Essays in Economics of Education and Skills

DOCTORAL DISSERTATION

BY

Lucas Gortazar

Facultad de Economía y Empresa

Departamento de Fundamentos del Análisis Económico II

Universidad del País Vasco-Euskal Herriko Unibertsitatea

Supervisor: Sara de la Rica

June, 2018

Acknowledgements

I would like to thank the Department of Mathematics of the University of the Basque Country, the Department of Mathematics of University College London, and CEMFI Research Center for their critical contribution to my pre-doctoral training. Furthermore, I am also truly thankful to the Basque Government and Fundación la Caixa for the scholarships for my post-graduate Masters at UCL and CEMFI respectively. I am highly indebted to the Department of Fundamentos de Análisis Económico II at the University of the Basque Country for providing key support in this dissertation. Finally, I want to thank Fedea Research Center for their generosity in hosting me while conducting this research.

This journey started ten years ago, at the beginning of the Great Recession, when I first started reading and learning about economics and social sciences. Since then, many people have played a key role in leading me to this point and although the signing is individual, I owe this work to a great group of people. It is an act of justice to express my gratitude to them for their generosity and kindness in the fulfillment of this thesis, which I wish they would also consider as theirs.

I would like to take this opportunity to thank Elena Iñarra for her initial help and enthusiasm and for building a personalized bridge to navigate from mathematics to social sciences. To Lucho Breceda and Lluís Gascón for our endless conversations and readings, back in 2009.

I want to thank the CEMFI crowd who walked a tough path together with me, with especial mention to Luis Diez-Catalán, Daniel Kutner, Avichal Mahajan and José Alonso.

I am also deeply grateful to Juan Manuel Moreno, Alberto Rodriguez and many other great professionals at the World Bank who offered me a great opportunity to learn and who transmitted the rigor and art of how to understand and make education policy happen.

Special thanks go to Nacho Plata, for the constant challenge to test my priors. And to my co-authors David Mayor and Pepe Montalbán, for this beautiful relay race that we have come across together for the last months.

And above all, I would like to express my gratitude to my advisor Sara de la Rica, for her outstanding support, generosity, courage and confidence in my ability to achieve this project. Without her guidance and help, this work would not have been possible at all. Thank you!

This thesis is dedicated to my parents and my sister. They have always made me feel that everything I have achieved in life was as important for them as it was for me. I am especially proud to finish this project now, in the summer of 2018, in the 40th anniversary of my father's graduation in this same faculty.

Finally, and most important, the thesis is dedicated to Miriam, my partner since the very beginning of this journey. She should feel the results of this work as hers, as she is truly responsible for the passion and commitment to achieve my goals.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, social scientists have had access to a wide range of quantitative data from administrative records and surveys in order to better understand the dynamics of human capital formation at the school and the workplace. Skills development for students and workers are at the center of public policies. Understanding the drivers and consequences of education policies as well as the dynamics of the labor market is becoming a more achievable task thanks to those data sources and therefore contributes to a more rigorous approach of social sciences. At the same time, inequalities in student and workers opportunities persist and, if anything, recent social and economic developments have increasingly generated challenges for policy-makers in developed and developing countries.

In education policy, emerging questions are now being answered with a wealth of quantitative data from administrative sources as well as from national and international assessments. The field of economics of education has opened up new layers of promising research. Moreover, the discipline is now in a better position to answer old and new questions with broader experience in terms of identifying causal effects of specific policy developments. For example, the continuity of national and international assessments worldwide in recent decades has allowed a wide range of research questions on the determinants of learning outcomes to be addressed. At the same time, administrative data records collected by education authorities in charge of managing school networks serve as primary sources to understand student dynamics across the schools within a specific district. In short, more research is now available with respect to enrollment patterns, demand for schooling, or school network policies.

With respect to the economics of skills and jobs, changes in labor demand patterns in recent decades are raising new questions with regard to what to expect in terms of the future labor force and the new skills that are required in the workplace. Technological change, together with other social and economic forces, is altering the demand for new jobs, while other jobs are moved out of the labor market. Such changes are transforming the tasks performed by workers in advanced economies, generating winners and losers. In the last two decades, the research on skills and labor economics has sought to identify the skill requirements at work by means of surveys on work task requirements. This has helped to better understand more in depth the conceptualization of the new job task requirements presented in the seminal work by Autor et al. (2003).

This thesis seeks to contribute to these two lines of economic research with three essays of empirical analysis that exploit administrative and survey data from students and workers. Chapter 2, Chapter 3 and Chapter 4 correspond to each of the three essays and Chapter 5 revisits the findings from those three chapters, discusses the contributions and limitations of the analysis, and concludes with a more open discussion on the policy implications.

