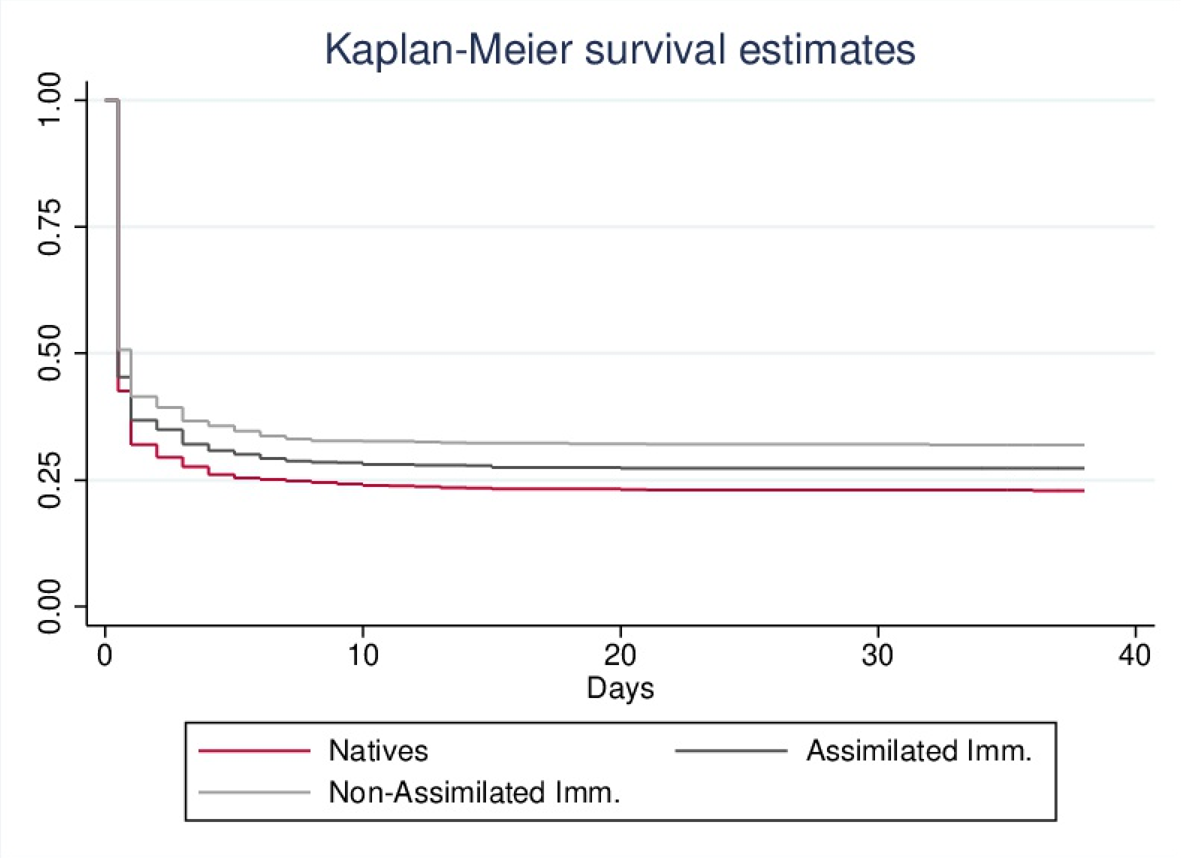
Notes: Nat, AI and NI stand, respectively, for natives, *assimilated* and *non-assimilated* immigrants. Standard deviations in parentheses: $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. p-value refers to the t-test for the difference in means for each pairwise comparison across treatment groups.

Figure 2.1 Foreign-Immigrant Share and Cultural Distance, by Nationality



Notes: The upper figure displays the share of population with certain nationality among the foreign-born population. The analysis considers the most-numerous 15 foreign nationalities in the region of Madrid in 2018. Latinos include citizens from the following nationalities: Colombia, Ecuador, Venezuela, Peru, Dominican Republic, Paraguay, Bolivia, Honduras. The lower panel displays a measure of cultural distance using the database from Spolaore & Wacziarg (2016) for these nationalities. This measure is constructed by averaging the genetic, linguistic and religious distances of the different nationalities *vis-à-vis* Spain. The individual distances were transformed into *z*-score to homogenize scales. To aggregate distances from Latin American countries into a single group, I use a population-weighted average by using the relative presence of migrants from each nationality as weights.

Figure 2.2 Kaplan-Meier survival estimates



3 Comparing Teacher and External Assessments: Are Boys, Immigrants, and Poorer Students Undergraded?

This paper was jointly co-authored with

Lucas Gortazar and Ainhoa Vega-Bayo

3.1 Introduction

Student assessments undertaken by teachers in the classroom are a fundamental pedagogic milestone for children progress. In their association with student promotion, teachers' grading criteria of summative assessments are critically relevant.[1] In particular, if teachers overuse their evaluations as a form of certification in basic education, students may end up accumulating various failed subjects, and thus repeating grades disproportionately. International comparative evidence suggests that the extent of these practices is unequal across different countries. For example, some education systems sharing the French influence (i.e., South Europe, French-speaking Africa, or Latin America) keep over-relying on excessively failing classroom grades in basic education. Furthermore, this phenomenon occurs even if the national level of skills, as measured by international assessments, is relatively high. For instance, countries like Belgium or Spain have high proportion of repeaters, similar to those in developing countries, even if the level of acquired skills is comparatively higher.

By increasing retention, these grading habits have relevant consequences on the equity and efficiency features of the education system. Repetition is associated with higher dropout, and remains as an ineffective and expensive education policy relative to other alternatives

[1]There are two different dimensions in teachers' assessments. On the one hand, there is a formative aspect, which is designed to provide feedback and to promote students' growth. On the other hand, there is a summative form of assessment, that is conceived to evaluate the acquisition of content knowledge and skills.

[Hattie (2008)]. Likewise, low grading decisions significantly influence students' confidence and self-perception [Bobba & Frisancho (2014)], and affect parental educational investment levels through the influence on parental beliefs about their children's ability and needs [Kinsler *et al.* (2014)]. At the same time, repetition tends to disproportionately affect disadvantaged students, immigrant students and boys, no matter their background skills. Consequently, grade retention policies become a source of vertical segregation by socioeconomic status [OECD (2016)], gender and immigrant origin [Ikeda & García (2014)]. Whether this phenomenon of repetition and low grading is a form of teacher cultural practices, or whether it comes from direct policy regulation is a topic widely discussed in the comparative policy literature [Eurydice (2017)]. With this motivation in mind, we ask the following questions: Do teachers provide accurate scores that are based solely on the educational performance of the student in a given subject test? If not, to what extent teachers' assessments hamper students from these socio-demographic groups?

In this essay, we focus on Spain, where approximately 30% of students repeat at least once by the end of compulsory education [Save the Children (2019)]. Despite high regional variation, the retention rate remains well-above the EU average (11%), even for the most advanced regions. For instance, in the Basque Country (our region of study) more than 20% of students have repeated at least once by the end of compulsory education [Save the Children (2019)]. Moreover, disadvantaged children, immigrant students and boys tend to disproportionately repeat relative to advantaged children, native students and girls, even after considering their level of skills [Save the Children Spain (2019)].

To answer the questions at hand, we examine the existence of distinct grading biases based on the several students' characteristics by primary and middle school teachers of the Basque Country. To do so, we study the presence of systematic differences between the scores of internal and external evaluations based on three key student characteristics: gender, national origin and socioeconomic background of the family.[2] The empirical analysis relies on comparing test scores from teacher-graded internal evaluations with the same student's results on the external diagnosis' assessments, which are standardized as well as quasi-blindly graded,

[2] In what remains, we follow Calsamiglia & Loviglio (2019) and define *grading bias* as the differential effect of a given student characteristic on internal assessments besides their cognitive skill, as measured by external evaluations. Consequently, our results should not be interpreted as direct evidence of teaching discrimination.

and hence serve as a counterfactual.[3] The non-blind nature of internal assessments allows us to evaluate the impact of certain intangibles (e.g. student skills not captured by the test or teacher biases) that the quasi-blind nature of external assessments are unlikely to capture.

Our analysis employs large student-level administrative data of the Basque education system. In particular, we have access to the census of two separate cohorts from students in the publicly-funded education network (i.e., one in primary and one in middle school). To evaluate the presence of grading biases between several subjects, we rely on available information of students' subject-specific performance in both internal and external assessments. The first evaluation type is decided by school teachers and provided at the end of the school year (although they are typically based on different assessments conducted throughout the whole school year). In contrast, the second assessment type corresponds to standardized diagnoses low-stake tests administered to students in 4^{th} grade of primary (age 10 approximately) and 2^{nd} year of middle school (equivalent to 8^{th} grade in the US, age 14 approximately). These are designed by a public authority and quasi-blindly graded during the months of April and May. To control for the sorting of students into schools and classrooms plus the nonrandom teacher-student assignments, our analysis relies on the use of class fixed effects.

The results of our main empirical analysis point towards the existence of significant grading biases from the teachers' part. First, we find that female students are significantly favored relative to male pupils. The results are highly significant for all subjects in middle school (between 0.224-0.385σ, depending on the subject, with our baseline specification) and in primary (0.105-0.275σ), with the exception of Math (0.013σ). Second, we find that first-generation immigrants obtain systematically lower internal grades than their native analogs, both in primary (0.215-0.316σ) and middle school (0.118-0.213σ). In contrast, we observe that second-generation immigrants suffer from smaller grade penalization. The effect is however only significant in primary education (0.078-0.237σ). Finally, we find a significant socioeconomic gradient in both primary and middle school, for all subjects. The results suggest a positive and strictly increasing effect of household income on the extent of overassessment.

To validate the stability of our results, we perform a number of robustness checks. In particular, we experiment with expanding the set of controls, with adopting a more flexible

[3]External assessments are graded by a marker with no affiliation with the school and no personal relationship with the student. However, these examinations also include an oral component. Thus, we cannot describe them as purely blind because examiners might deduce some of the students' socio-demographic characteristics.

functional form, and with employing several specifications of the dependent variable. The main qualitative and quantitative nature of the results withstand these checks. Additionally, we examine an alternative specification to address the endogeneity of external assessments. For this purpose, we use an instrumental variable approach. In the vein of Calsamiglia & Loviglio (2019) and Terrier (2020), we rely on the month of birth as our instrumental variable for the score in external standardized evaluations. Overall, the findings are robust to the use of IV; with the exception of first-generation immigrants in middle school.

We explore several mechanisms through which the observed assessment gaps may arise. First, we exclude the possibility that our results are driven by grading biases in specific parts of the ability distribution (with the exception of immigrants in middle school). Second, we find that the gender bias is significant in groups where girls outclass boys, and vice versa. Similarly, the results remain stable when we limit the sample to classes where the performance variance of girls is higher than that from boys. Altogether, this suggests that our findings are not primarily driven by statistical discrimination [Lavy (2008)]. Finally, we explore the role of student-specific unobservables (e.g., behavior in the classroom or effort). To this end, we evaluate whether substantial between-subjects grading bias heterogeneity exists. We achieve this by relying on within-student across-subject variation with the inclusion of student fixed effects. The findings confirm the presence of significant between-subjects heterogeneity. Overall, we believe that these results are consistent with contemporaneous studies that find stereotyping differences across subjects [Carlana (2019), Alesina *et al.* (2018)]. However, we cannot exclude that some other competing mechanisms are behind this result.

This study deepens the knowledge of the policy debate around grading and retention, and its relation to cultural factors and biases. By doing so, we contribute to the literature on grading biases that exploits the presence of systematic differences between non-blind and blind evaluations. In their seminal paper, Goldin & Rouse (2000) exploit the adoption of blind auditions for the symphony orchestra to study the existence of discrimination against female musicians. They find that blind auditions increased the female presence in the orchestra. Focusing on formal education, Lavy (2008) takes advantage of a natural experiment based on the matriculation exams of students in Israel and finds, on the contrary, that male students face significant discrimination in each subject. The evidence towards the presence of positively biased results towards female students has been found to be consistent ever since. Internal assessments are found to damage male relative to female students in Portugal [Ângelo & Reis (2017)], France [Terrier (2020)], Italy [Di Liberto & Casula (2016)] and several regions of Spain (i.e., Andalusia [Marcenaro-Gutierrez & Vignoles (2015)] and Catalonia [Calsamiglia & Loviglio (2019)]). As for immigrants, Burgess & Greaves (2013) find that while some ethnic minority groups are systematically provided with lower grades compared to their white peers, some minority groups are overassessed. At the same time, the evidence also hints at the existence of underassessment towards students with lower socioeconomic background in India [Hanna & Linden (2012)][4], France [Cosnefroy & Rocher (2004)] and Andalusia (Spain) [Marcenaro-Gutierrez & Vignoles (2015)].

The remainder of the chapter is organized as follows. Section 3.2 presents a simple theoretical model that captures the main elements of our identification strategy. Section 3.3 introduces the different data sources and sample restrictions. Section 3.4 describes our main empirical strategy. Section 3.5 displays the main results. Section 3.6 discusses the robustness checks. Section 3.7 provides evidence on mechanisms and Section 3.8 concludes.

3.2 A Simple Theoretical Model

In this section, we present a simple framework that illustrates our identification strategy. In particular, we have in mind the setting provided by Burgess & Greaves (2013). Let A_{is} be the underlying unobserved ability of student i in subject s. There are two measures of students' ability, for each subject: the external standardized assessments (hereafter, ext_{is}) and the internal teacher assessments (hereafter, int_{is}). We assume that ext_{is} are free from grading bias, but int_{is} are affected by teacher j's idiosyncratic assessing standards. More precisely, we adopt the following specification for ext_{is}:

$$ext_{is} = \alpha_s^{ext} A_{is} + \gamma_s^{ext} F_i + \phi_s e_{is}^{ext} + \varepsilon_{is}^{ext}, \tag{3.1}$$

where e_{is}^{ext} is the level of effort that student i exerts in the ext on subject s. We allow systematic factors correlated with student i's trait, F_i, to directly affect ext via γ_s^{ext}. For the sake of clarity,

[4] Although related to the above-mentioned literature strand in its aim, this paper does not rely on the presence of blind and non-blind assessments. In particular, the authors randomize the children's caste to exam cover sheets to uncover the presence of this discrimination in India.

we consider only one student trait in the model and treat it as gender (F_i = 1 if student i is female and 0 otherwise). In the latter analysis, we further study the presence of grading biases towards other distinct traits.[5] Finally, ε_{is}^{ext} is an individual random shock specific to external assessments.

In contrast, internal assessments follow a similar functional form, but are affected by teacher j's idiosyncratic grading biases. Formally,

$$int_{is} = \alpha_s^{int} A_{is} + \{\gamma_s^{int} + \tau_{js}\} F_i + \phi_s e_{is}^{int} + \varepsilon_{is}^{int}, \tag{3.2}$$

where τ_{js} reflects teacher j's distinct grading attitudes towards female students in subject s and is our main parameter of interest. Note that we allow the relationship between unobservables correlated with F_i to vary between ext and int. By doing so, our formulation captures two complementary confounding sources. On the one hand, we allow the same unobservable characteristics to differently affect ext and int. On the other, we enable different factors correlated with F_i to affect each assessment type. For example, gender might be correlated with the level of disrupting behavior exerted while in the classroom. This is an arguably relevant determinant of int, while it is unlikely to significantly affect ext.

In our data, we would ideally observe effort levels e_{is}^{int} and e_{is}^{ext}. Unfortunately, we do not have access to micro-data on item responses for any of the assessments nor behavioral information about students to proxy for student effort.[6] To circumvent this limitation, we assume that $e_{is}^k = \kappa_s^k A_{is} + \delta_s^k F_i + \upsilon_{is}^k$, where $k \in \{int, ext\}$ and υ_{is}^k is an idiosyncratic random error.

As a point of departure, we evaluate how the gap between between int and ext varies with F_i. Subtracting Equation (3.1) to (3.2), we have that:

$$d_{is} \equiv int_{is} - ext_{is} = \{\alpha_s + \phi_s \kappa_s\} A_{is} + \{\gamma_s + \phi_s \delta_s + \tau_{js}\} F_i + \zeta_{is}, \tag{3.3}$$

where ζ_{is} incorporates ε_{is}^k and υ_{is}^k for $k \in \{int, ext\}$. Equation (3.3) provides a basic setting to understand why a gender gap in d_i might arise. In particular, we observe that, conditional on

[5] Their inclusion in the model complicates the presentation, without providing further insights to the exposition.

[6] Zamarro *et al.* (2019), for instance, explore different proxy measures of effort in PISA by analyzing students' response patterns to the different test items. They find that these proxies explain between 32-38% of total unobserved variation in PISA scores across countries.

A_{is}, a gender gap can be attributed to three different features: **(a)** systematic distinct gender aptitudes that vary with the test taking environment (γ_s), **(b)** the responsiveness of effort to changes in the stakes that vary with F_i ($\phi_s \delta_s$) and **(c)** the grading attitudes of teacher j towards characteristic F_i (τ_{js}).

In what follows, we begin by studying the existence of gaps between int and ext for several observable student traits (i.e. gender, national origin and family income). To shed light on the mechanism, we then investigate whether disparities in d_{is} can be attributed to τ_{js}. This is achieved by studying the presence of variation in d across subjects. Section 3.7 presents the discussion and results of the analysis.

3.3 Data

We combine administrative data from two separate sources in order to perform the empirical analysis:

Data from the Department of Education of the Basque Government (*Hezkuntza Saila*).– It contains the enrollment records of the student population in public and semi-public schools. Besides typical enrollment information, this data source also incorporates rich personal information of students' characteristics: their date of birth, gender, country of origin, financial aid eligibility (based on students' household income), and whether the student has special needs. Furthermore, the data also records the students' legal guardians'/parents' country of birth. This information allows us to identify whether students are native, first- or second-generation immigrants.

In addition, the data also include students' records on the end-of-year final grade in each subject from the 2015/16 and 2016/17 academic years. In the Basque Country, each academic year is divided into three quarters for grading purposes. After each quarter, students within the same class are tested in a given subject on the same date. The end-of-year scores, that correspond to teachers' personal assessments, are the average of the three different quarter grades, per subject. These *internal* assessment scores adopt an integer-based scale, with values ranging between 0 and 10.

An unfortunate limitation of this data source is that student-teacher identifiers are unavailable: student enrollment information is limited to which school and grade the student is enrolled in, as well as the main language of instruction or "language model".[7] Although language model sorting gives *some* information on student-teacher assignment, the lack of true student-teacher identifiers in the data sources implies that we cannot delve deeper into some of the teacher-specific mechanisms behind the assessment gaps.

Data on external assessments from the Basque Institute for Research and Evaluation in Education (ISEI-IVEI).– The ISEI-IVEI provided us with data on the external assessments they have been carrying out to Basque students since 2009. The assessments were designed and carried out every two years to students in 4^{th} grade of primary school, and 2^{nd} year of middle school (the equivalent of 8^{th} grade in the US school system) of public and semi-public schools.

These external assessments, also called diagnostic evaluations, are low-stake tests that do not affect students' progress but are intended to provide information on the performance of the Basque educational system to the Department of Education of the Basque Government. Students' competences are assessed in Math, Science, English, Spanish and Basque. They complete the tests in their usual school, and hence there are no major alterations of the test-taking environment when compared to the internal assessments. In theory, all students enrolled in the mentioned grades at the time of the examination are required to take all five tests. However, if a student is sick or otherwise absent from school on the day of the test, they are not evaluated on another day. For the purposes of our analysis, we consider only students who take all five tests. These constitute 84.5% (82.2%) of primary (middle school) students in our sample.

Besides variables on the results obtained in each of the five external assessments for each student, the database from the ISEI-IVEI also includes other relevant information such as students' socioeconomic status and the classroom each student belongs to.

[7] Students in the Basque Country may receive lessons in non-language subjects such as Maths and Science in Spanish or Basque, or some subjects in Basque and some others in Spanish. This peculiarity of the region results in students being sorted into so-called "language models", which can be either model A (Spanish), B (Spanish and Basque), or D (Basque) depending on the language of instruction used for non-language subjects.

Student-level identifiers present on both data sources allow us to merge the two. Starting from the raw database, we implement a sequence of data restrictions necessary for the empirical analysis. First, since we need availability of both external and internal assessment scores, we restrict our sample to students enrolled in 4^{th} grade of primary school and 2^{nd} grade of secondary school during the 2016/17 academic year.