In Chapter 2, "The PISA shock in the Basque Country", I explore the reasons why the Basque Country region in Northern Spain experienced a great setback in the 2015 secondary education PISA (Programme for International Student Assessment) results. The Basque education system had been a top performer among Spanish regions in the past, both in national and international assessments, and its relevance in terms of post-secondary education enrollment is comparable to that of Northern European education systems. Hence, understanding this decline becomes a key question from a policy point of view, even so when the PISA results were accompanied by declines in regional assessments implemented by the Basque authorities (which validate the concerns of policy-makers and society regarding structural changes in the Basque education system). In particular, I explore the reasons of the decline by looking at socio-economic factors (given the impact of the Great Recession in the region), along with school factors in the Basque education system. I pay specific attention to factors that may be related to the very nature of the system, such as its linguistic education policy model, the coexistence of public and private provision of education services or the high rates of repetition experienced by the region compared to OECD and EU averages. I use decomposition techniques derived from the labor economics literature, to both understand average changes in performance between 2015 with respect to previous editions (2009 and 2012), as well as changes across the distribution of scores.

Chapter 3, ”School Choice, Student Mobility and School Segregation in Madrid”, addresses a key historical question from the economics of education literature, which relates school choice and student sorting. Over the last four decades, there has been extensive academic debate in the field of economics on whether more choice should be granted to students in school districts or school networks. Whereas proponents argue for the benefits of market-based education policies as drivers of improvement of outcomes, opponents claim that excessive choice leads to social sorting and segregation due to the diverse responses of households to reforms. While the first argument (improvement of performance) presents mixed evidence and points to the importance of contextual and implementation factors, the second discussion (sorting of students) seems to show evidence in the direction of more segregation of students across schools. Nevertheless, the evidence on this last argument presents validity issues: a recent literature discussion by Epple, Romano, and Urquiola (2017) argues that the evidence from large-scale programs has clear identification issues, as these are usually implemented country-wide. In my work with David Mayor and José Montalbán[1], we try to address this research gap by studying a large-scale reform in the administrative region of Madrid, which provides education services to more than one million students. In particular, the city of Madrid shifted in 2013 to an inter-district single school zone for students entering the system. By using a unique and high-quality administrative record of student application data, we study the response of households to the reform. We do so by, first, looking at mobility patterns of students, paying special attention to heterogeneous responses by social and immigrant status of households. Consequently, we measure the overall impact in terms of student segregation (or sorting) according to such student characteristics. To better describe the policy effects, we consider a diverse set of reasons (income of the district, school ownership, or student composition of demanded schools) that may be behind the changes in responses and the new composition of students across schools. An important contribution is that we are able use other municipalities that implemented the reform at the same time (or the year before) as a control group, in order to strengthen our identification strategy.

[1]This chapter is co-authored with David Mayor (Solchaga and Recio Asociados) and José Montalbán (Paris School of Economics). All authors contributed equally to the production and writing of the chapter.

In the research presented in Chapter 4, and jointly conducted with Sara de la Rica[2], we use the OECD's survey on adult skills PIAAC (Programme for International Assessment of Adult Competencies) to address differences on job tasks across a harmonized sample of 20 countries across the world for the first time by using very precise information on job contents at worker level. This implies an important measurement contribution of the task content of jobs, given that this literature usually builds from occupation-level data and overlooks within-occupation heterogeneity of the task content of jobs (Autor, 2013). Moreover, additional data on worker ability (which allows to control for unobserved heterogeneity) is available. First, as a validity analysis, the task measures constructed with PIAAC data are compared at the occupation level to those from O*NET, the most well-known data source in the literature. In addition, the relationship between job contents and individual and job characteristics as well as with computer use at work, both across and within occupations, is assessed. Finally, we consider the relationship between job tasks and wages, assuming that workers self-select into occupations based on comparative advantage.

[2]This chapter is co-authored with Sara de la Rica (UPV-EHU and Fedea). Both authors contributed equally to the production and writing of the chapter.

Chapter 2

The PISA "shock" in the Basque Country

2.1 Introduction

The cognitive skills of secondary school students in the Basque Country (Spain), measured by the OECD Programme for International Student Assessment (PISA), experienced a significant decline in 2015 relative to previous rounds. In particular, results published in December 2016 displayed two key relevant findings. First, 15-year old students of the Basque education system (born in 1999) obtained lower levels of reading, mathematics and scientific skills than their peers three and six years before, with statistically significant declines in mathematics (13 PISA points), and especially science (23 PISA points). Second, for the first time, this implied that the Basque regional education system was for the first time below the average mean scores in Spain for two out of the three competencies assessed in PISA (science and reading).

The PISA outcomes for the Basque country were not a surprise for the regional authorities and stakeholders. Following the introduction of a new Education Act (LOE (3/2006)), the Basque Country developed its own external diagnosis assessments (*Evaluación Diagnóstica*) in 2009, consisting of a competency-based test organized for both primary (Year 4) and secondary (Year 2 of lower secondary education, the equivalent to Year 8) students, which were run every two years. The 2015 results of such regional assessments which were published in early 2016 already showed a decline in several assessed competencies (language, mathematics and scientific competencies) for both primary and secondary students taking the test[1]. Although the cohorts of students taking the assessments did not correspond to the same one which sat in PISA (most of them being in Year 10 in 2015), these trends were in the same direction as the ones observed in PISA outcomes in 2015.