Second, we drop class groups with less than 15 or more than 32 students. Group sizes outside those ranges are unlikely, especially those larger than 32 students, since class sizes in the Basque Country are capped at 30 students with few exceptions. Thus, we believe that it is much more probable that these pupils are mismatched due to errors in the string-based classroom identifiers, which would result in unreliable estimates.[8]

Using this sample, we construct the standardized test scores for both the external and internal grades for each student in each of the five subjects. Finally, we exclude the students whose grades are not available for every subject or have missing data about their personal characteristics (26.9% of observations). By imposing such a restriction, we hold the estimation sample constant throughout the analysis.[9] Therefore, our findings are comparable, and the observed differences across subjects cannot be attributed to specification or sample adjustments. As a result, the final sample consists of 31,183 observations, distributed between primary (15,802 pupils) and middle school (15,381 students).

Table 3.1 shows basic descriptive statistics of the main student characteristics used in the empirical analysis. Approximately 50% of the students in the sample are female, both in primary and middle school. Special need students account for 4.79% of the sample in primary school, while this number is slightly lower (3.65%) in middle school. The percentage of students that have suffered grade retention is around 1.6% and 3% respectively.

Data on students' parents' country of origin allows us to classify students as Native, 1^{st} generation immigrants, or 2^{nd} generation immigrants using dummy-coded variables. Around 5% of primary-school pupils in the sample do not have information on their parents' country of origin and are thus classified as unspecified. This number is lower, approximately 2.6%, for students in the middle school sample. Information on financial aid eligibility, which is based

[8]Based on this consideration, we drop 14.4% of the classes: 18.0% in primary and 10.4% in middle school.

[9]This sample restriction is inconsequential for the stability of the results. The observed findings barely change when we exploit the entire sample.

Table 3.1 Summary Statistics

	Panel A: Primary School			*Panel B:* Middle School		
VARIABLES	*N*	Mean	SD	*N*	Mean	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Female	15,802	0.491	0.500	15,381	0.501	0.500
Special Needs	15,802	0.0479	0.214	15,381	0.0365	0.187
Grade Retention	15,802	0.0162	0.126	15,381	0.0302	0.171
Home language (Basque)	15,802	0.259	0.438	15,381	0.205	0.403
Month of Birth	15,802	0.497	0.311	15,381	0.498	0.312
Native	15,802	0.878	0.328	15,381	0.920	0.272
Immigrant (1st gen)	15,802	0.0266	0.161	15,381	0.0413	0.199
Immigrant (2nd gen)	15,802	0.0418	0.200	15,381	0.0134	0.115
Undefined	15,802	0.0540	0.226	15,381	0.0256	0.158
Income = 1 (poorest)	15,802	0.209	0.406	15,381	0.166	0.372
Income = 2	15,802	0.0839	0.277	15,381	0.0838	0.277
Income = 3	15,802	0.128	0.334	15,381	0.130	0.336
Income = 4 (richest)	15,802	0.580	0.494	15,381	0.620	0.485
Lang Model = A (Spanish)	15,802	0.0415	0.199	15,381	0.0573	0.233
Lang Model = B (Bilingual)	15,802	0.227	0.419	15,381	0.236	0.425
Lang Model = D (Basque)	15,802	0.731	0.443	15,381	0.707	0.455

Notes: Grade Retention = 1 if the student suffered from grade retention previously. Month of birth is *linearized* between 0 and 1 (i.e. = 1 if born in January and = 0 in December). The set of dummy variables for income groups are defined using students' eligibility to different scholarships. The specific eligibility thresholds depend on household size and total household income. For instance, for a family of 4 people, the following thresholds are defined: Income = 1 ($< 17,567$€), Income = 2 ($17,567 - 22,820$€), Income = 3 ($22,820 - 34,352$€), Income = 4 ($> 34,352$€).

on household income thresholds, enables us to sort students into four different income groups (from low to high income), also dummy-coded.

Lastly, students are sorted, as mentioned, according to the language model or language of instruction. Note that the vast majority of students (more than 70% in both primary and middle school) are enrolled in the D (Basque) language model; although only around 20-25% of students in the sample speak Basque at home.

3.4 Empirical Strategy

We now turn to empirically evaluate the differences between int and ext. Our analysis departs from the relationship established by Equation (3.3). Given that the ability level is unobserved, we *parametrize* A_{is} with ext_{is}. As argued by Burgess & Greaves (2013), this seems like a logical replacement. First, given the quasi-blind nature of external assessments, ext is arguably a less noisy signal of A than int. Second, int only adopts integer values between 0 and 10. Thus, using ext instead of int on the right-hand side provides more precise estimates, given that it employs a continuous scale and displays a higher variability. Altogether, this yields our main empirical specification:

$$d_{is} = \rho ext_{is} + \alpha fem_i + \beta \mathbf{origin}_i + \xi \mathbf{income}_i + \delta graderet_i + \gamma special_i + \chi_{c(i)} + u_{is}, \quad (3.4)$$

where student i attending class c receives scores int_{is} and ext_{is}, respectively, from internal and external evaluations in subject s. In our analysis, we employ two complementary dependent variables to assess the stability of our results. First, we use student i's continuous gap in the z-scores of int relative to ext in subject s. Second, we examine $P(int_{is} > ext_{is})$ with a Linear Probability Model. Regressors include a dummy for gender (fem_i), a vector of indicators for student origin ($\mathbf{origin}_i$, with dummies for first- or second-generation immigrant or unspecified origin), a vector of dummies with student's household income ($\mathbf{income}_i$), an indicator on whether the student suffered from grade retention ($graderet_i$), and another indicator variable that defines whether the student has special needs ($special_i$). Finally, the specification also includes class fixed effects ($\chi_{c(i)}$).

For presentation purposes, we display separate results for primary and middle school for each subject. To assess the robustness of our findings, we expand the empirical specifications by including other personal characteristics. More specifically, we add a set of dummies that indicates the tercile of Socioeconomic and Cultural Index, ISEC, of the student's family and an indicator on whether the language spoken at home with their family is Basque. Standard errors are clustered at the class level to allow for unobserved correlation across students attending the same class.

With the inclusion class fixed effects and by exploiting students' variation within a classroom, we alleviate some relevant concerns regarding the nonrandom allocation of students to schools and sorting of pupils into classes and teachers. First, the final assignment of students to public and semi-public schools prioritizes the admissions of children from the catchment area of the school.[10] Consequently, schools frequently echo neighborhood or municipality characteristics. Second, the education system is organized around three different models based on the instructional language. As a bilingual region, parents have the possibility of choosing their most preferred language model; Basque or Spanish, or a mix of both as their child's language of instruction. Thus, the presence of these different schemes prevents the random allocation of students across classes in the schools that offer several linguistic models. Third, teacher-student assignment might also be nonrandom. There is substantial evidence that teachers are not randomly assigned to students, not even within schools [Clotfelter *et al.* (2005); Aaronson *et al.* (2007); Rothstein (2010); Jackson (2014); Qureshi & Ost (2020)]. Fourth, teachers distinctively grade their students as a result of the average performance of the class. In particular, Calsamiglia & Loviglio (2019) show that students get penalized if they are allocated to a class with better peers. Finally, the existence of different attitudes towards grading might also bias the results. While some teachers might be more compassionate, others may set stricter rules and tend to *underassess* their students.

3.5 Results

Tables 3.2 through 3.6 show the main results of the regression analysis discussed in the empirical specification. As mentioned, the main analysis focuses on the link between internal and external assessments of students from 4^{th} year of primary education (9 years old) and 2^{nd} year of middle school (13 years old).

Table 3.2 shows the results of Equation (3.4) with the two complementary dependent variables and additional personal characteristics, separately for primary and middle school, for the Maths subject. While the left panel displays the results from regressions that use the continuous difference between int and ext as the dependent variable, the right panel employs a

[10]Prior to the academic year 2018-19, students were allocated based on a centralized system that employed the Parallel Mechanism. In January 2018, the Basque government reformed the assignment system and moved towards a Deferred Acceptance (DA) mechanism.

LPM specification. In what follows, we will focus on discussing the results from the right panel. The findings do not significantly change with the use of LPM or the inclusion of additional explanatory variables. Tables 3.3, 3.4, 3.5 and 3.6 show the analogous results for the remaining subjects (Science, Basque, Spanish and English, respectively).

Overall, our results suggest that female students obtain significantly higher scores on internal assessments compared to external assessments. This happens for both primary and middle school students, and for all subjects, with the exception of Math in primary school. Results are robust to both considerations of the dependent variable and the inclusion of additional covariates. Interestingly, the effects are not only larger in language skills, but also in middle school for all subjects. For example, being female increases the score obtained in the internal assessment of Science by around 0.10σ in primary education, and by around 0.28σ in middle school. In contrast, the analogous effect is 0.275σ in Spanish for primary school, but it is 0.385σ in middle school. These differences are highly significant in statistical terms ($p < .01$) for the remaining subjects.

Next, the evidence indicates that there are strong negative effects ($p < .01$) for first-generation immigrant students in primary school (between -0.174σ and -0.279σ, depending on the subject, based on the specification from column 3). There is also a strong but less negative effect for second generation immigrants in all subjects (between -0.113σ and -0.143σ), with the exception of English (-0.032σ). In middle school, the negative effects for first-generation students are also observed for all subjects, whereas this is not the case for second-generation immigrants.

Students from families with a higher income level show strong positive effects both in primary and middle school. These effects are highly significant in statistical terms ($p < .01$) and increase monotonically with income. For instance, for primary students in Basque the effect is 0.126σ for the second lowest income group, whereas it is 0.245σ for the highest one. These observed differences remain across all subjects and the two dependent variable specifications, and are robust to the inclusion of additional student characteristics.

Both special needs students and those who have previously suffered from grade retention obtain significantly lower scores in the internal assessments relative to the external ones. This negative effect is observed in all the regressions of the main analysis and it is significant at

Table 3.2 Main Results - Math

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.438***	-0.455***	-0.455***	-0.448***	-0.466***	-0.469***	-0.194***	-0.202***	-0.202***	-0.211***	-0.219***	-0.220***
	(0.00714)	(0.00721)	(0.00722)	(0.00762)	(0.00762)	(0.00761)	(0.00403)	(0.00407)	(0.00408)	(0.00446)	(0.00444)	(0.00445)
Female	0.0133	0.00621	0.00654	0.289***	0.281***	0.281***	-0.000774	-0.00412	-0.00395	0.132***	0.129***	0.129***
	(0.0129)	(0.0128)	(0.0128)	(0.0137)	(0.0137)	(0.0136)	(0.00800)	(0.00798)	(0.00798)	(0.00793)	(0.00790)	(0.00788)
Immigrant (1st gen)	-0.245***	-0.210***	-0.200***	-0.125***	-0.0868**	-0.0716**	-0.0831***	-0.0666***	-0.0617**	-0.105***	-0.0879***	-0.0802***
	(0.0420)	(0.0414)	(0.0415)	(0.0362)	(0.0359)	(0.0359)	(0.0245)	(0.0245)	(0.0245)	(0.0203)	(0.0202)	(0.0202)
Immigrant (2nd gen)	-0.166***	-0.122***	-0.113***	-0.0479	0.00100	0.0102	-0.0978***	-0.0770***	-0.0728***	-0.0488	-0.0269	-0.0221
	(0.0361)	(0.0362)	(0.0361)	(0.0664)	(0.0661)	(0.0660)	(0.0204)	(0.0204)	(0.0204)	(0.0349)	(0.0349)	(0.0348)
Undefined	-0.137***	-0.116***	-0.113***	-0.0286	-0.0155	-0.0106	-0.0640***	-0.0541***	-0.0526***	0.00573	0.0116	0.0141
	(0.0273)	(0.0271)	(0.0269)	(0.0462)	(0.0453)	(0.0451)	(0.0177)	(0.0176)	(0.0175)	(0.0284)	(0.0280)	(0.0277)
Income = 2	0.163***	0.142***	0.141***	0.114***	0.0983***	0.0973***	0.0655***	0.0558***	0.0550***	0.0410**	0.0341**	0.0336**
	(0.0265)	(0.0264)	(0.0264)	(0.0277)	(0.0275)	(0.0275)	(0.0165)	(0.0165)	(0.0165)	(0.0161)	(0.0160)	(0.0160)
Income = 3	0.251***	0.211***	0.208***	0.149***	0.113***	0.110***	0.107***	0.0883***	0.0870***	0.0685***	0.0523***	0.0506***
	(0.0240)	(0.0237)	(0.0237)	(0.0238)	(0.0239)	(0.0240)	(0.0147)	(0.0147)	(0.0147)	(0.0136)	(0.0136)	(0.0136)
Income = 4	0.361***	0.293***	0.285***	0.213***	0.152***	0.142***	0.157***	0.125***	0.121***	0.0953***	0.0682***	0.0634***
	(0.0193)	(0.0192)	(0.0192)	(0.0195)	(0.0203)	(0.0203)	(0.0116)	(0.0118)	(0.0118)	(0.0111)	(0.0113)	(0.0113)
Special Needs	-0.496***	-0.490***	-0.489***	-0.202***	-0.192***	-0.189***	-0.205***	-0.201***	-0.201***	-0.0931***	-0.0885***	-0.0873***
	(0.0313)	(0.0312)	(0.0313)	(0.0348)	(0.0346)	(0.0347)	(0.0177)	(0.0176)	(0.0176)	(0.0211)	(0.0210)	(0.0211)
Grade Retention	-0.334***	-0.305***	-0.302***	-0.463***	-0.437***	-0.433***	-0.184***	-0.170***	-0.169***	-0.193***	-0.181***	-0.179***
	(0.0469)	(0.0476)	(0.0476)	(0.0429)	(0.0430)	(0.0429)	(0.0283)	(0.0285)	(0.0285)	(0.0232)	(0.0233)	(0.0232)
2nd Tertile (ISEC)		0.165***	0.163***		0.117***	0.112***		0.0745***	0.0731***		0.0549***	0.0524***
		(0.0134)	(0.0134)		(0.0160)	(0.0161)		(0.00847)	(0.00846)		(0.00889)	(0.00890)
3rd Tertile (ISEC)		0.219***	0.215***		0.215***	0.206***		0.104***	0.102***		0.0946***	0.0902***
		(0.0145)	(0.0145)		(0.0167)	(0.0168)		(0.00930)	(0.00930)		(0.00973)	(0.00978)
Home language (Basque)			0.0697***			0.131***			0.0348***			0.0669***
			(0.0176)			(0.0196)			(0.0108)			(0.0122)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.427	0.436	0.437	0.414	0.422	0.424	0.291	0.298	0.299	0.301	0.306	0.307
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator (= 1if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1.

Table 3.3 Main Results - Science

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.529***	-0.550***	-0.551***	-0.527***	-0.548***	-0.551***	-0.221***	-0.230***	-0.230***	-0.227***	-0.236***	-0.237***
	(0.00788)	(0.00781)	(0.00783)	(0.00801)	(0.00801)	(0.00800)	(0.00428)	(0.00429)	(0.00431)	(0.00422)	(0.00428)	(0.00430)
Female	0.105***	0.102***	0.102***	0.285***	0.278***	0.279***	0.0380***	0.0369***	0.0371***	0.123***	0.120***	0.121***
	(0.0133)	(0.0132)	(0.0132)	(0.0143)	(0.0141)	(0.0140)	(0.00755)	(0.00751)	(0.00753)	(0.00769)	(0.00769)	(0.00767)
Immigrant (1st gen)	-0.310***	-0.271***	-0.260***	-0.155***	-0.110***	-0.0915**	-0.127***	-0.112***	-0.107***	-0.0592***	-0.0419**	-0.0355*
	(0.0438)	(0.0434)	(0.0436)	(0.0395)	(0.0393)	(0.0392)	(0.0236)	(0.0235)	(0.0236)	(0.0208)	(0.0209)	(0.0209)
Immigrant (2nd gen)	-0.201***	-0.153***	-0.143***	-0.0977	-0.0394	-0.0279	-0.114***	-0.0956***	-0.0919***	-0.0408	-0.0184	-0.0145
	(0.0378)	(0.0376)	(0.0377)	(0.0706)	(0.0698)	(0.0696)	(0.0205)	(0.0206)	(0.0206)	(0.0337)	(0.0332)	(0.0332)
Undefined	-0.136***	-0.113***	-0.109***	-0.0259	-0.0103	-0.00427	-0.0457**	-0.0367**	-0.0354**	0.0212	0.0272	0.0293
	(0.0282)	(0.0274)	(0.0272)	(0.0416)	(0.0405)	(0.0404)	(0.0178)	(0.0178)	(0.0178)	(0.0263)	(0.0261)	(0.0262)
Income = 2	0.199***	0.176***	0.174***	0.106***	0.0881***	0.0869***	0.0661***	0.0572***	0.0566***	0.0447***	0.0378**	0.0374**
	(0.0265)	(0.0263)	(0.0264)	(0.0286)	(0.0282)	(0.0282)	(0.0151)	(0.0150)	(0.0150)	(0.0156)	(0.0156)	(0.0155)
Income = 3	0.301***	0.256***	0.253***	0.147***	0.105***	0.101***	0.127***	0.109***	0.108***	0.0620***	0.0455***	0.0442***
	(0.0237)	(0.0234)	(0.0235)	(0.0246)	(0.0245)	(0.0245)	(0.0136)	(0.0136)	(0.0136)	(0.0138)	(0.0138)	(0.0139)
Income = 4	0.404***	0.325***	0.316***	0.235***	0.163***	0.151***	0.176***	0.145***	0.141***	0.0970***	0.0687***	0.0647***
	(0.0194)	(0.0193)	(0.0196)	(0.0203)	(0.0211)	(0.0211)	(0.0111)	(0.0110)	(0.0112)	(0.0112)	(0.0116)	(0.0116)
Special Needs	-0.481***	-0.472***	-0.472***	-0.366***	-0.353***	-0.350***	-0.171***	-0.168***	-0.168***	-0.142***	-0.137***	-0.136***
	(0.0319)	(0.0317)	(0.0318)	(0.0389)	(0.0386)	(0.0387)	(0.0181)	(0.0181)	(0.0181)	(0.0197)	(0.0195)	(0.0196)
Grade Retention	-0.579***	-0.545***	-0.542***	-0.670***	-0.641***	-0.636***	-0.239***	-0.226***	-0.225***	-0.256***	-0.245***	-0.243***
	(0.0514)	(0.0514)	(0.0513)	(0.0460)	(0.0458)	(0.0457)	(0.0281)	(0.0280)	(0.0281)	(0.0207)	(0.0207)	(0.0207)
2nd Tertile (ISEC)		0.165***	0.162***		0.131***	0.125***		0.0619***	0.0607***		0.0483***	0.0462***
		(0.0148)	(0.0148)		(0.0163)	(0.0163)		(0.00877)	(0.00879)		(0.00902)	(0.00903)
3rd Tertile (ISEC)		0.261***	0.257***		0.253***	0.242***		0.102***	0.100***		0.0991***	0.0955***
		(0.0161)	(0.0161)		(0.0170)	(0.0170)		(0.00946)	(0.00946)		(0.00967)	(0.00965)
Home language (Basque)			0.0802***			0.160***			0.0316***			0.0542***
			(0.0168)			(0.0197)			(0.0104)			(0.0112)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.498	0.509	0.510	0.475	0.484	0.487	0.338	0.344	0.344	0.340	0.345	0.347
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator ($=$ 1if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** $p<.01$, ** $p<.05$, * $p<.1$.