The PISA assessment has shaped public debates and promoted a new paradigm in the global education discussion and the governance of education systems (see Sellar and Lingard (2014)). Since it was first conducted in 2000, there have been numerous examples of education systems being highly influenced by unexpected PISA results through the public discussion and the policy direction taken by governments around the world (see Breakspear (2012)). The results in the Basque Country (which has participated with enough schools to generate its own representative sample since 2003) generated significant attention in the media for weeks, and the main issue raised by the coverage was to understand the reasons for the decline.

This paper explores such reasons and sheds light on the factors behind the decline, by looking at student, school and system factors. I construct a harmonized dataset for the Basque country for the five rounds in which the region has participated so far (2003, 2006, 2009, 2012, and 2015). The PISA student assessment has been historically accompanied by a student and school questionnaire, hence allowing for an analysis of the weight of underlying factors in shaping student outcomes. The periodicity of certain key student and school background questionnaire items is not safely guaranteed, and hence requires common items to be selected across years for consistent comparisons.

The Basque education system is characterized by two school networks (public and privately run) of equal size which are both funded by the Department of Education of the Basque Government. Being a multilingual region (Basque and Spanish, with the latter being the main language of use), the system of instructionis organized into three different linguistic models. Finally, the system has experienced two important changes in the last decade, such as the arrival of a large number of immigrant families and the social and economic consequences of the Great Recession.

Considering all these and other factors, I find that the decline in test scores in 2015, relative to previous editions is mostly unrelated to changes in socioeconomic characteristics of students. Conversely, I find three important factors to be behind the decline

[1] See ISEI-IVEI (2016b) and ISEI-IVEI (2016a).

in student outcomes since 2009 and 2012: the rate of student repeaters, the percentage of students taking the test in a language different to the one at home, and the changes in the perceived disciplinary climate and student behavior at school by the school principal. Depending on the baseline year (2009 or 2012) and the domain of comparison (science, mathematics or reading), the estimated difference effects account for between 20% and 90% of the total observed decline. Hence, an important proportion of the decline remains unexplained.

The rest of the paper is organized as follows. Section 2.2 discusses the literature related to changes in education quality in education systems using cross-section data. Section 2.3 describes the Basque education system and its most recent context. The PISA data collection process and the construction of harmonized dataset is detailed in Section 2.4. The results and its analysis are shown in Section 2.5. Section 2.6 provides a final discussion and concludes.

2.2 Related Literature

This paper is well related to the literature which has sought to address country changes (or cross-country differences) in learning outcomes. The literature mostly uses data from international assessments, such as PIRLS, TIMSS or PISA, and analyzes differences in student outcomes across countries or years by isolating the explanatory power of each observable factor that can affect student learning. Ammermueller (2007) finds ambiguous results in explaining the differences in PISA performance between Finland and Germany. Additionally, other authors have studied the gap differences between countries or regions (Nieto and Ramos (2015); Ojima and Von Below (2010)). Other studies have analyzed the nature of certain observable gaps within countries: Ramos et al. (2012) analyze the dynamics of rural-urban gaps in Colombia using PISA data and find that most of the differences are attributable to family characteristics as opposed to those of the school. Finally, another group of studies has looked at the differences across years for specific countries. Barrera-Osorio et al. (2011) look at the increase in PISA performance in Indonesia between 2003 and 2006 and find that almost the entire test score increase is explained by the returns to characteristics, mostly related to student age. However, the authors find that the adequate supply of teachers also plays a role in test score changes. In Bulgaria, Gortazar et al. (2014) look at the large improvements experienced in PISA scores between 2006 and 2012: they find that the improvements of student socioeconomic characteristics as well as the improvements in school resources account for most of the performance changes.

The decomposition methods have often been used in the labor economics literature to understand the nature of wage gaps or wage dynamics across time. In the case of the education literature, decomposition studies have also looked at mean differences in student learning outcomes using the Oaxaca (1973) method. Nevertheless, from an education policy perspective, decomposition methods at the mean may end up missing key information that is fundamental to understand student learning dynamics. Recent empirical methods in the labor economics literature have brought new econometric techniques to analyze wage gaps and wage inequality by computing counter-factual decompositions throughout the whole wage distribution (DiNardo et al. (1996), Machado and Mata (2005) and Firpo et al. (2007)). These methods have the advantage of introducing in a less restrictive assumption on the relation between observable independent variables and the outcome variable of interest: that is, that factor differences across groups or changes across years may not be affecting outcomes similarly across the distribution.

A similar assumption can hold for the education production function literature: that education policies and student and family factors may have a different influence on learning across student performance distribution, and that the changes across years in such factors may not be the same for all students. For example, it may well be the case that after the increase of immigrant population in the Basque school system, the learning outcomes of immigrant and non-immigrant students may have changed due to changes in the peer effect mechanisms, the composition of immigrant population (with different cultural and language backgrounds) at schools, or the organization of teaching within schools with a sudden large share of immigrant students. Hence, while student covariates may change across years (a larger share of immigrant population), the relation between immigrant population and student outcomes may change from a distributional perspective too.