Table 3.4 Main Results - Basque

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.410***	-0.428***	-0.435***	-0.443***	-0.462***	-0.472***	-0.191***	-0.199***	-0.203***	-0.213***	-0.221***	-0.225***
	(0.00758)	(0.00764)	(0.00764)	(0.00752)	(0.00755)	(0.00760)	(0.00457)	(0.00460)	(0.00464)	(0.00457)	(0.00456)	(0.00460)
Female	0.180***	0.178***	0.181***	0.325***	0.324***	0.327***	0.0818***	0.0811***	0.0822***	0.148***	0.147***	0.149***
	(0.0115)	(0.0115)	(0.0114)	(0.0133)	(0.0131)	(0.0131)	(0.00740)	(0.00740)	(0.00736)	(0.00789)	(0.00784)	(0.00783)
Immigrant (1st gen)	-0.305***	-0.274***	-0.245***	-0.213***	-0.177***	-0.149***	-0.153***	-0.138***	-0.124***	-0.0719***	-0.0578***	-0.0451**
	(0.0393)	(0.0381)	(0.0383)	(0.0359)	(0.0357)	(0.0355)	(0.0248)	(0.0244)	(0.0244)	(0.0202)	(0.0200)	(0.0198)
Immigrant (2nd gen)	-0.237***	-0.198***	-0.173***	-0.149**	-0.102	-0.0846	-0.137***	-0.118***	-0.106***	-0.0541	-0.0360	-0.0280
	(0.0349)	(0.0351)	(0.0349)	(0.0640)	(0.0634)	(0.0637)	(0.0206)	(0.0206)	(0.0205)	(0.0336)	(0.0335)	(0.0336)
Undefined	-0.135***	-0.115***	-0.107***	-0.0281	-0.0160	-0.00690	-0.0603***	-0.0511***	-0.0468***	0.000385	0.00507	0.00920
	(0.0264)	(0.0260)	(0.0254)	(0.0404)	(0.0406)	(0.0390)	(0.0178)	(0.0177)	(0.0176)	(0.0212)	(0.0212)	(0.0205)
Income = 2	0.149***	0.130***	0.126***	0.0671***	0.0529**	0.0515**	0.0615***	0.0526***	0.0506***	0.0242	0.0186	0.0180
	(0.0230)	(0.0230)	(0.0231)	(0.0258)	(0.0253)	(0.0252)	(0.0156)	(0.0156)	(0.0157)	(0.0157)	(0.0156)	(0.0156)
Income = 3	0.217***	0.181***	0.174***	0.166***	0.132***	0.126***	0.115***	0.0977***	0.0943***	0.0588***	0.0454***	0.0428***
	(0.0211)	(0.0210)	(0.0211)	(0.0231)	(0.0232)	(0.0232)	(0.0139)	(0.0139)	(0.0139)	(0.0137)	(0.0138)	(0.0137)
Income = 4	0.330***	0.268***	0.245***	0.213***	0.156***	0.139***	0.147***	0.117***	0.106***	0.0855***	0.0626***	0.0548***
	(0.0186)	(0.0183)	(0.0182)	(0.0181)	(0.0187)	(0.0186)	(0.0114)	(0.0115)	(0.0116)	(0.0108)	(0.0112)	(0.0112)
Special Needs	-0.498***	-0.492***	-0.492***	-0.207***	-0.198***	-0.196***	-0.229***	-0.226***	-0.226***	-0.0864***	-0.0833***	-0.0824***
	(0.0303)	(0.0302)	(0.0304)	(0.0340)	(0.0341)	(0.0341)	(0.0178)	(0.0178)	(0.0179)	(0.0203)	(0.0204)	(0.0204)
Grade Retention	-0.440***	-0.414***	-0.407***	-0.589***	-0.566***	-0.559***	-0.240***	-0.227***	-0.224***	-0.233***	-0.224***	-0.221***
	(0.0388)	(0.0384)	(0.0381)	(0.0456)	(0.0453)	(0.0450)	(0.0291)	(0.0288)	(0.0287)	(0.0210)	(0.0211)	(0.0209)
2nd Tertile (ISEC)		0.132***	0.124***		0.114***	0.106***		0.0651***	0.0616***		0.0402***	0.0365***
		(0.0136)	(0.0136)		(0.0150)	(0.0149)		(0.00886)	(0.00883)		(0.00901)	(0.00901)
3rd Tertile (ISEC)		0.211***	0.201***		0.203***	0.188***		0.0992***	0.0940***		0.0821***	0.0756***
		(0.0146)	(0.0144)		(0.0158)	(0.0157)		(0.00926)	(0.00920)		(0.00929)	(0.00932)
Home language (Basque)			0.212***			0.248***			0.102***			0.112***
			(0.0175)			(0.0184)			(0.0106)			(0.0119)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.471	0.479	0.486	0.451	0.458	0.465	0.325	0.331	0.335	0.331	0.334	0.339
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator (= 1if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1.

Table 3.5 Main Results - Spanish

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.491***	-0.505***	-0.505***	-0.550***	-0.570***	-0.569***	-0.210***	-0.217***	-0.217***	-0.238***	-0.246***	-0.246***
	(0.00750)	(0.00749)	(0.00750)	(0.00710)	(0.00708)	(0.00707)	(0.00429)	(0.00426)	(0.00427)	(0.00373)	(0.00372)	(0.00372)
Female	0.275***	0.269***	0.269***	0.385***	0.380***	0.380***	0.116***	0.113***	0.113***	0.153***	0.151***	0.151***
	(0.0130)	(0.0130)	(0.0130)	(0.0136)	(0.0135)	(0.0135)	(0.00753)	(0.00749)	(0.00749)	(0.00761)	(0.00757)	(0.00757)
Immigrant (1st gen)	-0.316***	-0.279***	-0.279***	-0.118***	-0.0729**	-0.0571	-0.128***	-0.112***	-0.113***	-0.0592***	-0.0413**	-0.0347*
	(0.0398)	(0.0394)	(0.0395)	(0.0371)	(0.0370)	(0.0370)	(0.0245)	(0.0244)	(0.0244)	(0.0195)	(0.0194)	(0.0194)
Immigrant (2nd gen)	-0.188***	-0.144***	-0.144***	-0.0133	0.0399	0.0491	-0.0863***	-0.0668***	-0.0676***	-0.0305	-0.00923	-0.00536
	(0.0351)	(0.0352)	(0.0351)	(0.0622)	(0.0612)	(0.0611)	(0.0211)	(0.0211)	(0.0211)	(0.0350)	(0.0346)	(0.0345)
Undefined	-0.120***	-0.0969***	-0.0969***	-0.0920**	-0.0775*	-0.0726*	-0.0607***	-0.0505***	-0.0508***	-0.0136	-0.00785	-0.00580
	(0.0270)	(0.0266)	(0.0266)	(0.0400)	(0.0396)	(0.0394)	(0.0165)	(0.0161)	(0.0162)	(0.0242)	(0.0238)	(0.0237)
Income = 2	0.183***	0.162***	0.162***	0.118***	0.0998***	0.0988***	0.0715***	0.0619***	0.0621***	0.0583***	0.0510***	0.0506***
	(0.0261)	(0.0261)	(0.0261)	(0.0274)	(0.0271)	(0.0270)	(0.0149)	(0.0149)	(0.0149)	(0.0150)	(0.0150)	(0.0150)
Income = 3	0.241***	0.201***	0.201***	0.158***	0.116***	0.113***	0.117***	0.0985***	0.0987***	0.0599***	0.0434***	0.0420***
	(0.0235)	(0.0235)	(0.0236)	(0.0255)	(0.0255)	(0.0255)	(0.0137)	(0.0137)	(0.0137)	(0.0138)	(0.0139)	(0.0139)
Income = 4	0.330***	0.261***	0.260***	0.252***	0.182***	0.172***	0.140***	0.109***	0.110***	0.100***	0.0724***	0.0682***
	(0.0193)	(0.0194)	(0.0195)	(0.0206)	(0.0211)	(0.0211)	(0.0112)	(0.0115)	(0.0116)	(0.0106)	(0.0111)	(0.0111)
Special Needs	-0.536***	-0.528***	-0.528***	-0.255***	-0.244***	-0.239***	-0.219***	-0.215***	-0.215***	-0.0861***	-0.0815***	-0.0796***
	(0.0323)	(0.0322)	(0.0323)	(0.0347)	(0.0345)	(0.0346)	(0.0175)	(0.0175)	(0.0175)	(0.0188)	(0.0187)	(0.0187)
Grade Retention	-0.460***	-0.428***	-0.428***	-0.629***	-0.599***	-0.594***	-0.195***	-0.181***	-0.181***	-0.218***	-0.206***	-0.204***
	(0.0475)	(0.0470)	(0.0470)	(0.0445)	(0.0442)	(0.0443)	(0.0267)	(0.0267)	(0.0267)	(0.0211)	(0.0212)	(0.0212)
2nd Tertile (ISEC)		0.147***	0.147***		0.124***	0.119***		0.0681***	0.0685***		0.0509***	0.0485***
		(0.0142)	(0.0142)		(0.0156)	(0.0156)		(0.00900)	(0.00903)		(0.00872)	(0.00874)
3rd Tertile (ISEC)		0.222***	0.222***		0.240***	0.230***		0.0981***	0.0986***		0.0952***	0.0911***
		(0.0149)	(0.0149)		(0.0166)	(0.0166)		(0.00926)	(0.00927)		(0.00915)	(0.00918)
Home language (Basque)			0.000900			0.130***			-0.00683			0.0546***
			(0.0170)			(0.0207)			(0.0102)			(0.0113)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.488	0.496	0.496	0.482	0.491	0.493	0.340	0.346	0.346	0.346	0.352	0.353
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator (= 1if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1.

Table 3.6 Main Results - English

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.380***	-0.395***	-0.395***	-0.318***	-0.333***	-0.333***	-0.159***	-0.167***	-0.167***	-0.183***	-0.189***	-0.189***
	(0.00793)	(0.00797)	(0.00798)	(0.00723)	(0.00727)	(0.00727)	(0.00469)	(0.00466)	(0.00467)	(0.00493)	(0.00491)	(0.00490)
Female	0.151***	0.149***	0.149***	0.224***	0.223***	0.223***	0.0740***	0.0728***	0.0730***	0.120***	0.120***	0.120***
	(0.0120)	(0.0120)	(0.0120)	(0.0115)	(0.0115)	(0.0115)	(0.00770)	(0.00768)	(0.00768)	(0.00781)	(0.00778)	(0.00777)
Immigrant (1st gen)	-0.215***	-0.182***	-0.174***	-0.136***	-0.107***	-0.0918***	-0.0774***	-0.0606***	-0.0572**	-0.0783***	-0.0667***	-0.0578***
	(0.0406)	(0.0398)	(0.0397)	(0.0337)	(0.0336)	(0.0336)	(0.0238)	(0.0234)	(0.0234)	(0.0213)	(0.0213)	(0.0214)
Immigrant (2nd gen)	-0.0782**	-0.0393	-0.0324	-0.156***	-0.119**	-0.110**	-0.0359*	-0.0160	-0.0130	-0.0743**	-0.0594*	-0.0542
	(0.0359)	(0.0360)	(0.0360)	(0.0531)	(0.0528)	(0.0529)	(0.0217)	(0.0219)	(0.0219)	(0.0343)	(0.0343)	(0.0343)
Undefined	-0.0524*	-0.0339	-0.0314	-0.0154	-0.00595	-0.00123	-0.0374**	-0.0279	-0.0268	-0.00646	-0.00272	4.94e-05
	(0.0274)	(0.0270)	(0.0270)	(0.0366)	(0.0367)	(0.0366)	(0.0185)	(0.0184)	(0.0184)	(0.0301)	(0.0301)	(0.0299)
Income = 2	0.188***	0.170***	0.169***	0.0836***	0.0731***	0.0720***	0.101***	0.0925***	0.0920***	0.0358**	0.0315*	0.0309*
	(0.0227)	(0.0226)	(0.0227)	(0.0244)	(0.0242)	(0.0241)	(0.0151)	(0.0151)	(0.0151)	(0.0164)	(0.0164)	(0.0164)
Income = 3	0.240***	0.205***	0.203***	0.154***	0.129***	0.125***	0.116***	0.0990***	0.0981***	0.0711***	0.0610***	0.0591***
	(0.0213)	(0.0212)	(0.0212)	(0.0216)	(0.0217)	(0.0217)	(0.0137)	(0.0137)	(0.0137)	(0.0145)	(0.0147)	(0.0147)
Income = 4	0.299***	0.243***	0.237***	0.198***	0.157***	0.148***	0.133***	0.104***	0.101***	0.0885***	0.0718***	0.0662***
	(0.0181)	(0.0182)	(0.0182)	(0.0170)	(0.0176)	(0.0177)	(0.0115)	(0.0118)	(0.0118)	(0.0113)	(0.0117)	(0.0117)
Special Needs	-0.543***	-0.537***	-0.536***	-0.263***	-0.257***	-0.253***	-0.245***	-0.242***	-0.242***	-0.143***	-0.140***	-0.138***
	(0.0293)	(0.0291)	(0.0292)	(0.0333)	(0.0332)	(0.0332)	(0.0179)	(0.0180)	(0.0180)	(0.0218)	(0.0219)	(0.0218)
Grade Retention	-0.324***	-0.300***	-0.297***	-0.456***	-0.440***	-0.435***	-0.180***	-0.168***	-0.167***	-0.214***	-0.208***	-0.205***
	(0.0439)	(0.0439)	(0.0439)	(0.0414)	(0.0416)	(0.0414)	(0.0286)	(0.0286)	(0.0285)	(0.0235)	(0.0236)	(0.0235)
2nd Tertile (ISEC)		0.139***	0.136***		0.0945***	0.0892***		0.0644***	0.0633***		0.0339***	0.0309***
		(0.0139)	(0.0139)		(0.0138)	(0.0138)		(0.00908)	(0.00908)		(0.00942)	(0.00944)
3rd Tertile (ISEC)		0.183***	0.180***		0.147***	0.137***		0.0982***	0.0966***		0.0605***	0.0550***
		(0.0143)	(0.0143)		(0.0142)	(0.0143)		(0.00930)	(0.00930)		(0.00978)	(0.00982)
Home language (Basque)			0.0567***			0.126***			0.0249**			0.0737***
			(0.0167)			(0.0168)			(0.0105)			(0.0120)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.460	0.466	0.467	0.426	0.431	0.433	0.305	0.310	0.311	0.330	0.332	0.334
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator (= 1if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** $p<.01$, ** $p<.05$, * $p<.1$.

$p < .01$. Interestingly, the negative effects for these two student characteristics are the largest ones observed.

The inclusion of the terciles of the Socioeconomic and Cultural Index of the student's family yields results similar to those observed with the income indicators: there are strong positive effects (i.e. higher scores obtained in the internal assessment) with higher terciles of the index.

Lastly, speaking Basque at home has a significantly positive effect for all subjects except for Spanish. The size of the effect is relatively small compared to other student characteristics for Maths, Science, and English. As one would expect, it is more quantitatively relevant for the Basque subject.

3.6 Robustness Checks

First, to assess the robustness of our results, we begin by experimenting with different functional forms and specifications. In Equation (3.4), we rely on the linearity assumption of d with respect to ext. To relax this assumption, we adopt a cubic specification on ext. Results are displayed in Tables 3.13-3.17 in Appendix 3.9. Overall, we find that adopting a more flexible form does not significantly alter the qualitative nor quantitative nature of our findings. For example, the coefficient for female students of primary education in English is 0.145σ (see Column 1 of Table 3.17), while it is 0.151σ with our main specification.

Second, our internal grades have a discrete support but the external assessments follow a continuous grading scheme. As a consequence, the score distributions of both evaluations are not highly comparable. To test that our results do not reflect this discrepancy, we exploit two alternative measures of ext and int. On the one hand, we repeat the analysis by focusing on the GPA instead of looking at each subject independently. By relying on the use of GPA, we increase variability of the dependent variable, and enhance the similarity between both score distributions. We find that the main qualitative nature of the results remains (see Table 3.18). In particular, results show that there exists a significant and positive gender gap, a significant socioeconomic gradient in both education levels, and a negative immigrant gap in primary school. On the other hand, we replicate our previous analysis by using z-scores that use the class-specific mean and standard deviation for both assessments.[11]. After redefining our z-scores using class level statistics, we find that the sizes and directions of the coefficients are very similar for all the subjects and observable traits.

Finally, we estimate regression (3.4) by using an IV specification. By replacing unobserved ability level A with ext in the right-hand side, we have that $cov(ext, u) \neq 0$. To avoid an endogeneity problem, we adopt an instrumental variable approach. As our instrument for external assessments, we use the age of the child at enrollment, which has been widely employed in the literature that exploits external and internal assessments to identify the presence of teacher bias [e.g. Terrier (2020); Calsamiglia & Loviglio (2019)].

The presence of a strict birth date threshold that determines when children can access compulsory education introduces significant age heterogeneity within a classroom.[12] Substantial evidence suggests that the date of birth is a relevant determinant of human capital formation. Using international data, Bedard & Dhuey (2006) find that older students display better academic performance than the youngest students (4-12 percentiles in grade four and 2-9 percentiles in grade eight). They find that these effects are highly persistent and have long-lasting effects in adulthood. In Catalonia (Spain), Calsamiglia & Loviglio (2016) show that the enrollment age significantly affects both internal and external assessments. Interestingly, the impact is homogeneous for students with different socioeconomic status and the effects are statistically significant across the ability distribution. [13] With regards to the functional form of the effect, Fenoll *et al.* (2019) find that the average scores are linear in the month of birth in Italy.

The exclusion restriction of the instrument is not violated insofar as the month of birth, after controlling for the aforementioned covariates, only affects the assessment gap through the level of cognitive ability, as measured by external evaluations. One possible source of concern for our identification strategy is that teachers might discern the initial maturity disparities, and consequently treat students differently based on their age. In opposition to this hypothesis, Elder & Lubotsky (2009) find in the context of the US that younger students suffer from

[11] Results are available upon request. We omit the results in the present document for the sake of brevity.

[12] The date cutoff in Spain is January 1st for everyone. This date is very rigid and not open to parental or school discretion.

[13] In contrast, using Spanish data, Berniell & Estrada (2020) find that college-educated parents adjust their investment in children education, and consequently, mitigate the month-of-birth penalty compared to their non-college-educated peers.

grade retention more (5 percentage points) and have a higher probability of being diagnosed with ADHD or other learning disabilities (3%). In Spain, Calsamiglia & Loviglio (2016) also find that younger students have a higher probability of dropping out. Overall, this indicates that teachers treat students within a classroom as highly homogeneous, and expect analogous performance from them, independently of their month of birth.

Tables 3.19-3.23 present the results for this IV analysis, separately for each subject. In each table, Panel A displays the outcomes of the first stage, and Panel B of the second stage. We observe that age at enrollment has a sizeable effect on the external assessments, both in primary and middle school. Furthermore, we find that the effect is highly significant ($p < .01$) for all subjects in every specification. From the results, we see that month of birth has a persistent, albeit decreasing, impact. For example, in Basque, the estimated effect is 0.341σ in primary (Table 3.21-Column 1) and 0.222σ in middle school (Column 4). The size of the maturity effects are very similar between subjects.

With regards to the second stage, we find that, once instrumented, the effect of ext on d becomes no longer significant for students in middle school. However, the observed gaps for most of the studied traits remain significantly stable. In particular, the female coefficient barely changes. The estimated immigrant bias is smaller in middle school, and its level of significance decreases in every subject. The nature of the findings are nevertheless constant in primary education. Finally, we find that the presence of a significant positive socioeconomic gradient remains, between subjects and for both education levels.

3.7 Discussion of the Mechanism

3.7.1 Estimated bias across the ability distribution

It is possible that our results are primarily driven by the presence of bias in some explicit parts of the ability distribution. Thus, one relevant question is whether the estimated gaps are constant across the support of this distribution. To tackle this concern, we compute the quartiles of ability using the external assessments of each subject. We then estimate the gender, origin and socioeconomic bias for every quartile separately using our original specification.

Table 3.7 Estimated gender bias by quartiles of ability distribution and subject

	Int - Ext					Prob(Int >Ext)				
	Math	Science	Basque	Spanish	English	Math	Science	Basque	Spanish	English
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Primary school										
1st quartile	0.0190	0.0928***	0.148***	0.271***	0.163***	0.00407	0.0228*	0.0656***	0.0855***	0.0753***
	(0.0242)	(0.0254)	(0.0223)	(0.0229)	(0.0238)	(0.0139)	(0.0132)	(0.0137)	(0.0124)	(0.0139)
2nd quartile	0.0440*	0.109***	0.193***	0.270***	0.155***	0.00983	0.0554***	0.0892***	0.121***	0.0806***
	(0.0242)	(0.0245)	(0.0224)	(0.0247)	(0.0232)	(0.0153)	(0.0155)	(0.0148)	(0.0149)	(0.0151)
3rd quartile	0.0224	0.116***	0.188***	0.253***	0.160***	0.00493	0.0461***	0.108***	0.147***	0.0968***
	(0.0236)	(0.0247)	(0.0227)	(0.0245)	(0.0227)	(0.0156)	(0.0147)	(0.0158)	(0.0158)	(0.0155)
4rd quartile	-0.0601***	0.0899***	0.188***	0.284***	0.107***	-0.0355**	0.0246*	0.0601***	0.0987***	0.0320**
	(0.0224)	(0.0234)	(0.0208)	(0.0233)	(0.0195)	(0.0139)	(0.0132)	(0.0145)	(0.0136)	(0.0138)
N	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802
R^2	0.440	0.511	0.487	0.497	0.476	0.306	0.346	0.341	0.354	0.319
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel B: Middle school										
1st quartile	0.234***	0.223***	0.246***	0.350***	0.165***	0.0608***	0.0569***	0.0733***	0.0795***	0.0773***
	(0.0272)	(0.0269)	(0.0249)	(0.0237)	(0.0227)	(0.0134)	(0.0131)	(0.0121)	(0.0121)	(0.0129)
2nd quartile	0.307***	0.304***	0.294***	0.394***	0.218***	0.148***	0.135***	0.158***	0.194***	0.135***
	(0.0243)	(0.0257)	(0.0245)	(0.0243)	(0.0221)	(0.0157)	(0.0151)	(0.0160)	(0.0161)	(0.0154)
3rd quartile	0.308***	0.325***	0.409***	0.413***	0.266***	0.176***	0.189***	0.218***	0.211***	0.153***
	(0.0249)	(0.0257)	(0.0242)	(0.0269)	(0.0214)	(0.0158)	(0.0165)	(0.0161)	(0.0159)	(0.0159)
4rd quartile	0.279***	0.263***	0.352***	0.363***	0.240***	0.134***	0.101***	0.138***	0.119***	0.108***
	(0.0237)	(0.0252)	(0.0247)	(0.0269)	(0.0204)	(0.0141)	(0.0132)	(0.0147)	(0.0137)	(0.0135)
N	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381
R^2	0.425	0.487	0.467	0.493	0.435	0.315	0.349	0.348	0.358	0.340
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1. Covariates include include the full set of controls in Tables 3.2-3.6. Quartiles of ability distribution are derived from the distribution of ext of the subject being studied.