The Firpo et al. (2007) decomposition has already been used in the decomposition literature with education production functions. For example, Lounkaew (2013) uses a Firpo-Fortin-Lemieux approach to decompose differences between rural and urban student performance in Thailand using PISA 2009 data. Moreover, Gouss and Ledonn (2014) look at the changes in the inequities of performance in France between 2000 and 2009 following a similar approach. They find that disadvantaged socioeconomic background has become even more penalizing to learning in 2009 compared to 2000. In addition, two educational policies seem to explain a large part of the rise in the decrease of low-achieving students: changes in sorting practices (through repetition policies) and changes in special education policies. Finally, the study from Gortazar et al. (2014) in Bulgaria extends the mean decomposition estimation to quantiles, and finds that among low-performing students, the improvement of school resources accounted for a large share of the increase, whereas among high-performing students, the improvement of socioeconomic conditions and the changes in peer composition at schools accounted for most of the learning improvements.

2.3 The Basque Education System and its Context

This section describes the key elements of the Basque education system, the recent trends and background factors which have occurred in the last years, and the extent to which these may have had an incidence in student learning outcomes. The Basque school system is a self-managed and self-financed model through its own tax system[2], although it is subject to Spanish regulation for basic issues[3]. In the school year 2014/2015, the system served 369,000 students between kindergarten and upper secondary (both general and vocational) education[4]. The system is organized around two distinct school networks of equal size, both publicly funded, and one being publicly managed (*Centros Públicos*) whereas the other is privately run (*Centros concertados*) [5].

Although the admission criteria for students are similar in both school networks and depend on public regulation set by the Basque government, there are three important

[2]The three provinces of the Basque Country, along with the region of Navarra, are granted constitutional rights to have their own tax system, through which they finance their public services and for which they pay an estimated amount covering the proportion of competencies that are not transferred by the central government.

[3]According to the Spanish Constitution, the Spanish central government regulates the conditions for granting and providing education diplomas and develops the basic norms for the implementation of Article 27, which recognized simultaneously the right to education and the freedom of education and teaching.

[4]Source: EUSTAT.

[5]Moreover, a minority of 0.8% of students attend privately funded schools, which represents the lowest share among all Spanish autonomous communities.

differences between the two school networks. First, the school management criteria differ: privately run schools are legal entities with budgetary autonomy to distribute all resources (given a set of rules fixed by the public administration) within the school and to hire and fire staff following the standard labor code. Conversely, public schools are managed by the Department of Education, with all staff mostly being civil servants[6], and the principal being a civil servant teacher appointed by a committee comprising representatives of the education authority and the school board in question. The second main difference is that the access to privately run schools is in practice not fully free of charge: although parents are not meant to pay for basic education services, in practice, the lack of sufficient public funding generates the needs of those schools to obtain alternative sources of funding (Rogero-García and Andrés-Candelas (2018)). In particular, given that charging fees in these schools is not legal, schools expect households to contribute with private donations or symbolic fees which end up acting as an entry barrier. This means *de facto* that the privately run school network of schools acts as a semi-public service rather than a full public service, a definition (semi-public) that we adopt following Calsamiglia and Güell (2018)[7]. Partially because of this, the third difference has to do with the differences in student composition: privately run schools serve on average a more socially advantaged population and a much smaller share of immigrant children, whereas public schools serve a more disadvantaged student population, including most of the immigrant children in the Basque Country[8].

In recent years, the system has experienced three important phenomena that have affected the structure of the Basque education and hence, may have impacted the results on student outcomes. First, as a bilingual region where two languages (Spanish and Basque) coexist, the Basque country organizes its education system along three different school streams in terms of language of instruction: the A model (Spanish as the main language of instruction and Basque being taught as a single subject), the B model (which balances the weight of both languages in terms of hours of instruction) and the D model (Basque being the main language of instruction, and Spanish being taught as a single subject). Since the early 2000, the number of schools with Basque as main language of instruction (model D) has increased significantly as part of an effort by the Basque public administration to increase the percentage of bilingual citizens. By 2015,

[6] A small proportion of public school teachers hold fixed-term appointments.

[7] According to the Basque statistical agency EUSTAT on the recent data available in 2012, parents paid an average annual fee of €707.6 for basic education services (without considering complementary activities or services) to semi-public schools in the Basque education system.

[8] Source: Departamento de Educación Política Lingüística y Cultura: Basque Government (2016)

almost two thirds of students in lower secondary schools studied under the Basque language immersion model (through the D model).

This means that, given the linguistic demography of the region (with almost 75% of the population declaring Spanish as the mother tongue and main language used at home[9]), an increasing share of students with Spanish as mother tongue attend schools with a Basque immersion program under the D model. Whereas the advantages of being fluent in both languages have been identified from the cognitive psychology literature (bilingual education, especially early in life, has a long-term positive impact on cognitive development[10]) and labor economics literature (speaking the two languages increases job opportunities in adult life[11]), these can be offset in the short-term by a decrease of learning outcomes, given the positive effects that mother tongue instruction (usually in primary) has on short-term academic outcomes (Ivlevs and King (2014)). Moreover, beyond the language of instruction, the language used to administer the test can also have implications on student outcomes. This is what is found in the different exploratory analyses conducted by the Basque Institute for Research and Evaluation in Education (ISEI-IVEI), a public agency answering to the Basque Department of Education. Using data from previous PISA rounds and the regional assessments, students enrolled in the linguistic D model with Spanish as the language used at home display better outcomes when taking the test in Spanish as opposed to taking it in Basque (ISEI-IVEI (2004); ISEI-IVEI (2012)).

The other major background event in the Basque education system is the surge in foreign-born children living in the Basque country and therefore using the Basque education system. In particular, the proportion of immigrant students rose from 2% in 2002 to almost 8.4% in 2015[12], with many of these students coming from countries whose language has not Latin origin and they therefore have lower cultural language links with the native population[13]. This phenomenon has increased the pressure on the system to provide quality education to a more diverse student population, and at the

[9]Source:EUSTAT

[10]See Costa and Sebastián-Gallés (2014)

[11]Cappellari and Di Paolo (2015) find positive results to bilingual education in Catalonia (Spain), a region which like the Basque Country also implemented Catalan immersion programs at schools. Prior to that, Angrist and Lavy (1997) find positive returns to bilingual education in the labor market in Morocco.

[12]Source: EUSTAT

[13]Source: Basque Government Departamento de Educación Política Lingüística y Cultura: Basque Government (2016). In the school year 2014/2015, there were 35,804 foreign-born students in the basic education schools, out of which only 8,899 (around 25%) were born in Latin American countries and 5,270 were born in EU countries (around 15%), whereas 14,436 (more than 40%) where born in African countries.

same time may have altered the peer effects dynamics between students from different backgrounds, as well as the enrollment dynamics of families when the vast majority of children enter schools at age 3.

Even though the Basque Country is one of the richest autonomous communities in Spain, the Great Recession also greatly impacted the Basque economy (and hence, on employment and public finances) when compared to other European countries or regions. According to the Basque Government's statistical Agency (EUSTAT) the regional GDP decreased by 7.5 percentage points between 2008 and 2013, whereas the unemployment rate boomed from 3.5% to 16.6% in the same period. In terms of education financing, the decline of economic activity and the pressure on autonomous communities from Spanish and European authorities to meet deficit goals led to a decrease in the regional public executed education budget of %15.5 between 2009 and 2015[14]. These two factors (unemployment and budget cuts) may have had significantly negatively impacted learning outcomes that could first be observed in education assessments such like PISA 2015, and would not be a surprise according to recent evidence. On the one hand, evidence from another autonomous community in Spain (Catalonia) has shown a causal relation between parental job loss and the decline of student performance (Ruiz-Valenzuela (2015)). On the other hand, Jakcson et al. (2018) have recently shown that the spending cuts that have occurred in the US during the Great Recession are causally linked to a decrease in student outcomes and graduation rates.

The changes in these important factors (linguistic models, immigration, economic crisis) may have been important in shaping the results of the student population in the Basque education system observed in national and international assessments. Beyond these, another two important characteristics that are idiosyncratic to the Basque education system are its large share of pre-primary enrollment and the prevalence of grade repetition. In the Basque country, more than 90% of children start kindergarten as early as age 2, and almost 50% of children are already attending childcare centers at age 1[15]. These numbers place the Basque country region as high as Denmark or Iceland, the OECD countries with the largest enrollment rates in kindergarten(OECD (2016a)[16]).

[14] According to the Basque Department of Finance, the public executed budget in education by the Basque Government, which includes services from kindergarten to tertiary education, decreased from €3,021 million in 2009 to €2,552 million in 2015, with the lowest investment made in 2014, with €2,441 million. For more information see `aurrekontuak.irekia.euskadi.eus`.

[15] Source: Ministry of Education, Culture and Sports (Spain).

[16] Source: See Education at a Glance 2017

The other key feature of the system is grade repetition. Grade repetition is a prevalent policy in Southern European education systems due to cultural inheritance of assessment practices at the school level (Eurydice (2011)), and the Basque Country is not exempted from large share of repeaters, both when compared to OECD and EU standards. Grade repetition has shown to be an inefficient policy, which does not obtain expected results in terms of providing adequate opportunities to students lagging behind (Hattie (2008); Manacorda (2012); Jacob and Lefgren (2009)). García-Pérez et al. (2014) find for the Spanish case that grade retention has a negative impact on educational outcomes. At the same time, in Spain in general, and in the Basque Country in particular, students from the most disadvantaged backgrounds are the ones that bear the negative costs of repeating, as the policy disproportionately affects this group, no matter what their cognitive skills are (OECD (2014)).