Table 3.8 Estimated origin bias by quartiles of ability distribution and subject

	Int - Ext					Prob(Int >Ext)				
	Math	Science	Basque	Spanish	English	Math	Science	Basque	Spanish	English
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Primary school										
1st quartile	-0.165***	-0.180***	-0.129***	-0.233***	-0.122***	-0.0757***	-0.0893***	-0.0644***	-0.0743***	-0.0655**
	(0.0409)	(0.0418)	(0.0378)	(0.0424)	(0.0429)	(0.0242)	(0.0249)	(0.0249)	(0.0263)	(0.0259)
2nd quartile	-0.126**	-0.176***	-0.244***	-0.272***	-0.116**	-0.119***	-0.139***	-0.163***	-0.169***	-0.0613**
	(0.0522)	(0.0510)	(0.0437)	(0.0473)	(0.0483)	(0.0305)	(0.0307)	(0.0265)	(0.0294)	(0.0304)
3rd quartile	-0.167**	-0.185***	-0.250***	-0.133**	-0.0453	-0.0585	-0.0929***	-0.149***	-0.0847**	0.0213
	(0.0690)	(0.0617)	(0.0527)	(0.0560)	(0.0573)	(0.0384)	(0.0316)	(0.0342)	(0.0341)	(0.0354)
4rd quartile	-0.123	-0.252**	-0.293***	-0.0831	0.0167	0.0691	-0.0233	-0.0754**	0.0448	0.0570
	(0.0809)	(0.104)	(0.0763)	(0.0765)	(0.0671)	(0.0440)	(0.0383)	(0.0372)	(0.0310)	(0.0384)
N	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802
R^2	0.439	0.511	0.487	0.497	0.476	0.306	0.346	0.347	0.355	0.319
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel B: Middle school										
1st quartile	-0.0280	0.00225	-0.0468	0.0223	-0.0958*	-0.0678**	-0.00460	0.0113	0.000157	-0.0675**
	(0.0507)	(0.0533)	(0.0507)	(0.0525)	(0.0489)	(0.0270)	(0.0243)	(0.0263)	(0.0263)	(0.0284)
2nd quartile	-0.124**	-0.0981	-0.207***	0.0190	-0.101*	-0.128***	-0.0284	-0.0921***	-0.0119	-0.0725**
	(0.0613)	(0.0642)	(0.0534)	(0.0561)	(0.0546)	(0.0367)	(0.0369)	(0.0339)	(0.0332)	(0.0361)
3rd quartile	-0.0431	-0.0843	-0.197**	-0.119*	-0.148**	-0.0182	-0.0767*	-0.111***	-0.113***	-0.0660*
	(0.0668)	(0.0788)	(0.0765)	(0.0647)	(0.0609)	(0.0401)	(0.0407)	(0.0400)	(0.0380)	(0.0378)
4rd quartile	0.0212	-0.252***	-0.172*	-0.131*	-0.0208	0.00818	-0.0569*	-0.0145	-0.0125	0.0209
	(0.0992)	(0.0833)	(0.0978)	(0.0789)	(0.0705)	(0.0477)	(0.0331)	(0.0470)	(0.0346)	(0.0396)
N	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381
R^2	0.425	0.487	0.466	0.493	0.435	0.313	0.341	0.346	0.356	0.339
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1. Covariates include include the full set of controls in Tables 3.2-3.6. Quartiles of ability distribution are derived from the distribution of ext of the subject being studied.

Table 3.9 Estimated socioeconomic bias by quartiles of ability distribution and subject

	Int - Ext					Prob(Int >Ext)				
	Math	Science	Basque	Spanish	English	Math	Science	Basque	Spanish	English
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Primary school										
1st quartile	-0.243***	-0.256***	-0.159***	-0.185***	-0.206***	-0.102***	-0.117***	-0.0778***	-0.0752***	-0.101***
	(0.0253)	(0.0258)	(0.0242)	(0.0259)	(0.0249)	(0.0151)	(0.0146)	(0.0151)	(0.0149)	(0.0150)
2nd quartile	-0.229***	-0.267***	-0.226***	-0.178***	-0.189***	-0.132***	-0.151***	-0.117***	-0.115***	-0.0822***
	(0.0312)	(0.0295)	(0.0249)	(0.0282)	(0.0260)	(0.0190)	(0.0181)	(0.0173)	(0.0174)	(0.0172)
3rd quartile	-0.210***	-0.230***	-0.190***	-0.242***	-0.119***	-0.0942***	-0.131***	-0.110***	-0.120***	-0.0283
	(0.0297)	(0.0300)	(0.0284)	(0.0300)	(0.0272)	(0.0185)	(0.0178)	(0.0179)	(0.0182)	(0.0180)
4rd quartile	-0.153***	-0.172***	-0.143***	-0.140***	-0.121***	-0.0162	-0.0290*	-0.0207	-0.0175	-0.0338*
	(0.0326)	(0.0331)	(0.0289)	(0.0296)	(0.0294)	(0.0177)	(0.0164)	(0.0170)	(0.0158)	(0.0193)
N	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802	15,802
R^2	0.438	0.509	0.486	0.495	0.474	0.306	0.346	0.341	0.354	0.318
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Panel B: Middle school										
1st quartile	-0.0466	-0.0309	-0.0958***	-0.0446*	-0.0851***	-0.00919	-0.0189	-0.0164	0.00238	-0.0329**
	(0.0284)	(0.0283)	(0.0254)	(0.0265)	(0.0245)	(0.0154)	(0.0140)	(0.0137)	(0.0138)	(0.0152)
2nd quartile	-0.134***	-0.148***	-0.110***	-0.103***	-0.121***	-0.0892***	-0.0716***	-0.0679***	-0.0430**	-0.0805***
	(0.0292)	(0.0302)	(0.0262)	(0.0288)	(0.0255)	(0.0179)	(0.0186)	(0.0176)	(0.0180)	(0.0182)
3rd quartile	-0.120***	-0.112***	-0.142***	-0.171***	-0.133***	-0.0529***	-0.0606***	-0.0766***	-0.116***	-0.0689***
	(0.0313)	(0.0310)	(0.0299)	(0.0312)	(0.0299)	(0.0181)	(0.0189)	(0.0184)	(0.0178)	(0.0204)
4rd quartile	-0.107***	-0.163***	-0.126***	-0.203***	-0.136***	-0.0510***	-0.0384**	-0.0191	-0.0283	-0.0223
	(0.0346)	(0.0346)	(0.0343)	(0.0376)	(0.0317)	(0.0196)	(0.0182)	(0.0188)	(0.0174)	(0.0196)
N	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381	15,381
R^2	0.424	0.487	0.466	0.492	0.434	0.313	0.347	0.346	0.356	0.339
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1. Covariates include include the full set of controls in Tables 3.2-3.6. Quartiles of ability distribution are derived from the distribution of ext of the subject being studied.

Table 3.7 presents the results of the estimated gender bias for each quartile of the ability distribution. We observe that the estimated female positive bias is significant in all segments of ability, with the exception of Math in primary education. This finding mirrors the main results of the analysis. In this case, we observe that the estimated bias is virtually zero in the first three quartiles of the score distribution, but it is significant and negative in the fourth quartile. Overall, the size of female bias remains fairly stable across the score distribution in the five subjects. Altogether, the evidence suggests that there is significant positive female bias at every point of the ability distribution, with the exception of Math. Thus, we can safely conclude that our results on gender bias are not driven by the presence of heterogeneous behaviors based on the score levels.

Alternatively, Table 3.8 displays the results of the analogous analysis by national origin. Given the limited presence of second-generation immigrants in the sample, we combine both types of immigrant students into a single group to enhance the power of our results. In primary school, we see that the negative origin bias in Science and Basque is significant in every point of the ability distribution. In the remaining subjects (Math, Spanish and English), the effect is more negative in the lower quartiles. In middle school, however, the effects do not show consistent patterns. The negative immigrant biases are concentrated in certain sections of the ability distribution. More specifically, we find highly significant results ($p < .05$) in the second quartile for Math, fourth quartile in Science, in the middle section of the distribution in Basque and in the third quartile of English.

Finally, Table 3.9 presents the estimated biases based on the household income of the student. To ease the exposition of the findings, we combine the lowest two income levels and the highest two into two separate groups. The results suggest that a negative socioeconomic bias exists for every ability level. We find that the negative biases are stronger in the lower points of the distribution, with the exception of Spanish. Surprisingly, this pattern is reversed in middle school. We can nevertheless conclude that the estimated socioeconomic bias is statistically significant for every ability quartile.

3.7.2 Does the estimated gender bias reflect statistical discrimination?

One reason why a positive female bias could emerge is the presence of statistical discrimination. As put by Lavy (2008), two explanations support this hypothesis. First, teachers might differently grade boys and girls as a result of expecting a higher average performance of female students. Second, teachers may display heterogeneous grading behaviors if they believe that the internal assessments are less reliable for boys. For example, male students can be perceived as cheating more than their female peers.

To evaluate this hypothesis, we replicate our analysis separately for classes where the average performance (as measured by ext) of female students is higher than that of males and vice versa. Table 3.10 displays the results of Equation (3.4) for the two separate samples. Both subsamples yield positive significant female bias in both education levels for every subject (with the exception of Math in primary which is, once again, in line with the main results). We observe that, in all instances, the effect is stronger in classes where female students outperform their male peers. Altogether, these results suggest that, although we cannot completely rule out the possibility of statistical discrimination, it is not the main channel through which the observed gender disparity operates.

In the same spirit, we also look at whether differences in the achievement variance across genders might affect our results. To test this possibility, we similarly divide our sample according to whether students attend a class where the performance variance (in ext) is higher for female students than for males. For this purpose, we use the gender-specific variance observed in a classroom in each specific subject. Results are displayed in Table 3.11. Overall, we find that significant positive female biases exist in both samples for every subject. Furthermore, the results show that the estimated assessment gaps are similar between both subsamples. This suggests that disparities in the gender performance variance do not play a substantial role in explaining our findings.

3.7.3 Exploiting within-student between-subjects variation

In Section 3.2, we established with a simple model that the observed gaps between int and ext could be caused by various factors. In particular, we mentioned the potential role of **(1)** systematic differences in the test-taking aptitudes that are correlated with a given trait (γ_s),

Table 3.10 Estimated gender bias by gender average performance in ext

	Int - Ext									
	Math		Science		Basque		Spanish		English	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Panel A: Primary school					
Female	0.0311	-0.00676	0.110***	0.0870***	0.189***	0.172***	0.275***	0.264***	0.155***	0.147***
	(0.0190)	(0.0175)	(0.0163)	(0.0222)	(0.0140)	(0.0197)	(0.0188)	(0.0182)	(0.0146)	(0.0210)
N	6,672	9,130	9,969	5,833	10,656	5,146	7,313	8,489	10,158	5,644
R^2	0.451	0.428	0.512	0.508	0.494	0.468	0.481	0.509	0.470	0.463
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fem >Male	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
					Panel B: Middle school					
Female	0.291***	0.277***	0.294***	0.262***	0.333***	0.319***	0.383***	0.388***	0.232***	0.207***
	(0.0207)	(0.0190)	(0.0187)	(0.0218)	(0.0154)	(0.0274)	(0.0171)	(0.0226)	(0.0135)	(0.0234)
N	7,049	8,332	8,879	6,502	11,527	3,854	9,787	5,594	11,100	4,281
R^2	0.424	0.426	0.483	0.491	0.464	0.469	0.487	0.503	0.433	0.434
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fem >Male	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗

Notes: OLS regressions use standard errors clustered at the class level. *** $p<.01$, ** $p<.05$, * $p<.1$. Covariates include include the full set of controls in Tables 3.2-3.6. Performance is measured using ext of the particular subject. Fem >Male stands for classes where the average performance in external assessments is higher for female than male students.

Table 3.11 Estimated gender bias by gender differences in the variance of ext

	Int - Ext									
	Math		Science		Basque		Spanish		English	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Primary school										
Female	-0.0120	0.0200	0.113***	0.0947***	0.177***	0.183***	0.267***	0.271***	0.111***	0.183***
	(0.0193)	(0.0171)	(0.0192)	(0.0181)	(0.0168)	(0.0154)	(0.0197)	(0.0172)	(0.0176)	(0.0162)
N	6,518	9,284	7,057	8,745	7,113	8,689	7,273	8,529	7,603	8,199
R^2	0.443	0.433	0.506	0.514	0.495	0.478	0.492	0.501	0.458	0.476
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fem >Male	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗
Panel B: Middle school										
Female	0.320***	0.255***	0.275***	0.282***	0.348***	0.307***	0.389***	0.373***	0.226***	0.221***
	(0.0218)	(0.0174)	(0.0208)	(0.0189)	(0.0179)	(0.0190)	(0.0177)	(0.0204)	(0.0167)	(0.0158)
N	6,226	9,155	6,646	8,735	8,095	7,286	7,870	7,511	7,624	7,757
R^2	0.432	0.420	0.486	0.488	0.470	0.460	0.503	0.483	0.432	0.434
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fem >Male	✓	✗	✓	✗	✓	✗	✓	✗	✓	✗

Notes: OLS regressions use standard errors clustered at the class level. *** $p<.01$, ** $p<.05$, * $p<.1$. Covariates include include the full set of controls in Tables 3.2-3.6. Performance variance is measured using ext of the particular subject. Fem >Male stands for classes where the variance of performance in external assessments is higher for female than male students.

(2) the distinctive responsiveness to changes in the stakes across groups ($\phi_s \delta_s$) and **(3)** how grading attitudes of teachers towards certain traits could be behind our findings (τ_{js}). We now turn to discuss some specific elements associated with the former two determinants, and to formally test whether our results can be attributed to the latter factor.

Among the different elements captured in γ_s, we consider two complementary features affecting the emergence of a gap between internal and external scores. First, students' behavior in the classroom significantly impacts teachers' assessments. A survey among teachers in charge of the curriculum in each Basque school reveals that bad classroom behavior is among the most important reasons that cause students' grade retention; a decision that is motivated by low teacher assessments [Arregi *et al.* (2009)]. Thus, our findings may be a result of the distinct relevance attached to pupils' behavior in internal assessments and the presence of group disparities in student behavior. Second, another source of concern is that the evaluation format differs between the two assessments. While our external assessments typically rely on an item-based approach, the design of internal assessments are decided by the teachers themselves and do not necessarily share the same structure. Ben-Shakhar & Sinai (1991) provide evidence that female tendencies towards guessing less in multiple-choice tests grant male students an advantage in these type of exams. Consequently, for example, our results on positive gender bias might therefore be a product of this consideration.

In addition, it is also plausible that there are systematic differences in how certain groups respond to changes in the stakes of the assessments. In this sense, existing evidence suggests that significant heterogeneity exists in the amount of effort that student groups exert, based on the stakes of the examination. For instance, female students are found to devote more effort than their male peers in low stake examinations (for example, O'Neil *et al.* (2005) and Eklöf (2007) using TIMSS data). Interestingly, following a reform in the Spanish university entrance exam, Arenas & Calsamiglia (2020) also find that female performance was negatively affected in tests for which the stakes increased more. This indicates that male students might not take the low-stake external assessments as seriously, which would in turn affect our results.

To test for the presence of teacher differential grading attitudes towards students' traits, we now turn to investigate the variation in the observed gaps across subjects, by using a Diff-in-Diffs type of analysis. To this end, we replicate the previous analysis by pooling the scores of the five available subjects, and testing whether the estimated coefficients are equal

Table 3.12 Hypothesis test for the presence of stereotyping, using within-student between-subject variation

PERSONAL TRAIT	Female (1)	Immigrant (2)	Income (3)
Panel A - **Dependent:** $int - ext$			
p-value	0.000	0.000	0.000
F-statistic	22.666	17.762	15.031
Panel B - **Dependent:** $Pr(int > ext)$			
p-value	0.000	0.000	0.000
F-statistic	14.997	8.983	9.270

Notes: This table reports the equality test of the observed gap between int and ext for a given trait, across subjects. The sample consists of primary school students. We run regressions using as explanatory variables: subject-specific z-score, student fixed effects, subject fixed effects, subject $\times$ gender, subject $\times$ national origin and subject $\times$ income. Panel A (B) reports the outcome of the regression using the continuous gap (dummy indicator) as dependent variable. Each observation corresponds to a student-subject pair, i.e. each student has 5 observations, one per each subject. Standard errors are clustered at the classroom level. Immigrant dummy pools both first- and second-generation immigrants. Income is included as a linear term, as opposed to a set of dummies. The inclusion of student fixed effects control for student-specific constant effects on both int and ext across subjects. The omitted baseline subject category is Math. We reject the null hypothesis that: (i) interaction terms of subject and female dummy are equal, (ii) interaction terms of subject and immigrant dummy are equal, and (iii) interaction terms of subject and income are equal.

across subjects for the same trait group. For this part, we focus on primary school. A teacher in primary education typically provides instruction to a given classroom in all core subjects. Therefore, in the spirit of Burgess & Greaves (2013), we explore the variation between subjects for a given student-teacher match for each personal trait.

To exploit the presence of five observations per student, we now use student fixed effects in the regression. This allows us to control for the aforementioned confounding channels. First, by focusing on primary education, we reduce the possibility that a student engages in different types of behaviors across subjects, given that they are typically assigned to the same teacher. Second, we are able to substantially control for effort responsiveness and the behavior in the classroom. The reason is that the inclusion of individual fixed effects allows to absorb a significant portion of the unobserved variation stemming from class behavior and effort.[14]

Table 3.12 reports the F-statistic and associated p-value of the equality test for the interaction terms between subjects and each trait, within students. We find that there exist substantial trait $\times$ subject differences in the gap between internal and external assessments,

[14]This is true if students similarly modify their effort levels when they move from the internal to the external assessment in the different subjects.

even after the incorporation of student fixed effects. The results are significant ($p < .01$) for gender, immigrant and socioeconomic origin, using both considerations of the dependent variable. Altogether, this finding suggests that teachers display disparate grading attitudes between subjects and personal characteristics that are largely unexplained by student-specific considerations, like classroom behavior and effort responsiveness.

One potential explanation for the observed heterogeneity in assessment gaps between subjects is the presence of stereotyping on the part of teachers. We believe that our findings are consistent with this hypothesis. Recent research has found that teachers display significant heterogeneity in gender and immigrant stereotyping across subjects [Alesina *et al.* (2018); Carlana (2019)]. Unfortunately, we cannot directly measure the extent of stereotyping through personally testing or inquiring teachers, like in the aforementioned papers, or in Alan *et al.* (2018) and Alan *et al.* (2020). Therefore, we cannot rule out the existence of other reasons that explain these empirical results. Although we are not aware of these alternative explanations, the interpretation that teachers' stereotypes account for the observed findings need to be considered cautiously.

3.8 Conclusion

In this essay, we examine whether teachers' assessments particularly hinder certain socio-demographic groups that suffer from higher retention levels: boys, immigrant students, and lower income families. Using data from the Basque Country (Spain), we use quasi-blindly graded external evaluations and non-blind internal assessments to uncover the presence of significant assessment gaps for the above-mentioned groups. We find that, controlling for class fixed effects, these groups are significantly underassessed relative to girls, native students, and pupils with higher income. We perform a number of checks to study the stability of our results. The qualitative nature of the findings remain.