2.4 Data and Variables

2.4.1 The PISA Dataset

I use the data from the Programme for International Student Assessment (PISA) in 2015, a triennial assessment which provides information on countries and economies' education systems. It does by means of an assessment taken by 15-year olds which focus on three domains (reading, mathematics and science), and which does not only ascertain whether students can reproduce knowledge; it also examines how well students can extrapolate, reflect, evaluate and communicate what they have learned and can apply it in unfamiliar settings, in different contexts inside and outside the school. This approach is set to be key in the context of modern economies, which rewards individuals not for what they know, but for what they can do with it.

In the PISA 2015 wave, for example, 71 countries and an additional set of subnational regions (such as the Basque Country) participated. The data included a country random sample of usually 100 to 150 schools per country, with around 30 students per school responding to a two-hour test with multiple choice and open questions. In each edition, the test focused on one of the three domains of study: in 2000 it started with the reading domain, in 2003 it continued with the mathematics domain and in 2006 it moved to the scientific domain. Following this order, the 2015 edition focused for the second time on the scientific domain.

PISA, in the same way as other large scale studies, uses an Item Response Theory (IRT) approach to transform the student item responses into competency scales. It does so by providing different plausible values for each subject area derived from taking random draws out of the distribution resulted from the IRT scaling process. This is because not all students respond to all items and domains (in order to keep their attention on the test) and hence the OECD derives an estimated probability distribution out of the items answered. For the first time, PISA 2015 edition displayed a model resulting in 10 plausible values for each of the three domains, an improvement with respect to previous editions, as it enlarged the usual 5 plausible values of the previous PISA rounds. I use plausible values in the analysis when feasible, which allows me to produce more consistent standard errors. The estimates that I produce in the next section are hence derived from computing average estimates of each parameter obtained in regression models on plausible value scores, as well as the adjusted standard errors of these estimates[17].

Moreover, PISA provides rich data from the student and school background questionnaires. Students answer questions related to their family and home, their views about life, their school, their schedule and learning time, their learning experience and their views about learning. Principals, on the other hand, answer questions related to the school background information, the school management, the teaching staff, the assessment and evaluation practices, the organization of learning across student groups, and the school disciplinary and behavioral climate. These two questionnaires cover a wide range of questions for students and principals to answer. However, the questionnaires have not been identical across years: in particular, many student and principal questions focus on the domain of study at each PISA round (reading, mathematics or science) and the learning and teaching practices around that domain, which means that many items of the questionnaire are not comparable across all years. Other questions simply change the item answers by restructuring them or substituting some of them by new adjusted categories. Hence, in order to construct a harmonized and comparable dataset, I restrict myself to use item questions that are strictly comparable across years, so that these are domain-free (or domain related for the case of science comparisons between 2006 and 2015) and they follow the same answer categories.

[17] I do so by using the REPEST command in STATA (Avvisati and Keslair (2016))

2.4.2 Variable Description

I examine several dimensions that affect learning and which have been well studied in previous empirical studies using national and international data sources. I also consider idiosyncratic variables of the Basque education system described in the previous section. To do this, I group variables by student characteristics, school characteristics and systemic characteristics.

With respect to student characteristics, I include student's gender, age (through month of birth), language at home, and the nationality of birth. I include the Index of Economic, Social, and Cultural Status (ESCS), constructed by the OECD based on parental education, occupation and home possessions, to consider the socioeconomic dimension. Moreover, I consider variables related to student's prior education history, such as the fact of having repeated one or two grades in the past. I do not include the variable of years in pre-primary education, as there was a change in the question structure for 2015[18].

The second group of variables includes school characteristics, behavior and climate and learning and teaching practices. First, I consider the average socioeconomic status of students in the school, which I compute by calculating the average score of the ESCS Index for all students in a specific school. I also consider student wellbeing variables, such as items related to learning anxiety, learning motivation and sense of belonging at the school. I consider teacher and student behavior variables. For the case of scientific questions (which I observe in 2006 and 2015 because of the focus on the science domain), I include variables related to perceptions of scientific learning, reported by students, such as self-efficacy in science, or science learning enjoyment.

The third group of variables considers elements which are observed at the school level but determined at the system level. I first pick the two main idiosyncratic variables, which at the same time are the two variables through which the sample stratification is conducted: the ownership of the school (public or semi-public) and the linguistic model of instruction (A, B or D) in the school. Moreover, I consider other dimensions such as the level of autonomy of the school (measured by the two OECD

[18]The attending pre-primary school variable in 2015 is organized in terms of age of entrance to kindergarten. This represents a change with respect to previous editions, where this variable was framed by the number of years of schooling before primary education, and hence does not allow for rigorous comparisons.

indexes of autonomy on resources and pedagogic and curriculum issues), the school location (whether it is urban or rural), the school size or the level of school resources (which allows changes in the school budget in the years of the economic crisis to be proxied).

2.4.3 Harmonizing Variables and Indexes

The data preparation requires comparable variables or indexes to be used. To exploit variation in several dimensions, I use OECD's indexes on different student and school dimensions, which transform a set of questionnaire items into one single index dimension through a parametric model[19]. I exploit the variation of such indexes to better identify differences across years. These indexes (the Index of Economic, Social and Cultural Status (ESCS) is the best-known) have the goal of synthesizing different questionnaire items around a latent dimension. They are created independently for each PISA round. They usually take a mean value of zero and a standard deviation of one for the yearly distribution of OECD countries. Hence, these are constructed as relative scales for which the values taken by students on a specific country or region directly depend on the students' responses in other countries that same year. This means these scales are not comparable within the same education system that from a time perspective.

I use an alternative approach to make these indexes comparable for a given education system, such as the Basque country. First, I identify item questions that are repeated across years, so that indexes are time-consistent. Although in general I follow the OECD recommendations, for some cases I have to discard certain items which are not present in different years. Beyond a change to the index values because of the rescaling, this implies certain ordinal discrepancies with respect to the values given by OECD official indexes. That is, for example, in 2012, it may be the case that a Basque student is placed in the 99^{th} percentile of the ESCS index in the Basque country according to the OECD index value, but is then moved to the 98^{th} percentile in 2012 when I re-calibrate the index to make it time-consistent. Nevertheless, as a robustness check, I provide correlations between the re-calibrated index values and the OECD index values for each index and year.

[19] A more detailed description of each index item can be found in the Appendix

Once the items are identified, I follow a similar approach as OECD (OECD (2012)), and compute indexes using Item Response Theory (IRT) models, Rasch models (a particular version of IRT) or just the aggregation of positive responses (for the school responsibility indexes). I perform this calibration by computing parameters jointly for all years in the Basque country with equal weight for each year. Hence, the parameters transforming items into indexes are common for all Basque students and schools which have participated in PISA 2015 and all the previous editions.

For the case of the ESCS index, the index is computed through a Rash model (Rasch (1960)) of dichotomized items [20] based on three sub-indexes which, following the OECD recommendations, are aggregated through a principal component analysis. These three sub-indexes include parental education (PARED: the highest educational level of parents was also recoded into estimated number of years of schooling), parental job occupation (HISEI: highest occupational status of parents measured by the ISEI occupation score) and a home possessions index (HOMEPOS, a summary index including items from different dimensions of goods owned at home [21]). I use imputation methods for missing values of one of the three ESCS sub-dimensions[22].

Table 2.1 shows the structure of the ESCS index (socioeconomic status) once it has been re-calibrated for the Basque Country. The first two columns show the differences of the two Indexes (the OECD and the re-calibrated Index). While the OECD Index decreased between 2012 and 2015, this was due to the OECD full sample values rather than to relevant changes in the Basque Country. In particular, results display a stagnation of the re-calibrated index between 2012 and 2015, but not a decline. In particular, the ESCS components show a decrease in the index of home possessions, an increase of the Index occupational status, and a very mild decrease of the highest parental education level. What can be observed is a decline of socioeconomic conditions of the

[20] The items usually take several categorical responses, but following OECD recommendation, I reduce those categories to only two.

[21] The HOMEPOS index is derived of items of three sub-dimensions: WEALTH (possessions of durable goods at home, such as cars, TVs, computers, rooms at home), CULTPOS (cultural possessions such as classical literature, poetry or works of art) and HEDRES (home educational resources such as a desk, computer, a room to study, textbooks or educational software) as well as books at home. I recode all the four, five or six categorical variables into two-level categorical variables to compute a Rasch model. A more detailed description of the HOMEPOS Index construction can be found in Table 2.8 in the Appendix.

[22] Regarding the missing values for students with missing data for only one variable of the three sub-dimensions, I follow OECD recommendation (OECD (2012)) and impute predicted values plus a random component based on a regression on the missing variable of the other two variables. Variables with imputed values were then used for a principal component analysis. If there were missing data on more than one variable, ESCS was not computed for that case and a missing value was assigned for ESCS.

students in the lowest percentiles of the distribution since 2009, particularly $p10$ and to a lesser extent $p25$, whereas there are small increases of socioeconomic conditions of students in the median and the upper part of the ESCS-distribution. Overall, it can be seen that since 2009 the standard deviation of the re-calibrated ESCS Index has increased, which is consistent with other social and economic indicators. Hence, one would expect an increase in outcomes dispersion related to the observed increases in dispersion statistics of student ESCS. Finally, there should be an important note of caution when looking at this indicator in the Basque Country (and in general, Souther European economies): the fact that the index does not capture the employment status of parents overlooks important variation of income and labor dynamics in households, given the large surges in unemployment rates observed in recent years. In the Basque Country, this pattern is even larger in households with children (Save the Children Spain (2017)).