Then, we investigate some potential mechanisms that could explain the observed patterns. We find that the results are not primarily driven by students from certain segments of the ability distribution. At the same time, the findings suggest that the detected gender gaps are not primarily driven by the presence of statistical discrimination. Finally, we focus on studying whether certain student-specific unobservables, like student behavior, are behind our results. Controlling for student fixed effects, we find that significant differences across socio-demographic groups and subjects exist. Overall, we believe that these patterns are consistent with recent papers that find distinct stereotyping behavior across subjects. However, we are not able to test this directly, and thus remain cautious with regards to the adequacy of this interpretation.

This study contributes to the policy debate of assessment policy and its relation to grade retention, pointing to the cultural hypothesis of why some countries make students (and in particular, students of certain socio-demographic groups) disproportionately repeat. By showing substantive teacher bias by gender, socioeconomic status and immigrant origin in grading students, results suggest that removing such biases in a positive direction (hence favoring boys, immigrants and low-SES students) would simultaneously contribute to promote efficiency (by reducing student grade failure and retention, an ineffective policy) and equity (by narrowing the retention gaps) in education systems. Given that biases are likely to stem from stereotypes, beliefs and long-standing school practices, reducing them would require a combination of a wise regulatory framework incentivizing more positive assessment overall, as well as targeting interventions aiming at modifying teacher stereotypes and beliefs over certain students' true abilities and skills [Alesina *et al.* (2018)].

3.9 Appendix

THE FOLLOWING PAGES INCLUDE TABLES THAT CORRESPOND TO THE ROBUSTNESS CHECKS DISCUSSED IN SECTION 3.6

Table 3.13 Results - Math (Cubic specification)

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.360***	-0.377***	-0.378***	-0.427***	-0.447***	-0.449***	-0.165***	-0.174***	-0.174***	-0.226***	-0.235***	-0.236***
	(0.0111)	(0.0111)	(0.0111)	(0.0130)	(0.0129)	(0.0129)	(0.00761)	(0.00758)	(0.00759)	(0.00743)	(0.00728)	(0.00729)
External Assessment (Squared)	-0.00160	-0.00173	-0.00231	0.0297***	0.0292***	0.0286***	-0.00575**	-0.00582**	-0.00611**	-0.00557*	-0.00577*	-0.00607*
	(0.00482)	(0.00480)	(0.00480)	(0.00593)	(0.00591)	(0.00593)	(0.00289)	(0.00289)	(0.00290)	(0.00329)	(0.00330)	(0.00331)
External Assessment (Cubic)	-0.0278***	-0.0278***	-0.0276***	-0.0120***	-0.0116***	-0.0114***	-0.00960***	-0.00959***	-0.00951***	0.00692***	0.00712***	0.00720***
	(0.00289)	(0.00286)	(0.00287)	(0.00402)	(0.00398)	(0.00399)	(0.00188)	(0.00187)	(0.00187)	(0.00206)	(0.00205)	(0.00206)
Female	0.0117	0.00455	0.00484	0.292***	0.284***	0.284***	-0.00178	-0.00513	-0.00499	0.132***	0.128***	0.128***
	(0.0129)	(0.0128)	(0.0128)	(0.0137)	(0.0137)	(0.0136)	(0.00799)	(0.00797)	(0.00797)	(0.00792)	(0.00790)	(0.00787)
Immigrant (1st gen)	-0.244***	-0.209***	-0.199***	-0.128***	-0.0896**	-0.0746**	-0.0821***	-0.0656***	-0.0606**	-0.105***	-0.0882***	-0.0804***
	(0.0421)	(0.0414)	(0.0416)	(0.0361)	(0.0358)	(0.0358)	(0.0243)	(0.0244)	(0.0244)	(0.0203)	(0.0203)	(0.0203)
Immigrant (2nd gen)	-0.163***	-0.118***	-0.110***	-0.0440	0.00487	0.0139	-0.0966***	-0.0759***	-0.0716***	-0.0492	-0.0272	-0.0225
	(0.0361)	(0.0362)	(0.0362)	(0.0669)	(0.0665)	(0.0665)	(0.0204)	(0.0204)	(0.0204)	(0.0348)	(0.0348)	(0.0347)
Undefined	-0.133***	-0.112***	-0.109***	-0.0299	-0.0168	-0.0119	-0.0624***	-0.0525***	-0.0509***	0.00596	0.0119	0.0144
	(0.0271)	(0.0269)	(0.0267)	(0.0463)	(0.0453)	(0.0452)	(0.0177)	(0.0175)	(0.0174)	(0.0283)	(0.0279)	(0.0276)
Income = 2	0.160***	0.139***	0.137***	0.114***	0.0986***	0.0977***	0.0637***	0.0540***	0.0532***	0.0414**	0.0346**	0.0341**
	(0.0264)	(0.0263)	(0.0264)	(0.0276)	(0.0274)	(0.0274)	(0.0165)	(0.0165)	(0.0165)	(0.0161)	(0.0160)	(0.0160)
Income = 3	0.249***	0.209***	0.206***	0.148***	0.112***	0.109***	0.106***	0.0871***	0.0858***	0.0692***	0.0529***	0.0513***
	(0.0239)	(0.0236)	(0.0236)	(0.0238)	(0.0239)	(0.0240)	(0.0147)	(0.0147)	(0.0147)	(0.0136)	(0.0136)	(0.0136)
Income = 4	0.359***	0.291***	0.283***	0.212***	0.151***	0.142***	0.156***	0.124***	0.120***	0.0960***	0.0689***	0.0639***
	(0.0193)	(0.0191)	(0.0192)	(0.0195)	(0.0203)	(0.0203)	(0.0116)	(0.0118)	(0.0118)	(0.0111)	(0.0113)	(0.0113)
Special Needs	-0.489***	-0.482***	-0.481***	-0.218***	-0.208***	-0.205***	-0.199***	-0.196***	-0.195***	-0.0907***	-0.0860***	-0.0846***
	(0.0310)	(0.0310)	(0.0311)	(0.0348)	(0.0346)	(0.0347)	(0.0177)	(0.0176)	(0.0176)	(0.0211)	(0.0211)	(0.0211)
Grade Retention	-0.325***	-0.296***	-0.293***	-0.466***	-0.440***	-0.436***	-0.180***	-0.166***	-0.165***	-0.193***	-0.181***	-0.179***
	(0.0471)	(0.0478)	(0.0478)	(0.0428)	(0.0428)	(0.0427)	(0.0284)	(0.0286)	(0.0286)	(0.0231)	(0.0232)	(0.0232)
2nd Tertile (ISEC)		0.165***	0.162***		0.118***	0.113***		0.0743***	0.0728***		0.0549***	0.0524***
		(0.0134)	(0.0133)		(0.0160)	(0.0161)		(0.00848)	(0.00847)		(0.00888)	(0.00889)
3rd Tertile (ISEC)		0.219***	0.215***		0.214***	0.206***		0.104***	0.102***		0.0949***	0.0905***
		(0.0146)	(0.0146)		(0.0167)	(0.0169)		(0.00931)	(0.00931)		(0.00972)	(0.00976)
Home language (Basque)			0.0697***			0.129***			0.0353***			0.0673***
			(0.0175)			(0.0195)			(0.0108)			(0.0122)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.430	0.440	0.441	0.416	0.424	0.425	0.293	0.300	0.301	0.301	0.306	0.308
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator ($= 1$if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** $p <.01$, ** $p <.05$, * $p <.1$.

Table 3.14 Results - Science (Cubic specification)

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.471***	-0.495***	-0.496***	-0.497***	-0.521***	-0.525***	-0.210***	-0.220***	-0.220***	-0.253***	-0.262***	-0.264***
	(0.0118)	(0.0117)	(0.0117)	(0.0125)	(0.0124)	(0.0124)	(0.00765)	(0.00761)	(0.00763)	(0.00696)	(0.00693)	(0.00695)
External Assessment (Squared)	-0.0190***	-0.0192***	-0.0200***	-0.00134	-0.00227	-0.00336	-0.0129***	-0.0130***	-0.0133***	-0.0119***	-0.0122***	-0.0126***
	(0.00465)	(0.00458)	(0.00460)	(0.00542)	(0.00537)	(0.00537)	(0.00258)	(0.00257)	(0.00258)	(0.00281)	(0.00280)	(0.00280)
External Assessment (Cubic)	-0.0215***	-0.0203***	-0.0202***	-0.0107***	-0.01000***	-0.00947***	-0.00379**	-0.00333*	-0.00327*	0.0102***	0.0104***	0.0106***
	(0.00309)	(0.00303)	(0.00303)	(0.00342)	(0.00340)	(0.00337)	(0.00178)	(0.00178)	(0.00178)	(0.00161)	(0.00161)	(0.00161)
Female	0.103***	0.100***	0.101***	0.286***	0.279***	0.279***	0.0374***	0.0363***	0.0366***	0.121***	0.118***	0.118***
	(0.0133)	(0.0131)	(0.0131)	(0.0143)	(0.0141)	(0.0140)	(0.00755)	(0.00751)	(0.00753)	(0.00771)	(0.00770)	(0.00768)
Immigrant (1st gen)	-0.308***	-0.269***	-0.258***	-0.154***	-0.110***	-0.0908**	-0.126***	-0.111***	-0.107***	-0.0581***	-0.0406*	-0.0339
	(0.0438)	(0.0434)	(0.0437)	(0.0395)	(0.0394)	(0.0393)	(0.0236)	(0.0235)	(0.0237)	(0.0207)	(0.0207)	(0.0207)
Immigrant (2nd gen)	-0.199***	-0.152***	-0.142***	-0.0956	-0.0374	-0.0257	-0.115***	-0.0963***	-0.0923***	-0.0392	-0.0164	-0.0123
	(0.0380)	(0.0377)	(0.0378)	(0.0707)	(0.0698)	(0.0697)	(0.0206)	(0.0206)	(0.0206)	(0.0334)	(0.0329)	(0.0328)
Undefined	-0.135***	-0.113***	-0.109***	-0.0252	-0.00950	-0.00333	-0.0459***	-0.0370**	-0.0356**	0.0225	0.0286	0.0308
	(0.0281)	(0.0274)	(0.0271)	(0.0417)	(0.0406)	(0.0405)	(0.0178)	(0.0178)	(0.0177)	(0.0261)	(0.0258)	(0.0259)
Income = 2	0.196***	0.173***	0.172***	0.107***	0.0887***	0.0873***	0.0643***	0.0555***	0.0547***	0.0425***	0.0354**	0.0349**
	(0.0263)	(0.0261)	(0.0262)	(0.0286)	(0.0282)	(0.0282)	(0.0151)	(0.0149)	(0.0150)	(0.0156)	(0.0155)	(0.0155)
Income = 3	0.297***	0.252***	0.249***	0.147***	0.104***	0.100***	0.124***	0.107***	0.106***	0.0611***	0.0444***	0.0430***
	(0.0236)	(0.0233)	(0.0234)	(0.0245)	(0.0244)	(0.0245)	(0.0136)	(0.0136)	(0.0136)	(0.0138)	(0.0139)	(0.0139)
Income = 4	0.402***	0.323***	0.314***	0.234***	0.162***	0.150***	0.174***	0.143***	0.140***	0.0973***	0.0687***	0.0645***
	(0.0193)	(0.0192)	(0.0195)	(0.0202)	(0.0211)	(0.0210)	(0.0111)	(0.0110)	(0.0112)	(0.0112)	(0.0116)	(0.0116)
Special Needs	-0.470***	-0.462***	-0.461***	-0.364***	-0.350***	-0.347***	-0.163***	-0.160***	-0.160***	-0.137***	-0.132***	-0.130***
	(0.0321)	(0.0318)	(0.0319)	(0.0390)	(0.0388)	(0.0389)	(0.0182)	(0.0181)	(0.0182)	(0.0196)	(0.0195)	(0.0195)
Grade Retention	-0.569***	-0.536***	-0.533***	-0.669***	-0.640***	-0.635***	-0.235***	-0.222***	-0.221***	-0.255***	-0.244***	-0.242***
	(0.0518)	(0.0518)	(0.0517)	(0.0459)	(0.0457)	(0.0456)	(0.0282)	(0.0282)	(0.0283)	(0.0209)	(0.0208)	(0.0208)
2nd Tertile (ISEC)		0.163***	0.160***		0.130***	0.124***		0.0614***	0.0601***		0.0487***	0.0465***
		(0.0148)	(0.0148)		(0.0162)	(0.0163)		(0.00876)	(0.00878)		(0.00900)	(0.00900)
3rd Tertile (ISEC)		0.260***	0.255***		0.252***	0.242***		0.102***	0.100***		0.100***	0.0965***
		(0.0161)	(0.0161)		(0.0170)	(0.0170)		(0.00943)	(0.00943)		(0.00961)	(0.00960)
Home language (Basque)			0.0820***			0.159***			0.0332***			0.0566***
			(0.0168)			(0.0197)			(0.0104)			(0.0112)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.501	0.511	0.512	0.475	0.485	0.487	0.339	0.345	0.345	0.342	0.348	0.349
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator ($=1$if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level.*** $p<$.01, ** $p<$.05, * $p<$.1.

Table 3.15 Results - Basque (Cubic specification)

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.351***	-0.372***	-0.378***	-0.399***	-0.419***	-0.429***	-0.161***	-0.170***	-0.173***	-0.215***	-0.223***	-0.227***
	(0.0121)	(0.0120)	(0.0120)	(0.0118)	(0.0117)	(0.0117)	(0.00790)	(0.00791)	(0.00792)	(0.00724)	(0.00720)	(0.00720)
External Assessment (Squared)	0.0151***	0.0135**	0.00866	0.0319***	0.0315***	0.0273***	0.0125***	0.0118***	0.00953***	0.00968***	0.00939***	0.00750**
	(0.00534)	(0.00525)	(0.00528)	(0.00587)	(0.00589)	(0.00587)	(0.00324)	(0.00322)	(0.00322)	(0.00356)	(0.00355)	(0.00352)
External Assessment (Cubic)	-0.0247***	-0.0235***	-0.0231***	-0.0219***	-0.0215***	-0.0208***	-0.0134***	-0.0128***	-0.0126***	-0.000703	-0.000559	-0.000219
	(0.00349)	(0.00346)	(0.00346)	(0.00350)	(0.00349)	(0.00346)	(0.00204)	(0.00203)	(0.00204)	(0.00193)	(0.00193)	(0.00192)
Female	0.179***	0.178***	0.180***	0.324***	0.323***	0.326***	0.0817***	0.0810***	0.0819***	0.148***	0.147***	0.149***
	(0.0116)	(0.0115)	(0.0114)	(0.0132)	(0.0131)	(0.0130)	(0.00741)	(0.00740)	(0.00736)	(0.00787)	(0.00783)	(0.00782)
Immigrant (1st gen)	-0.305***	-0.274***	-0.245***	-0.214***	-0.178***	-0.150***	-0.153***	-0.138***	-0.125***	-0.0730***	-0.0589***	-0.0462**
	(0.0392)	(0.0381)	(0.0382)	(0.0358)	(0.0356)	(0.0354)	(0.0245)	(0.0242)	(0.0241)	(0.0202)	(0.0199)	(0.0198)
Immigrant (2nd gen)	-0.238***	-0.199***	-0.174***	-0.144**	-0.0978	-0.0806	-0.137***	-0.119***	-0.107***	-0.0539	-0.0358	-0.0280
	(0.0346)	(0.0348)	(0.0347)	(0.0637)	(0.0631)	(0.0633)	(0.0203)	(0.0204)	(0.0203)	(0.0336)	(0.0334)	(0.0336)
Undefined	-0.134***	-0.115***	-0.106***	-0.0278	-0.0157	-0.00649	-0.0598***	-0.0507***	-0.0464***	-0.000419	0.00429	0.00848
	(0.0263)	(0.0259)	(0.0253)	(0.0403)	(0.0405)	(0.0390)	(0.0177)	(0.0177)	(0.0176)	(0.0212)	(0.0211)	(0.0205)
Income = 2	0.151***	0.132***	0.128***	0.0688***	0.0546**	0.0529**	0.0631***	0.0541***	0.0519***	0.0248	0.0192	0.0185
	(0.0230)	(0.0230)	(0.0231)	(0.0257)	(0.0253)	(0.0252)	(0.0155)	(0.0156)	(0.0156)	(0.0157)	(0.0156)	(0.0156)
Income = 3	0.217***	0.182***	0.175***	0.167***	0.133***	0.127***	0.116***	0.0987***	0.0952***	0.0593***	0.0458***	0.0432***
	(0.0211)	(0.0210)	(0.0211)	(0.0231)	(0.0232)	(0.0232)	(0.0139)	(0.0138)	(0.0138)	(0.0137)	(0.0138)	(0.0137)
Income = 4	0.331***	0.269***	0.246***	0.213***	0.156***	0.139***	0.147***	0.118***	0.107***	0.0857***	0.0628***	0.0551***
	(0.0186)	(0.0183)	(0.0182)	(0.0181)	(0.0186)	(0.0186)	(0.0114)	(0.0115)	(0.0116)	(0.0108)	(0.0112)	(0.0112)
Special Needs	-0.501***	-0.494***	-0.492***	-0.220***	-0.211***	-0.207***	-0.232***	-0.229***	-0.228***	-0.0921***	-0.0888***	-0.0868***
	(0.0300)	(0.0299)	(0.0301)	(0.0340)	(0.0340)	(0.0341)	(0.0175)	(0.0175)	(0.0176)	(0.0203)	(0.0204)	(0.0204)
Grade Retention	-0.433***	-0.408***	-0.400***	-0.586***	-0.562***	-0.556***	-0.237***	-0.225***	-0.221***	-0.234***	-0.224***	-0.222***
	(0.0386)	(0.0382)	(0.0380)	(0.0452)	(0.0449)	(0.0447)	(0.0286)	(0.0284)	(0.0283)	(0.0209)	(0.0210)	(0.0208)
2nd Tertile (ISEC)		0.130***	0.123***		0.116***	0.107***		0.0644***	0.0609***		0.0409***	0.0371***
		(0.0136)	(0.0136)		(0.0150)	(0.0149)		(0.00884)	(0.00880)		(0.00900)	(0.00900)
3rd Tertile (ISEC)		0.209***	0.199***		0.202***	0.188***		0.0977***	0.0928***		0.0819***	0.0756***
		(0.0145)	(0.0144)		(0.0158)	(0.0157)		(0.00924)	(0.00919)		(0.00930)	(0.00932)
Home language (Basque)			0.212***			0.244***			0.100***			0.111***
			(0.0175)			(0.0184)			(0.0106)			(0.0119)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.473	0.481	0.487	0.453	0.460	0.466	0.327	0.332	0.337	0.331	0.335	0.339
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator ($= 1$if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table 3.16 Results - Spanish (Cubic specification)

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.453***	-0.467***	-0.467***	-0.531***	-0.552***	-0.552***	-0.201***	-0.206***	-0.207***	-0.258***	-0.266***	-0.266***
	(0.0116)	(0.0115)	(0.0115)	(0.0114)	(0.0113)	(0.0112)	(0.00777)	(0.00766)	(0.00767)	(0.00648)	(0.00640)	(0.00639)
External Assessment (Squared)	0.0153***	0.0147***	0.0147***	0.0180***	0.0187***	0.0184***	-0.000235	-0.000477	-0.000440	0.000391	0.000681	0.000550
	(0.00534)	(0.00524)	(0.00524)	(0.00537)	(0.00536)	(0.00534)	(0.00301)	(0.00299)	(0.00298)	(0.00269)	(0.00270)	(0.00269)
External Assessment (Cubic)	-0.0157***	-0.0163***	-0.0163***	-0.00933***	-0.00901***	-0.00871***	-0.00409**	-0.00435**	-0.00435**	0.00723***	0.00736***	0.00749***
	(0.00362)	(0.00361)	(0.00361)	(0.00330)	(0.00326)	(0.00324)	(0.00206)	(0.00206)	(0.00206)	(0.00154)	(0.00154)	(0.00153)
Female	0.276***	0.270***	0.270***	0.385***	0.380***	0.380***	0.116***	0.113***	0.113***	0.153***	0.151***	0.151***
	(0.0131)	(0.0130)	(0.0130)	(0.0136)	(0.0135)	(0.0134)	(0.00754)	(0.00749)	(0.00750)	(0.00760)	(0.00757)	(0.00757)
Immigrant (1st gen)	-0.316***	-0.279***	-0.279***	-0.117***	-0.0720*	-0.0564	-0.128***	-0.112***	-0.113***	-0.0599***	-0.0419**	-0.0353*
	(0.0402)	(0.0398)	(0.0398)	(0.0371)	(0.0370)	(0.0370)	(0.0245)	(0.0244)	(0.0244)	(0.0194)	(0.0194)	(0.0194)
Immigrant (2nd gen)	-0.188***	-0.144***	-0.144***	-0.0134	0.0399	0.0490	-0.0863***	-0.0668***	-0.0677***	-0.0311	-0.00975	-0.00586
	(0.0350)	(0.0350)	(0.0350)	(0.0621)	(0.0611)	(0.0611)	(0.0211)	(0.0210)	(0.0210)	(0.0349)	(0.0345)	(0.0344)
Undefined	-0.119***	-0.0963***	-0.0963***	-0.0941**	-0.0796**	-0.0747*	-0.0607***	-0.0505***	-0.0508***	-0.0128	-0.00698	-0.00488
	(0.0272)	(0.0268)	(0.0267)	(0.0399)	(0.0395)	(0.0393)	(0.0165)	(0.0162)	(0.0162)	(0.0242)	(0.0238)	(0.0236)
Income = 2	0.185***	0.164***	0.164***	0.119***	0.101***	0.0997***	0.0717***	0.0622***	0.0623***	0.0587***	0.0514***	0.0510***
	(0.0261)	(0.0261)	(0.0261)	(0.0274)	(0.0271)	(0.0270)	(0.0149)	(0.0149)	(0.0149)	(0.0150)	(0.0150)	(0.0150)
Income = 3	0.242***	0.202***	0.202***	0.159***	0.117***	0.114***	0.117***	0.0986***	0.0989***	0.0602***	0.0436***	0.0421***
	(0.0235)	(0.0235)	(0.0236)	(0.0255)	(0.0255)	(0.0255)	(0.0137)	(0.0137)	(0.0137)	(0.0138)	(0.0139)	(0.0139)
Income = 4	0.332***	0.262***	0.262***	0.252***	0.182***	0.172***	0.140***	0.109***	0.110***	0.101***	0.0730***	0.0688***
	(0.0193)	(0.0194)	(0.0195)	(0.0206)	(0.0211)	(0.0211)	(0.0112)	(0.0115)	(0.0116)	(0.0106)	(0.0111)	(0.0111)
Special Needs	-0.541***	-0.533***	-0.533***	-0.270***	-0.259***	-0.255***	-0.218***	-0.214***	-0.215***	-0.0858***	-0.0814***	-0.0794***
	(0.0322)	(0.0321)	(0.0321)	(0.0346)	(0.0345)	(0.0345)	(0.0174)	(0.0173)	(0.0173)	(0.0189)	(0.0188)	(0.0188)
Grade Retention	-0.458***	-0.427***	-0.427***	-0.630***	-0.600***	-0.595***	-0.194***	-0.180***	-0.181***	-0.218***	-0.206***	-0.204***
	(0.0472)	(0.0467)	(0.0468)	(0.0443)	(0.0440)	(0.0441)	(0.0266)	(0.0266)	(0.0267)	(0.0211)	(0.0212)	(0.0212)
2nd Tertile (ISEC)		0.147***	0.147***		0.125***	0.119***		0.0682***	0.0685***		0.0515***	0.0492***
		(0.0142)	(0.0142)		(0.0155)	(0.0156)		(0.00900)	(0.00904)		(0.00872)	(0.00873)
3rd Tertile (ISEC)		0.222***	0.222***		0.241***	0.231***		0.0983***	0.0988***		0.0956***	0.0914***
		(0.0149)	(0.0149)		(0.0166)	(0.0166)		(0.00927)	(0.00929)		(0.00914)	(0.00916)
Home language (Basque)			0.000145			0.129***			-0.00676			0.0551***
			(0.0170)			(0.0207)			(0.0102)			(0.0113)
Observations	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.489	0.497	0.497	0.483	0.491	0.493	0.341	0.346	0.346	0.347	0.352	0.354
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator $(= 1\text{if Int}>\text{Ext})$ in the right panel. OLS regressions use standard errors clustered at the class level.*** $p <.01$, ** $p <.05$, * $p <.1$.

Table 3.17 Results - English (Cubic specification)

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.279***	-0.296***	-0.296***	-0.282***	-0.298***	-0.300***	-0.141***	-0.150***	-0.150***	-0.160***	-0.167***	-0.167***
	(0.0106)	(0.0107)	(0.0107)	(0.0119)	(0.0120)	(0.0119)	(0.00759)	(0.00753)	(0.00754)	(0.00793)	(0.00786)	(0.00785)
External Assessment (Squared)	-0.0228***	-0.0234***	-0.0234***	-0.0152**	-0.0153**	-0.0153**	-0.00980***	-0.0101***	-0.0101***	-0.0187***	-0.0188***	-0.0188***
	(0.00549)	(0.00546)	(0.00547)	(0.00614)	(0.00612)	(0.00609)	(0.00341)	(0.00339)	(0.00339)	(0.00399)	(0.00399)	(0.00398)
External Assessment (Cubic)	-0.0272***	-0.0263***	-0.0262***	-0.0129***	-0.0124***	-0.0118***	-0.00376**	-0.00329*	-0.00328*	-0.00700***	-0.00682**	-0.00648**
	(0.00240)	(0.00239)	(0.00239)	(0.00422)	(0.00422)	(0.00420)	(0.00179)	(0.00178)	(0.00179)	(0.00266)	(0.00266)	(0.00266)
Female	0.145***	0.143***	0.143***	0.223***	0.222***	0.222***	0.0725***	0.0714***	0.0715***	0.119***	0.119***	0.119***
	(0.0119)	(0.0119)	(0.0119)	(0.0115)	(0.0115)	(0.0115)	(0.00769)	(0.00766)	(0.00766)	(0.00781)	(0.00778)	(0.00777)
Immigrant (1st gen)	-0.204***	-0.172***	-0.165***	-0.135***	-0.106***	-0.0912***	-0.0751***	-0.0586**	-0.0552**	-0.0774***	-0.0661***	-0.0574***
	(0.0406)	(0.0398)	(0.0397)	(0.0337)	(0.0336)	(0.0336)	(0.0238)	(0.0234)	(0.0234)	(0.0212)	(0.0213)	(0.0213)
Immigrant (2nd gen)	-0.0745**	-0.0367	-0.0299	-0.158***	-0.121**	-0.112**	-0.0350	-0.0153	-0.0123	-0.0761**	-0.0615*	-0.0564
	(0.0352)	(0.0354)	(0.0353)	(0.0529)	(0.0527)	(0.0528)	(0.0216)	(0.0218)	(0.0218)	(0.0343)	(0.0343)	(0.0344)
Undefined	-0.0520*	-0.0340	-0.0316	-0.0157	-0.00635	-0.00170	-0.0371**	-0.0277	-0.0266	-0.00698	-0.00333	-0.000611
	(0.0271)	(0.0266)	(0.0266)	(0.0367)	(0.0367)	(0.0366)	(0.0185)	(0.0184)	(0.0184)	(0.0303)	(0.0303)	(0.0301)
Income = 2	0.187***	0.170***	0.169***	0.0824***	0.0720***	0.0710***	0.101***	0.0919***	0.0913***	0.0346**	0.0304*	0.0299*
	(0.0226)	(0.0225)	(0.0225)	(0.0244)	(0.0241)	(0.0241)	(0.0151)	(0.0151)	(0.0151)	(0.0164)	(0.0164)	(0.0164)
Income = 3	0.234***	0.201***	0.199***	0.151***	0.126***	0.123***	0.115***	0.0973***	0.0964***	0.0688***	0.0589***	0.0570***
	(0.0211)	(0.0210)	(0.0210)	(0.0215)	(0.0216)	(0.0216)	(0.0136)	(0.0137)	(0.0137)	(0.0145)	(0.0146)	(0.0147)
Income = 4	0.292***	0.237***	0.231***	0.196***	0.155***	0.146***	0.131***	0.102***	0.0991***	0.0867***	0.0700***	0.0646***
	(0.0180)	(0.0180)	(0.0180)	(0.0169)	(0.0175)	(0.0176)	(0.0115)	(0.0118)	(0.0118)	(0.0113)	(0.0116)	(0.0117)
Special Needs	-0.514***	-0.508***	-0.507***	-0.251***	-0.245***	-0.241***	-0.237***	-0.234***	-0.234***	-0.128***	-0.126***	-0.124***
	(0.0293)	(0.0291)	(0.0292)	(0.0335)	(0.0335)	(0.0334)	(0.0179)	(0.0179)	(0.0180)	(0.0219)	(0.0219)	(0.0219)
Grade Retention	-0.307***	-0.283***	-0.281***	-0.450***	-0.435***	-0.430***	-0.176***	-0.163***	-0.162***	-0.208***	-0.202***	-0.199***
	(0.0442)	(0.0441)	(0.0441)	(0.0415)	(0.0416)	(0.0414)	(0.0288)	(0.0288)	(0.0287)	(0.0234)	(0.0235)	(0.0234)
2nd Tertile (ISEC)		0.135***	0.133***		0.0917***	0.0866***		0.0639***	0.0628***		0.0314***	0.0284***
		(0.0139)	(0.0139)		(0.0138)	(0.0138)		(0.00908)	(0.00908)		(0.00939)	(0.00941)
3rd Tertile (ISEC)		0.178***	0.174***		0.146***	0.137***		0.0973***	0.0957***		0.0604***	0.0551***
		(0.0141)	(0.0141)		(0.0142)	(0.0143)		(0.00928)	(0.00928)		(0.00976)	(0.00980)
Home language (Basque)			0.0554***			0.124***			0.0246**			0.0725***
			(0.0166)			(0.0168)			(0.0105)			(0.0120)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.470	0.477	0.477	0.427	0.432	0.434	0.306	0.312	0.312	0.332	0.334	0.336
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator (= 1if Int>Ext) in the right panel. OLS regressions use standard errors clustered at the class level. *** $p<.01$, ** $p<.05$, * $p<.1$.

Table 3.18 Results - GPA

	Int - Ext						Prob (Int >Ext)					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
External Assessment	-0.309***	-0.324***	-0.325***	-0.307***	-0.324***	-0.327***	-0.167***	-0.176***	-0.177***	-0.175***	-0.185***	-0.186***
	(0.00695)	(0.00699)	(0.00702)	(0.00635)	(0.00640)	(0.00639)	(0.00472)	(0.00476)	(0.00478)	(0.00451)	(0.00450)	(0.00449)
Female	0.148***	0.146***	0.146***	0.323***	0.321***	0.322***	0.0851***	0.0835***	0.0838***	0.171***	0.170***	0.170***
	(0.0102)	(0.0101)	(0.0101)	(0.0121)	(0.0120)	(0.0120)	(0.00744)	(0.00743)	(0.00744)	(0.00829)	(0.00823)	(0.00820)
Immigrant (1st gen)	-0.268***	-0.243***	-0.231***	-0.123***	-0.0957***	-0.0766**	-0.121***	-0.105***	-0.0989***	-0.0609***	-0.0457**	-0.0347*
	(0.0344)	(0.0339)	(0.0341)	(0.0322)	(0.0321)	(0.0321)	(0.0239)	(0.0237)	(0.0238)	(0.0211)	(0.0210)	(0.0210)
Immigrant (2nd gen)	-0.153***	-0.122***	-0.112***	-0.0895	-0.0531	-0.0415	-0.110***	-0.0912***	-0.0858***	-0.0320	-0.0119	-0.00518
	(0.0288)	(0.0289)	(0.0289)	(0.0587)	(0.0582)	(0.0582)	(0.0199)	(0.0200)	(0.0199)	(0.0369)	(0.0367)	(0.0366)
Undefined	-0.109***	-0.0943***	-0.0906***	-0.0379	-0.0283	-0.0222	-0.0506***	-0.0413**	-0.0394**	-0.0132	-0.00792	-0.00441
	(0.0221)	(0.0218)	(0.0216)	(0.0346)	(0.0344)	(0.0342)	(0.0173)	(0.0171)	(0.0171)	(0.0260)	(0.0259)	(0.0261)
Income = 2	0.175***	0.160***	0.159***	0.100***	0.0891***	0.0879***	0.0921***	0.0832***	0.0823***	0.0601***	0.0540***	0.0534***
	(0.0201)	(0.0201)	(0.0202)	(0.0245)	(0.0242)	(0.0241)	(0.0151)	(0.0152)	(0.0152)	(0.0165)	(0.0164)	(0.0163)
Income = 3	0.240***	0.212***	0.209***	0.152***	0.126***	0.122***	0.130***	0.112***	0.111***	0.0975***	0.0830***	0.0806***
	(0.0190)	(0.0189)	(0.0190)	(0.0214)	(0.0214)	(0.0215)	(0.0134)	(0.0134)	(0.0134)	(0.0147)	(0.0148)	(0.0148)
Income = 4	0.317***	0.270***	0.260***	0.206***	0.162***	0.151***	0.169***	0.140***	0.134***	0.117***	0.0925***	0.0857***
	(0.0160)	(0.0159)	(0.0161)	(0.0173)	(0.0180)	(0.0180)	(0.0112)	(0.0114)	(0.0114)	(0.0115)	(0.0120)	(0.0120)
Special Needs	-0.450***	-0.448***	-0.447***	-0.171***	-0.168***	-0.165***	-0.232***	-0.231***	-0.230***	-0.0673***	-0.0657***	-0.0640***
	(0.0269)	(0.0268)	(0.0268)	(0.0293)	(0.0293)	(0.0293)	(0.0182)	(0.0182)	(0.0182)	(0.0197)	(0.0197)	(0.0196)
Grade Retention	-0.380***	-0.361***	-0.358***	-0.576***	-0.560***	-0.555***	-0.217***	-0.205***	-0.203***	-0.284***	-0.275***	-0.272***
	(0.0398)	(0.0398)	(0.0398)	(0.0383)	(0.0381)	(0.0379)	(0.0295)	(0.0295)	(0.0295)	(0.0209)	(0.0209)	(0.0208)
2nd Tertile (ISEC)		0.122***	0.118***		0.0892***	0.0832***		0.0708***	0.0690***		0.0444***	0.0409***
		(0.0116)	(0.0116)		(0.0137)	(0.0137)		(0.00881)	(0.00882)		(0.00935)	(0.00938)
3rd Tertile (ISEC)		0.156***	0.151***		0.158***	0.147***		0.0978***	0.0951***		0.0904***	0.0843***
		(0.0120)	(0.0120)		(0.0147)	(0.0147)		(0.00897)	(0.00896)		(0.00974)	(0.00981)
Home language (Basque)			0.0847***			0.162***			0.0452***			0.0935***
			(0.0142)			(0.0172)			(0.0106)			(0.0124)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
R^2	0.454	0.461	0.463	0.378	0.384	0.388	0.313	0.318	0.319	0.276	0.280	0.283
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Dependent variables are the difference between internal and external evaluations in the left panel and it is a binary indicator (= 1if Int>Ext) in the right panel. GPA is the average of Math, Science, Basque, Spanish and English. OLS regressions use standard errors clustered at the class level. *** p <.01, ** p <.05, * p <.1.

Table 3.19 IV Results - Math

	Panel A: First stage						Panel B: Second stage					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Month of Birth	0.372***	0.367***	0.367***	0.211***	0.206***	0.207***						
	(0.0244)	(0.0241)	(0.0241)	(0.0235)	(0.0231)	(0.0232)						
External Assessment							-0.137**	-0.140***	-0.139***	-0.0557	-0.0579	-0.0573
							(0.0535)	(0.0539)	(0.0539)	(0.100)	(0.102)	(0.102)
Female	-0.101***	-0.111***	-0.110***	-0.0446***	-0.0587***	-0.0582***	0.0452***	0.0425***	0.0430***	0.307***	0.305***	0.306***
	(0.0161)	(0.0157)	(0.0158)	(0.0160)	(0.0156)	(0.0156)	(0.0139)	(0.0141)	(0.0141)	(0.0152)	(0.0158)	(0.0157)
Immigrant (1st gen)	-0.229***	-0.160***	-0.155***	-0.310***	-0.227***	-0.208***	-0.179***	-0.163***	-0.154***	-0.00286	0.00576	0.0141
	(0.0472)	(0.0463)	(0.0464)	(0.0396)	(0.0396)	(0.0396)	(0.0446)	(0.0435)	(0.0436)	(0.0496)	(0.0449)	(0.0442)
Immigrant (2nd gen)	-0.148***	-0.0664*	-0.0618	-0.0942	0.00300	0.0138	-0.120***	-0.0998***	-0.0925***	-0.0103	0.000353	0.00515
	(0.0397)	(0.0389)	(0.0389)	(0.0656)	(0.0642)	(0.0646)	(0.0361)	(0.0353)	(0.0353)	(0.0681)	(0.0677)	(0.0677)
Undefined	-0.112***	-0.0718**	-0.0702**	-0.0276	-0.00147	0.00429	-0.102***	-0.0922***	-0.0895***	-0.0184	-0.0156	-0.0130
	(0.0348)	(0.0330)	(0.0330)	(0.0516)	(0.0507)	(0.0498)	(0.0289)	(0.0284)	(0.0284)	(0.0477)	(0.0473)	(0.0474)
Income = 2	0.117***	0.0779**	0.0771**	0.0879***	0.0552*	0.0540*	0.128***	0.117***	0.116***	0.0796***	0.0762**	0.0756**
	(0.0313)	(0.0308)	(0.0309)	(0.0314)	(0.0306)	(0.0305)	(0.0283)	(0.0279)	(0.0279)	(0.0306)	(0.0300)	(0.0300)
Income = 3	0.186***	0.110***	0.109***	0.147***	0.0712**	0.0672**	0.194***	0.175***	0.172***	0.0918***	0.0838***	0.0820***
	(0.0292)	(0.0292)	(0.0292)	(0.0291)	(0.0287)	(0.0287)	(0.0275)	(0.0261)	(0.0261)	(0.0299)	(0.0269)	(0.0268)
Income = 4	0.311***	0.172***	0.168***	0.284***	0.156***	0.144***	0.266***	0.238***	0.231***	0.101***	0.0881***	0.0828***
	(0.0225)	(0.0224)	(0.0228)	(0.0224)	(0.0227)	(0.0228)	(0.0257)	(0.0219)	(0.0219)	(0.0358)	(0.0271)	(0.0266)
Special Needs	-0.557***	-0.524***	-0.524***	-0.666***	-0.623***	-0.619***	-0.324***	-0.320***	-0.318***	0.0622	0.0656	0.0680
	(0.0385)	(0.0383)	(0.0383)	(0.0443)	(0.0429)	(0.0431)	(0.0432)	(0.0423)	(0.0423)	(0.0770)	(0.0743)	(0.0738)
Grade Retention	-0.445***	-0.376***	-0.374***	-0.343***	-0.281***	-0.275***	-0.199***	-0.185***	-0.182***	-0.328***	-0.322***	-0.319***
	(0.0545)	(0.0535)	(0.0535)	(0.0422)	(0.0420)	(0.0419)	(0.0555)	(0.0544)	(0.0543)	(0.0568)	(0.0539)	(0.0536)
2nd Tertile (ISEC		0.213***	0.211***		0.220***	0.214***		0.0978***	0.0950***		0.0271	0.0241
		(0.0187)	(0.0187)		(0.0182)	(0.0181)		(0.0183)	(0.0182)		(0.0286)	(0.0281)
3rd Tertile (ISEC)		0.431***	0.428***		0.416***	0.405***		0.0824***	0.0781***		0.0446	0.0393
		(0.0193)	(0.0193)		(0.0190)	(0.0189)		(0.0278)	(0.0277)		(0.0464)	(0.0455)
Home language (Basque)			0.0379*			0.154***			0.0584***			0.0680***
			(0.0221)			(0.0243)			(0.0187)			(0.0251)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
Number of clusters							845	845	845	834	834	834
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Left panel presents the first-stage regression (dependent variable: normalized score of ext). Right panel includes the second-stage regression (dependent variable: normalized score of int). *** p <.01, ** p <.05, * p <.1. Instrument for external assessment is month of birth. It is normalized as to have value 1 if the student is born in January and 0 in December. Both regressions use class fixed effects and standard errors that are clustered at the class level.

Table 3.20 IV Results - Science

	Panel A: First stage						Panel B: Second stage					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Month of Birth	0.355***	0.351***	0.352***	0.216***	0.211***	0.212***						
	(0.0227)	(0.0225)	(0.0224)	(0.0239)	(0.0235)	(0.0236)						
External Assessment							-0.182***	-0.186***	-0.185***	-0.0279	-0.0295	-0.0287
							(0.0549)	(0.0552)	(0.0551)	(0.106)	(0.108)	(0.108)
Female	0.151***	0.142***	0.143***	0.0690***	0.0550***	0.0554***	0.0538***	0.0516***	0.0517***	0.251***	0.250***	0.250***
	(0.0156)	(0.0153)	(0.0152)	(0.0153)	(0.0151)	(0.0150)	(0.0157)	(0.0154)	(0.0154)	(0.0174)	(0.0168)	(0.0168)
Immigrant (1st gen)	-0.246***	-0.180***	-0.166***	-0.269***	-0.187***	-0.170***	-0.228***	-0.209***	-0.203***	-0.0210	-0.0134	-0.00264
	(0.0495)	(0.0484)	(0.0485)	(0.0426)	(0.0425)	(0.0425)	(0.0479)	(0.0471)	(0.0471)	(0.0544)	(0.0502)	(0.0495)
Immigrant (2nd gen)	-0.227***	-0.149***	-0.136***	-0.0254	0.0696	0.0794	-0.121***	-0.0974**	-0.0920**	-0.0842	-0.0748	-0.0686
	(0.0381)	(0.0380)	(0.0381)	(0.0747)	(0.0736)	(0.0738)	(0.0392)	(0.0380)	(0.0379)	(0.0769)	(0.0775)	(0.0775)
Undefined	-0.143***	-0.104***	-0.0997***	-0.0107	0.0148	0.0200	-0.0851***	-0.0737**	-0.0717**	-0.0214	-0.0189	-0.0156
	(0.0334)	(0.0326)	(0.0323)	(0.0552)	(0.0535)	(0.0527)	(0.0308)	(0.0301)	(0.0300)	(0.0470)	(0.0468)	(0.0472)
Income = 2	0.124***	0.0866***	0.0844***	0.0837**	0.0516	0.0505	0.155***	0.144***	0.143***	0.0648*	0.0618*	0.0611*
	(0.0305)	(0.0302)	(0.0303)	(0.0328)	(0.0320)	(0.0320)	(0.0281)	(0.0276)	(0.0277)	(0.0331)	(0.0325)	(0.0325)
Income = 3	0.185***	0.113***	0.109***	0.149***	0.0748***	0.0712**	0.235***	0.213***	0.212***	0.0731**	0.0662**	0.0638**
	(0.0282)	(0.0284)	(0.0283)	(0.0287)	(0.0286)	(0.0287)	(0.0266)	(0.0254)	(0.0255)	(0.0321)	(0.0292)	(0.0291)
Income = 4	0.334***	0.205***	0.193***	0.246***	0.119***	0.108***	0.287***	0.249***	0.244***	0.113***	0.101***	0.0944***
	(0.0229)	(0.0233)	(0.0234)	(0.0225)	(0.0227)	(0.0228)	(0.0266)	(0.0229)	(0.0229)	(0.0342)	(0.0268)	(0.0264)
Special Needs	-0.537***	-0.507***	-0.505***	-0.578***	-0.536***	-0.532***	-0.289***	-0.283***	-0.282***	-0.0748	-0.0715	-0.0684
	(0.0406)	(0.0401)	(0.0401)	(0.0431)	(0.0418)	(0.0419)	(0.0465)	(0.0454)	(0.0454)	(0.0730)	(0.0706)	(0.0702)
Grade Retention	-0.463***	-0.399***	-0.394***	-0.391***	-0.329***	-0.324***	-0.417***	-0.399***	-0.397***	-0.474***	-0.469***	-0.465***
	(0.0562)	(0.0563)	(0.0562)	(0.0425)	(0.0430)	(0.0430)	(0.0648)	(0.0634)	(0.0633)	(0.0659)	(0.0623)	(0.0620)
2nd Tertile (ISEC		0.224***	0.220***		0.205***	0.199***		0.0826***	0.0805***		0.0254	0.0216
		(0.0176)	(0.0175)		(0.0171)	(0.0170)		(0.0197)	(0.0196)		(0.0281)	(0.0277)
3rd Tertile (ISEC)		0.398***	0.391***		0.413***	0.403***		0.116***	0.112***		0.0379	0.0310
		(0.0185)	(0.0184)		(0.0191)	(0.0191)		(0.0271)	(0.0268)		(0.0483)	(0.0473)
Home language (Basque)			0.103***			0.139***			0.0433**			0.0877***
			(0.0224)			(0.0248)			(0.0183)			(0.0275)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
Number of clusters							845	845	845	834	834	834
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Left panel presents the first-stage regression (dependent variable: normalized score of ext). Right panel includes the second-stage regression (dependent variable: normalized score of int). *** p <.01, ** p <.05, * p <.1. Instrument for external assessment is month of birth. It is normalized as to have value 1 if the student is born in January and 0 in December. Both regressions use class fixed effects and standard errors that are clustered at the class level.

Table 3.21 IV Results - Basque

	Panel A: First stage						Panel B: Second stage					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Month of Birth	0.341***	0.336***	0.338***	0.222***	0.217***	0.217***						
	(0.0230)	(0.0229)	(0.0226)	(0.0233)	(0.0227)	(0.0227)						
External Assessment							-0.0775	-0.0798	-0.0766	0.0347	0.0360	0.0370
							(0.0559)	(0.0564)	(0.0561)	(0.0912)	(0.0933)	(0.0930)
Female	0.195***	0.185***	0.187***	0.280***	0.266***	0.267***	0.117***	0.115***	0.116***	0.192***	0.192***	0.192***
	(0.0159)	(0.0154)	(0.0153)	(0.0157)	(0.0154)	(0.0152)	(0.0160)	(0.0157)	(0.0157)	(0.0299)	(0.0293)	(0.0293)
Immigrant (1st gen)	-0.246***	-0.177***	-0.150***	-0.279***	-0.192***	-0.160***	-0.226***	-0.215***	-0.195***	-0.0796*	-0.0812*	-0.0672
	(0.0458)	(0.0449)	(0.0448)	(0.0374)	(0.0368)	(0.0363)	(0.0442)	(0.0429)	(0.0428)	(0.0482)	(0.0446)	(0.0433)
Immigrant (2nd gen)	-0.208***	-0.126***	-0.102***	-0.0827	0.0181	0.0366	-0.167***	-0.153***	-0.135***	-0.108	-0.111*	-0.102
	(0.0386)	(0.0381)	(0.0382)	(0.0619)	(0.0598)	(0.0598)	(0.0376)	(0.0367)	(0.0364)	(0.0677)	(0.0671)	(0.0672)
Undefined	-0.0976***	-0.0571*	-0.0484	-0.0486	-0.0215	-0.0116	-0.101***	-0.0945***	-0.0880***	-0.00565	-0.00617	-0.00184
	(0.0351)	(0.0338)	(0.0331)	(0.0532)	(0.0511)	(0.0506)	(0.0289)	(0.0285)	(0.0283)	(0.0442)	(0.0440)	(0.0434)
Income = 2	0.102***	0.0625**	0.0583**	0.100***	0.0664**	0.0643**	0.115***	0.108***	0.105***	0.0197	0.0204	0.0194
	(0.0287)	(0.0283)	(0.0284)	(0.0300)	(0.0292)	(0.0291)	(0.0247)	(0.0244)	(0.0244)	(0.0294)	(0.0289)	(0.0289)
Income = 3	0.213***	0.137***	0.130***	0.163***	0.0851***	0.0782***	0.145***	0.132***	0.126***	0.0877***	0.0895***	0.0864***
	(0.0279)	(0.0278)	(0.0276)	(0.0261)	(0.0260)	(0.0260)	(0.0254)	(0.0240)	(0.0240)	(0.0297)	(0.0270)	(0.0269)
Income = 4	0.369***	0.232***	0.210***	0.300***	0.167***	0.147***	0.207***	0.186***	0.168***	0.0697**	0.0732***	0.0643***
	(0.0218)	(0.0219)	(0.0219)	(0.0213)	(0.0216)	(0.0217)	(0.0280)	(0.0234)	(0.0226)	(0.0326)	(0.0251)	(0.0242)
Special Needs	-0.536***	-0.504***	-0.501***	-0.641***	-0.596***	-0.588***	-0.315***	-0.311***	-0.308***	0.103	0.103	0.107
	(0.0370)	(0.0371)	(0.0371)	(0.0438)	(0.0420)	(0.0419)	(0.0446)	(0.0437)	(0.0436)	(0.0692)	(0.0671)	(0.0663)
Grade Retention	-0.473***	-0.405***	-0.396***	-0.365***	-0.300***	-0.291***	-0.282***	-0.272***	-0.264***	-0.414***	-0.415***	-0.410***
	(0.0506)	(0.0508)	(0.0506)	(0.0394)	(0.0396)	(0.0397)	(0.0490)	(0.0472)	(0.0470)	(0.0627)	(0.0600)	(0.0595)
2nd Tertile (ISEC		0.229***	0.220***		0.226***	0.216***		0.0514***	0.0447**		0.00182	-0.00317
		(0.0173)	(0.0172)		(0.0174)	(0.0174)		(0.0191)	(0.0188)		(0.0263)	(0.0254)
3rd Tertile (ISEC)		0.420***	0.408***		0.432***	0.413***		0.0642**	0.0536**		-0.0135	-0.0225
		(0.0183)	(0.0183)		(0.0184)	(0.0181)		(0.0275)	(0.0267)		(0.0432)	(0.0414)
Home language (Basque)			0.196***			0.263***			0.143***			0.115***
			(0.0214)			(0.0245)			(0.0214)			(0.0330)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
Number of clusters							845	845	845	834	834	834
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Left panel presents the first-stage regression (dependent variable: normalized score of ext). Right panel includes the second-stage regression (dependent variable: normalized score of int). *** $p<.01$, ** $p<.05$, * $p<.1$. Instrument for external assessment is month of birth. It is normalized as to have value 1 if the student is born in January and 0 in December. Both regressions use class fixed effects and standard errors that are clustered at the class level.

Table 3.22 IV Results - Spanish

	Panel A: First stage						Panel B: Second stage					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Month of Birth	0.394***	0.390***	0.389***	0.236***	0.231***	0.231***						
	(0.0252)	(0.0251)	(0.0252)	(0.0256)	(0.0251)	(0.0251)						
External Assessment							-0.141***	-0.144***	-0.143***	-0.0317	-0.0319	-0.0306
							(0.0516)	(0.0518)	(0.0519)	(0.0952)	(0.0973)	(0.0971)
Female	-0.0535***	-0.0615***	-0.0622***	0.138***	0.123***	0.123***	0.296***	0.293***	0.294***	0.314***	0.314***	0.315***
	(0.0166)	(0.0162)	(0.0162)	(0.0164)	(0.0161)	(0.0161)	(0.0140)	(0.0140)	(0.0140)	(0.0203)	(0.0195)	(0.0195)
Immigrant (1st gen)	-0.112**	-0.0528	-0.0729	-0.172***	-0.0821**	-0.0875**	-0.280***	-0.264***	-0.256***	-0.0285	-0.0288	-0.0100
	(0.0542)	(0.0526)	(0.0526)	(0.0414)	(0.0415)	(0.0419)	(0.0404)	(0.0401)	(0.0403)	(0.0455)	(0.0437)	(0.0439)
Immigrant (2nd gen)	-0.156***	-0.0854**	-0.103**	-0.134**	-0.0296	-0.0328	-0.132***	-0.112***	-0.106***	0.0571	0.0566	0.0676
	(0.0426)	(0.0420)	(0.0421)	(0.0670)	(0.0660)	(0.0662)	(0.0363)	(0.0358)	(0.0359)	(0.0711)	(0.0694)	(0.0694)
Undefined	0.0136	0.0485	0.0422	-0.0272	0.000886	-0.000790	-0.123***	-0.113***	-0.111***	-0.0788*	-0.0790*	-0.0732
	(0.0366)	(0.0360)	(0.0362)	(0.0583)	(0.0563)	(0.0565)	(0.0294)	(0.0294)	(0.0294)	(0.0478)	(0.0478)	(0.0479)
Income = 2	0.0607*	0.0275	0.0306	0.0258	-0.00926	-0.00890	0.161***	0.151***	0.150***	0.105***	0.105***	0.104***
	(0.0317)	(0.0314)	(0.0314)	(0.0348)	(0.0337)	(0.0337)	(0.0278)	(0.0277)	(0.0278)	(0.0327)	(0.0326)	(0.0325)
Income = 3	0.148***	0.0834***	0.0888***	0.0991***	0.0181	0.0193	0.188***	0.169***	0.167***	0.106***	0.106***	0.102***
	(0.0294)	(0.0290)	(0.0290)	(0.0299)	(0.0296)	(0.0296)	(0.0259)	(0.0256)	(0.0257)	(0.0313)	(0.0299)	(0.0299)
Income = 4	0.254***	0.133***	0.149***	0.217***	0.0794***	0.0828***	0.240***	0.211***	0.205***	0.140***	0.140***	0.128***
	(0.0238)	(0.0237)	(0.0238)	(0.0249)	(0.0255)	(0.0255)	(0.0238)	(0.0218)	(0.0222)	(0.0320)	(0.0264)	(0.0265)
Special Needs	-0.568***	-0.539***	-0.541***	-0.664***	-0.618***	-0.619***	-0.332***	-0.327***	-0.326***	0.0929	0.0925	0.0980
	(0.0395)	(0.0392)	(0.0391)	(0.0442)	(0.0429)	(0.0429)	(0.0453)	(0.0443)	(0.0445)	(0.0749)	(0.0722)	(0.0722)
Grade Retention	-0.389***	-0.329***	-0.335***	-0.310***	-0.243***	-0.245***	-0.322***	-0.308***	-0.305***	-0.467***	-0.467***	-0.461***
	(0.0565)	(0.0566)	(0.0566)	(0.0448)	(0.0455)	(0.0454)	(0.0577)	(0.0568)	(0.0569)	(0.0618)	(0.0592)	(0.0593)
2nd Tertile (ISEC		0.165***	0.171***		0.239***	0.241***		0.0864***	0.0839***		-0.00405	-0.0108
		(0.0186)	(0.0185)		(0.0190)	(0.0190)		(0.0169)	(0.0170)		(0.0294)	(0.0295)
3rd Tertile (ISEC)		0.383***	0.392***		0.445***	0.448***		0.0822***	0.0784***		0.000320	-0.0117
		(0.0197)	(0.0197)		(0.0207)	(0.0208)		(0.0245)	(0.0250)		(0.0482)	(0.0485)
Home language (Basque)			-0.144***			-0.0446*			0.0538***			0.155***
			(0.0230)			(0.0256)			(0.0194)			(0.0242)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
Number of clusters							845	845	845	834	834	834
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Left panel presents the first-stage regression (dependent variable: normalized score of ext). Right panel includes the second-stage regression (dependent variable: normalized score of int). *** p <.01, ** p <.05, * p <.1. Instrument for external assessment is month of birth. It is normalized as to have value 1 if the student is born in January and 0 in December. Both regressions use class fixed effects and standard errors that are clustered at the class level.

Table 3.23 IV Results - English

	Panel A: First stage						Panel B: Second stage					
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Month of Birth	0.384***	0.379***	0.379***	0.187***	0.182***	0.182***						
	(0.0226)	(0.0225)	(0.0225)	(0.0219)	(0.0217)	(0.0217)						
External Assessment							-0.206***	-0.210***	-0.208***	0.0416	0.0443	0.0456
							(0.0471)	(0.0473)	(0.0473)	(0.0952)	(0.0982)	(0.0978)
Female	0.143***	0.135***	0.135***	0.242***	0.227***	0.227***	0.127***	0.125***	0.125***	0.137***	0.137***	0.137***
	(0.0160)	(0.0156)	(0.0156)	(0.0154)	(0.0151)	(0.0151)	(0.0138)	(0.0136)	(0.0136)	(0.0266)	(0.0259)	(0.0259)
Immigrant (1st gen)	-0.0184	0.0465	0.0450	-0.123***	-0.0317	-0.0318	-0.213***	-0.193***	-0.185***	-0.0920**	-0.0951***	-0.0798**
	(0.0460)	(0.0453)	(0.0454)	(0.0411)	(0.0403)	(0.0405)	(0.0394)	(0.0390)	(0.0390)	(0.0373)	(0.0354)	(0.0356)
Immigrant (2nd gen)	-0.0338	0.0434	0.0421	0.0823	0.188***	0.188***	-0.0716**	-0.0467	-0.0396	-0.185***	-0.189***	-0.180***
	(0.0371)	(0.0365)	(0.0366)	(0.0661)	(0.0643)	(0.0645)	(0.0339)	(0.0341)	(0.0341)	(0.0516)	(0.0547)	(0.0547)
Undefined	-0.0571*	-0.0190	-0.0194	-0.0202	0.00826	0.00825	-0.0418	-0.0297	-0.0271	-0.00862	-0.00959	-0.00487
	(0.0305)	(0.0301)	(0.0301)	(0.0492)	(0.0473)	(0.0474)	(0.0274)	(0.0269)	(0.0270)	(0.0393)	(0.0394)	(0.0397)
Income = 2	0.0975***	0.0608**	0.0611**	0.0966***	0.0610**	0.0610**	0.171***	0.159***	0.157***	0.0492*	0.0504*	0.0493*
	(0.0283)	(0.0282)	(0.0282)	(0.0311)	(0.0301)	(0.0301)	(0.0226)	(0.0223)	(0.0224)	(0.0279)	(0.0271)	(0.0270)
Income = 3	0.164***	0.0931***	0.0935***	0.166***	0.0840***	0.0840***	0.210***	0.187***	0.185***	0.0939***	0.0970***	0.0936***
	(0.0263)	(0.0262)	(0.0262)	(0.0272)	(0.0270)	(0.0270)	(0.0224)	(0.0214)	(0.0214)	(0.0277)	(0.0244)	(0.0244)
Income = 4	0.381***	0.251***	0.252***	0.323***	0.184***	0.184***	0.232***	0.196***	0.189***	0.0822**	0.0882***	0.0784***
	(0.0210)	(0.0209)	(0.0212)	(0.0216)	(0.0216)	(0.0216)	(0.0253)	(0.0214)	(0.0214)	(0.0348)	(0.0250)	(0.0250)
Special Needs	-0.463***	-0.432***	-0.432***	-0.653***	-0.607***	-0.607***	-0.460***	-0.454***	-0.452***	-0.0266	-0.0264	-0.0217
	(0.0364)	(0.0362)	(0.0362)	(0.0402)	(0.0389)	(0.0389)	(0.0364)	(0.0355)	(0.0356)	(0.0706)	(0.0683)	(0.0682)
Grade Retention	-0.444***	-0.379***	-0.379***	-0.458***	-0.390***	-0.390***	-0.247***	-0.229***	-0.226***	-0.290***	-0.292***	-0.287***
	(0.0476)	(0.0484)	(0.0484)	(0.0417)	(0.0417)	(0.0417)	(0.0485)	(0.0472)	(0.0472)	(0.0602)	(0.0567)	(0.0565)
2nd Tertile (ISEC		0.202***	0.203***		0.241***	0.241***		0.101***	0.0981***		0.00388	-0.00166
		(0.0168)	(0.0169)		(0.0169)	(0.0169)		(0.0169)	(0.0169)		(0.0283)	(0.0283)
3rd Tertile (ISEC)		0.404***	0.405***		0.452***	0.452***		0.108***	0.104***		-0.0243	-0.0342
		(0.0180)	(0.0181)		(0.0184)	(0.0185)		(0.0239)	(0.0239)		(0.0477)	(0.0476)
Home language (Basque)			-0.0111			-0.000354			0.0592***			0.126***
			(0.0205)			(0.0229)			(0.0165)			(0.0180)
N	15,802	15,802	15,802	15,381	15,381	15,381	15,802	15,802	15,802	15,381	15,381	15,381
Number of clusters							845	845	845	834	834	834
Class FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Education level	Primary	Primary	Primary	Middle	Middle	Middle	Primary	Primary	Primary	Middle	Middle	Middle

Notes: Left panel presents the first-stage regression (dependent variable: normalized score of ext). Right panel includes the second-stage regression (dependent variable: normalized score of int). *** p <.01, ** p <.05, * p <.1. Instrument for external assessment is month of birth. It is normalized as to have value 1 if the student is born in January and 0 in December. Both regressions use class fixed effects and standard errors that are clustered at the class level.

References

Aaronson, Daniel, Barrow, Lisa, & Sander, William. 2007. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, **25**(1), 95–135.

Abadie, Alberto, Athey, Susan, Imbens, Guido W, & Wooldridge, Jeffrey. 2017. *When Should You Adjust Standard Errors for Clustering?* Working Paper 24003. National Bureau of Economic Research.

Abdulkadiroğlu, Atila, & Sönmez, Tayfun. 2003. School Choice: A Mechanism Design Approach. *American Economic Review*, **93**(3), 729–747.

Abdulkadiroğlu, Atila, Pathak, Parag, Roth, Alvin E, & Sonmez, Tayfun. 2006. *Changing the Boston School Choice Mechanism*. Working Paper 11965. National Bureau of Economic Research.

Abdulkadiroğlu, Atila, Pathak, Parag A., Schellenberg, Jonathan, & Walters, Christopher R. 2020. Do Parents Value School Effectiveness? *American Economic Review*, **110**(5), 1502–39.

Abramitzky, Ran, Boustan, Leah Platt, & Eriksson, Katherine. 2016. *Cultural Assimilation during the Age of Mass Migration*. Working Paper 22381. National Bureau of Economic Research.

Acemoglu, Daron, & Angrist, Joshua. 2001. How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual 2000*, 9–74.

Adda, Jérôme, Pinotti, Paolo, & Tura, Giulia. 2020. There's More to Marriage than Love: The Effect of Legal Status and Cultural Distance on Intermarriages and Separations.

Agarwal, Nikhil, & Somaini, Paulo. 2018. Demand Analysis using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica*, **86**(2), 391–444.

Akerlof, George A, & Kranton, Rachel E. 2000. Economics and Identity. *The Quarterly Journal of Economics*, **115**(3), 715–753.

Akerlof, George A, & Kranton, Rachel E. 2002. Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, **40**(4), 1167–1201.

Alan, Sule, Ertac, Seda, & Mumcu, Ipek. 2018. Gender Stereotypes in the Classroom and Effects on Achievement. *Review of Economics and Statistics*, **100**(5), 876–890.

Alan, Sule, Duysak, Enes, Kubilay, Elif, Mumcu, Ipek, *et al.* 2020. *Social Exclusion and Ethnic Segregation in Schools: The Role of Teacher's Ethnic Prejudice*. Tech. rept.

Alesina, Alberto, Carlana, Michela, Ferrara, Eliana La, & Pinotti, Paolo. 2018. *Revealing Stereotypes: Evidence from Immigrants in Schools*. Tech. rept. National Bureau of Economic Research.

Algan, Yann, Mayer, Thierry, & Thoenig, Mathias. 2013. The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns Across France.

Allport, Gordon Willard. 1954. *The Nature of Prejudice*. Addison-Wesley Reading.

Almagro, Milena, & Andrés-Cerezo, David. 2020. The Construction of National Identities. *Theoretical Economics*, **15**(2), 763–810.

Ângelo, Catarina, & Reis, Ana Balcão. 2017. *Gender Gaps in Different Grading Systems*. Tech. rept. FEUNL Working Paper Series.

Anghel, Brindusa, Cabrales, Antonio, Sainz, Jorge, & Sanz, Ismael. 2015. Publicizing the Results of Standardized External Tests: Does it Have an Effect on School Outcomes? *IZA Journal of European Labor Studies*, **4**(1), 7.

Anghel, Brindusa, Cabrales, Antonio, & Carro, Jesus M. 2016. Evaluating a Bilingual Education Program in Spain: the Impact beyond Foreign Language Learning. *Economic Inquiry*, **54**(2), 1202–1223.

Arai, Mahmood, & Skogman Thoursie, Peter. 2009. Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings. *Journal of Labor Economics*, **27**(1), 127–147.

Arellano, Manuel, & Zamarro, Gema. 2007. The Choice between Public and Private Schools with or Without Subsidies in Spain. *Preliminary and incomplete preprint*.

Arenas, Andreu, & Calsamiglia, Caterina. 2020. The Design of University Entrance Exams and its Implications for Gender Gaps. *Unpublished Manuscript*.

Arregi, A, Martínez, P, Sainz, A, & Ugarriza, JR. 2009. *Efecto de las repeticiones de curso en el proceso de enseñanza-aprendizaje del alumnado*. Tech. rept. Gobierno Vasco: Instituto Vasco de Evaluación e Investigación Educativa.

Aspachs-Bracons, Oriol, Clots-Figueras, Irma, Costa-Font, Joan, & Masella, Paolo. 2008. Compulsory Language Educational Policies and Identity Formation. *Journal of the European Economic Association*, **6**(2-3), 434–444.

Bedard, Kelly, & Dhuey, Elizabeth. 2006. The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics*, **121**(4), 1437–1472.

Ben-Shakhar, Gershon, & Sinai, Yakov. 1991. Gender Differences in Multiple-Choice Tests: The Role of Differential Guessing Tendencies. *Journal of Educational Measurement*, **28**(1), 23–35.

Bergman, Peter, & McFarlin Jr, Isaac. 2020. *Education for All? A Nationwide Audit Study of Schools of Choice*. Discussion Paper 13007. Institute of Labor Economics (IZA).

Berniell, Inés, & Estrada, Ricardo. 2020. Poor Little Children: the Socioeconomic Gap in Parental Responses to School Disadvantage. *Labour Economics*, 101879.

Beuermann, Diether, Jackson, C Kirabo, Navarro-Sola, Laia, & Pardo, Francisco. 2019. What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output.

Billings, Stephen B, Deming, David J, & Rockoff, Jonah. 2014. School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg. *The Quarterly Journal of Economics*, **129**(1), 435–476.

Bisin, Alberto, & Tura, Giulia. 2019. *Marriage, Fertility, and Cultural Integration in Italy*. Tech. rept. National Bureau of Economic Research.

Bisin, Alberto, & Verdier, Thierry. 2001. The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic theory*, **97**(2), 298–319.

Bisin, Alberto, & Verdier, Thierry. 2011. The Economics of Cultural Transmission and Socialization. *Pages 339–416 of: Handbook of Social Economics*, vol. 1. Elsevier.

Black, Sandra. 1999. Do Better Schools Matter? Parental Valuation of Elementary Education. *The Quarterly Journal of Economics*, **114**(2), 577–599.

Bobba, Matteo, & Frisancho, Veronica. 2014. *Learning About Oneself: The Effects of Signaling Academic Ability on School Choice*. Tech. rept. Inter-American Development Bank.

Böhlmark, Anders, Holmlund, Helena, & Lindahl, Mikael. 2016. Parental Choice, Neighbourhood Segregation or Cream Skimming? An Analysis of School Segregation after a Generalized Choice Reform. *Journal of Population Economics*, **29**(4), 1155–1190.

Burgess, Simon, & Greaves, Ellen. 2013. Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics*, **31**(3), 535–576.

Burgess, Simon, Greaves, Ellen, Vignoles, Anna, & Wilson, Deborah. 2015. What Parents Want: School Preferences and school Choice. *Economic Journal*, **125**(587), 1262–1289.

Calsamiglia, Caterina, & Loviglio, Annalisa. 2016. Maturity and School outcomes in an Inflexible System: Evidence from Catalonia. *SERIEs*, 1–49.

Calsamiglia, Caterina, & Loviglio, Annalisa. 2019. Grading on a curve: When Having Good Peers Is Not Good. *Economics of Education Review*, **73**, 101916.

Calsamiglia, Caterina, Haeringer, Guillaume, & Klijn, Flip. 2010. Constrained School Choice: An Experimental Study. *American Economic Review*, **100**(4), 1860–74.

Calsamiglia, Caterina, Fu, Chao, & Güell, Maia. 2020. Structural Estimation of a Model of School Choices: The Boston Mechanism versus Its Alternatives. *Journal of Political Economy*, **128**(2), 642–680.

Card, David, Domnisoru, Ciprian, & Taylor, Lowell. 2018. *The Intergenerational Transmission of Human Capital: Evidence from the Golden Age of Upward mobility*. Working Paper 25000. National Bureau of Economic Research.

Carlana, Michela. 2019. Implicit Stereotypes: Evidence from Teachers' Gender Bias. *The Quarterly Journal of Economics*, **134**(3), 1163–1224.

Chen, Yan, & Kesten, Onur. 2017. Chinese College Admissions and School Choice Reforms: A Theoretical Analysis. *Journal of Political Economy*, **125**(1), 99–139.

Chetty, Raj, Friedman, John N, Hilger, Nathaniel, Saez, Emmanuel, Schanzenbach, Diane Whitmore, & Yagan, Danny. 2011. How Does Your Kindergarten Classroom Affect your Earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, **126**(4), 1593–1660.

Chetty, Raj, Friedman, John N, Saez, Emmanuel, Turner, Nicholas, & Yagan, Danny. 2017. *Mobility Report Cards: The Role of Colleges in Intergenerational Mobility*. Working Paper 23618. National Bureau of Economic Research.

Clotfelter, Charles T, Ladd, Helen F, & Vigdor, Jacob. 2005. Who Teaches Whom? Race and the Distribution of Novice Teachers. *Economics of Education review*, **24**(4), 377–392.

Clots-Figueras, Irma, & Masella, Paolo. 2013. Education, Language and Identity. *Economic Journal*, **123**(570), F332–F357.

Cosnefroy, Olivier, & Rocher, Thierry. 2004. Le Redoublement au Cours de la Scolarisé Obligatoire: Nouvelles Analyses, Mêmes Constats. *Éducation et Formations*, 73–82.

de la Rica, Sara, & Ortega, Francesc. 2012. Cultural Integration in Spain. *Chap. 5 of:* et al, Yann Algan (ed), *Cultural Integration of Immigrants in Europe*. Oxford University Press.

Di Liberto, Adriana, & Casula, Laura. 2016. *Teacher Assessments Versus Standardized Tests: Is Acting 'Girly' an Advantage?* Tech. rept. IZA Discussion Paper.

Diaz-Serrano, Luis, & Meix-Llop, Enric. 2016. Do Schools Discriminate Against Homosexual Parents? Evidence from a Randomized Correspondence Experiment. *Economics of Education Review*, **53**, 133–142.

Echeverria, Begoña. 2003. Schooling, Language, and Ethnic Identity in the Basque Autonomous Community. *Anthropology & Education Quarterly*, **34**(4), 351–372.

Eklöf, Hanna. 2007. Test-Taking Motivation and Mathematics Performance in TIMSS 2003. *International Journal of Testing*, **7**(3), 311–326.

Elder, Todd E, & Lubotsky, Darren H. 2009. Kindergarten Entrance Age and Children's Achievement Impacts of State Policies, Family Background, and Peers. *Journal of Human Resources*, **44**(3), 641–683.

Eurydice. 2017. Grade Retention During Compulsory Education in Europe.

Ewens, Michael, Tomlin, Bryan, & Wang, Liang Choon. 2014. Statistical discrimination or Prejudice? A Large Sample Field Experiment. *Review of Economics and Statistics*, **96**(1), 119–134.

Farré, Lídia, Ortega, Francesc, & Tanaka, Ryuichi. 2015. *Immigration and School Choices in the midst of the Great Recession*. Discussion Paper 9234. Institute of Labor Economics (IZA).

Fenoll, Ainhoa Aparicio, Campaniello, Nadia, & Monzon, Ignacio. 2019. Parental Love Is Not Blind: Identifying Selection into Early School Start.

Figlio, David N, & Lucas, Maurice E. 2004. What's In A Grade? School Report Cards and the Housing Market. *American Economic Review*, **94**(3), 591–604.

Fouka, Vasiliki. 2019. How Do Immigrants Respond to Discrimination? The Case of Germans in the US during World War I. *American Political Science Review*, **113**(2), 405–422.

Fouka, Vasiliki. 2020. Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I. *The Review of Economic Studies*, **87**(1), 204–239.

Fouka, Vasiliki, Mazumder, Soumyajit, & Tabellini, Marco. 2018. *From Immigrants to Americans: Race and Assimilation during the Great Migration*. Working Paper 19-018. Harvard Business School.

Fryer Jr, Roland G, & Levitt, Steven D. 2004. The Causes and Consequences of Distinctively Black Names. *The Quarterly Journal of Economics*, **119**(3), 767–805.

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

Gardeazabal, Javier. 2011. Linguistic Polarization and Conflict in the Basque Country. *Public Choice*, **149**(3-4), 405.

Giulietti, Corrado, Tonin, Mirco, & Vlassopoulos, Michael. 2019. Racial Discrimination in Local Public Services: A Field Experiment in the United States. *Journal of the European Economic Association*, **17**(1), 165–204.

Glaeser, Edward L. 2005. The Political Economy of Hatred. *The Quarterly Journal of Economics*, **120**(1), 45–86.

Glazerman, Steven, & Dotter, Dallas. 2017. Market Signals: Evidence on the Determinants and Consequences of School Choice from a Citywide Lottery. *Educational Evaluation and Policy Analysis*, **39**(4), 593–619.

Goldin, Claudia, & Rouse, Cecilia. 2000. Orchestrating Impartiality: The Impact of "blind" Auditions on Female Musicians. *American Economic Review*, **90**(4), 715–741.

Goldstein, Joshua R, & Stecklov, Guy. 2016. From Patrick to John F. Ethnic Names and Occupational Success in the Last Era of Mass Migration. *American Sociological Review*, **81**(1), 85–106.

Gortazar, Lucas, Mayor, David, & Montalban, José. 2020. *School Choice Priorities and School Segregation: Evidence from Madrid*. Working Paper 1/2020. Swedish Institute for Social Research.

Haeringer, Guillaume, & Klijn, Flip. 2009. Constrained School Choice. *Journal of Economic theory*, **144**(5), 1921–1947.

Hanna, Rema N, & Linden, Leigh L. 2012. Discrimination In Grading. *American Economic Journal: Economic Policy*, **4**(4), 146–68.

Harris, Douglas N, & Larsen, Matthew. 2015. What Schools Do Families Want (and Why). *Policy Brief (New Orleans, LA: Education Research Alliance for New Orleans*.

Hastings, Justine, Kane, Thomas J, & Staiger, Douglas O. 2009. Heterogeneous Preferences and the Efficacy of Public School Choice. *NBER Working Paper*, **2145**, 1–46.

Hastings, Justine S, & Weinstein, Jeffrey M. 2008. Information, School choice, and Academic Achievement: Evidence from Two Experiments. *The Quarterly Journal of Economics*, **123**(4), 1373–1414.

Hattie, John. 2008. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*.

He, Yinghua. 2016. Gaming the Boston School Choice Mechanism in Beijing.

Hortaçsu, Ali, & McAdams, David. 2010. Mechanism Choice and Strategic Bidding in Divisible Good Auctions: An Empirical Analysis of the Turkish Treasury Auction Market. *Journal of Political Economy*, **118**(5), 833–865.

Hoxby, Caroline, & Turner, Sarah. 2013. *Expanding College Opportunities for High-achieving, Low income Students*. Discussion Paper 12-014. Stanford Institute for Economic Policy Research.

Hwang, Sam. 2016. A Robust Redesign of High School Match. *EAI Endorsed Trans. Serious Games*, **3**(11), e5.

Ikeda, Miyako, & García, Emma. 2014. Grade Repetition: A Comparative Study of Academic and Non-Academic Consequences. *OECD Journal: Economic Studies*, **2013**(1), 269–315.

Jackson, C Kirabo. 2014. Teacher Quality at the High School Level: The Importance of Accounting for Tracks. *Journal of Labor Economics*, **32**(4), 645–684.

Kapor, Adam J., Neilson, Christopher A., & Zimmerman, Seth D. 2020. Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*, **110**(5), 1274–1315.

Kinsler, Josh, Pavan, Ronni, & DiSalvo, Richard. 2014. *Distorted beliefs and parental investment in children*. Tech. rept. University of Rochester, Human Capital and Economic Opportunity Working Group.

Kline, Patrick M, & Walters, Christopher R. 2020. Reasonable Doubt: Experimental Detection of Job-Level Employment Discrimination. *Econometrica*.

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

Lerman, Steven, & Manski, Charles. 1981. On the Use of Simulated Frequencies to Approximate Choice Probabilities. *Structural Analysis of Discrete Data with Econometric Applications*, **10**, 305–319.

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

Mizala, Alejandra, & Urquiola, Miguel. 2013. School Markets: The Impact of Information Approximating Schools' Effectiveness. *Journal of Development Economics*, **103**, 313–335.

Murillo, Francisco Javier, & Martínez-Garrido, Cynthia. 2018. Magnitud de la Segregación Escolar por Nivel Socioeconómico en España y sus Comunidades Autónomas y Comparación con los Países de la Unión Europea. *Revista de Sociología de la Educación-RASE*, **11**(1), 37–58.

Murillo, Francisco Javier, Martínez-Garrido, Cynthia, & Belavi, Guillermina. 2017. Segregación escolar por Origen Nacional en España. *OBETS: Revista de Ciencias Sociales*, **12**(2), 395–423.

Nelder, John A, & Mead, Roger. 1965. A Simplex Method for Function Minimization. *The Computer Journal*, **7**(4), 308–313.

OECD. 2016. PISA 2015 Results (Volume I). Excellence and Equity in Education.

O'Neil, Harold F, Abedi, Jamal, Miyoshi, Judy, & Mastergeorge, Ann. 2005. Monetary Incentives for Low-Stakes Tests. *Educational Assessment*, **10**(3), 185–208.

Pettigrew, Thomas F, & Tropp, Linda R. 2006. A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology*, **90**(5), 751.

Pfaff, Steven, Crabtreet, Charles, Kern, Holger L., & Holbein, John B. 2018. *Does Religious Bias Shape Access to Public Services? A Large-Scale Audit Experiment Among Street-Level Bureaucrats*. Working Paper. Center for Open Science.

Press, William H, Teukolsky, Saul A, Vetterling, William T, & Flannery, Brian P. 1997. *Numerical Recipes in Fortran 77 and Fortran 90: Source Code for Recipes and Example Programs*. Cambridge University Press Cambridge, UK.

Qureshi, Javaeria A, & Ost, Ben. 2020. The Role of Families in Student Sorting to Teachers. *Journal of Human Resources*, **55**(2), 470–503.

REDE, (Red por el Diálogo Educativo). 2020. La financiación del Sistema Educativo: Invertir en Calidad, Equidad e Inclusión.

Rothstein, Jesse. 2010. Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement. *The Quarterly Journal of Economics*, **125**(1), 175–214.

Rothstein, Jesse M. 2006. Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition Among Jurisdictions. *American Economic Review*, **96**(4), 1333–1350.

Save the Children. 2019. Todo lo que Debes Saber de PISA 2018 sobre Equidad: La Equidad Educativa en España y sus Comunidades Autónomas en PISA 2018.

Save the Children Spain. 2019. Todo lo que Debes Saber de PISA 2018 sobre Equidad: La Equidad Educativa en España y sus Comunidades Autónomas en PISA 2018. Anexo Euskadi.

Schindler, David, & Westcott, Mark. 2020. Shocking Racial Attitudes: Black G.I.s in Europe. *The Review of Economic Studies*, **(Forthcoming)**.

Schüller, Simone. 2015. Parental Ethnic Identity and Educational Attainment of Second-Generation Immigrants. *Journal of Population Economics*, **28**(4), 965–1004.

Shayo, Moses. 2009. A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review*, **103**(2), 147–174.

Söderström, Martin, & Uusitalo, Roope. 2010. School choice and Segregation: Evidence from an Admission Reform. *Scandinavian Journal of Economics*, **112**(1), 55–76.

Spolaore, Enrico, & Wacziarg, Romain. 2016. Ancestry, Language and Culture. *Pages 174–211 of: The Palgrave Handbook of Economics and Language*. Springer.

Terrier, Camille. 2020. Boys Lag Behind: How Teachers' Gender Biases Affect Student Achievement. *Economics of Education Review*, **77**, 101981.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Vega-Bayo, Ainhoa, & Mariel, Petr. 2019. A Discrete Choice Experiment Application to School Choice in the Basque Country. *Hacienda Pública Española*, **230**(3), 41–62.

Voigtländer, Nico, & Voth, Hans-Joachim. 2015. Nazi Indoctrination and Anti-Semitic Beliefs in Germany. *Proceedings of the National Academy of Sciences*, **112**(26), 7931–7936.

Zamarro, Gema, Hitt, Collin, & Mendez, Ildefonso. 2019. When Students Don't Care: Reexamining International Differences in Achievement and Student Effort. *Journal of Human Capital*, **13**(4), 519–552.

Zimmer, Ron W, & Guarino, Cassandra M. 2013. Is there Empirical Evidence that Charter Schools "push out" Low-performing Students? *Educational Evaluation and Policy Analysis*, **35**(4), 461–480.