Beyond ESCS, I re-compute the rest of the indexes, which can be seen in Table 2.2, and for the years for which the data is available: (i) Index of Sense of Belonging at school (BELONG); (ii) Index of Science Enjoyment (JOYSCIE); (iii) Index of Science Self-Efficacy (SCIEEFF); (iv) Index of School Autonomy on Resources (RESPRES); (v) Index of School Autonomy on Pedagogics and Curriculum (RESPCURR); (vi) Index of Instructional Leadership (LEADINST); (vii) Index of Instructional Improvement and Professional Development promoted by the principal (LEADPD); (viii) Index of Teacher Participation in Leadership (LEADTCH); (ix) Index of Student behavior (STUBEHA); (x) Index of Teacher Behavior (TEACHBEHA). A summary of the items used for each index can be found in Figure 2.3 of the Appendix.

TABLE 2.1: Evolution of ESCS and its Statistics in the Basque Country since 2003.

	OECD ESCS		Re-calibrated ESCS for Basque Country									
	Mean	S.D	Mean	HOMEPOS	HISEI	PARED	S.D	p10	p25	p50	p75	p90
2003	-0.10	0.94	-0.29	-0.16	45.21	12.13	1.020	-1.66	-1.02	-0.31	0.47	1.09
2006	-0.04	0.99	-0.08	0.01	47.92	12.58	0.990	-1.38	-0.81	-0.10	0.67	1.26
2009	-0.08	0.97	0.09	0.04	49.54	13.52	0.910	-1.15	-0.54	0.13	0.78	1.25
2012	0.03	0.93	0.14	0.03	51.30	13.69	0.980	-1.18	-0.61	0.21	0.96	1.41
2015	-0.25	1.08	0.15	0.00	52.49	13.67	1.020	-1.27	-0.64	0.23	1.04	1.45
Total	-0.09	0.00	-0.02	49.31	13.13	1.000	-1.35	-0.74	0.02	0.83	1.32	

Notes: Values are computed after re-scaling of all indexes. Re-calibrated ESCS index is the result of a principal component analysis, with replacement of missing values based on linear regressions on the other two observations.

TABLE 2.2: Index Items and Years.

Year	2003	2006	2009	2012	2015
ESCS	X	X	X	X	X
BELONG	X			X	X
JOYSCIE		X			X
SCIEEFF		X			X
RESPRES		X	X	X	X
RESPCURR		X	X	X	X
LEADINST				X	X
LEADPD				X	X
LEADTCH				X	X
STUBEHA	X		X	X	X
TEACHBEHA	X		X	X	X

I compute an IRT model for the Index of Sense of Belonging at School, the Index of Science Enjoyment at School, the Index of Science Self-Efficacy, the Index of Instructional Leadership, the Index of Instructional Improvement and Professional Development, the Index of Teacher Participation in Leadership, the Index of Student Behavior, and the Index of Teacher Behavior. For the school autonomy indexes, I follow the OECD procedure and compute the proportion of ”yes” answers for school governing board, principal or teachers reporting responsibility on specific issues vis a vis the yes responses for regional/local education authority or national educational authority with respect to the different dimensions of school management responsibility.

In Table 2.3, I display the year correlation between OECD indexes and the results of the recalibration process of the indexes for the Basque Country. As can be seen, the correlation is very high (above 0.9) for most dimensions, including ESCS, BELONG, JOYSCIE, SCIEEFF, LEADINST, LEADPD, LEADTCH, STUBEHA and TEACHBEHA. For the case of the indexes of school autonomy the results are slightly weaker, and usually stand between 0.8 and 0.9. For the case of ESCS, given the lack of some dimensions in 2003 I also recomputed the index as a robustness check only using 2006 and the following years (without considering students in 2003): the results are slightly better than the ones shown in Table 2.3, with year correlations with OECD Index values of 0.977 in 2006, 0.967 in 2009, 0.976 in 2012, and 0.967 in 2015.

TABLE 2.3: Comparison of re-calibrated Items and OECD Values, by Year.

Year	2003	2006	2009	2012	2015
ESCS	0.971	0.967	0.950	0.965	0.955
BELONG	0.978			0.866	0.981
JOYSCIE		0.966			0.997
SCIEEFF		0.936			0.954
RESPRES		0.874	0.944	0.840	0.915
RESPCURR		0.868	0.867	0.871	0.880
LEADINST				0.948	0.957
LEADPD				0.965	0.952
LEADTCH				0.995	0.996
STUBEHA	0.970		0.967	0.946	0.960
TEACHBEHA	0.916		0.918	0.881	0.969

Notes: Year correlations include all the sample of available data constructed by both OECD and in this re-calibration process in the Basque Country.

2.5 Results

2.5.1 Empirical Strategy

I compute linear and non-linear models to decompose changes in student outcomes in the Basque Country between 2015 and previous years 2003, 2006, 2009 and 2012. I use both Oaxaca-Blinder decomposition (Oaxaca (1973)) for the mean and the method proposed by Firpo et al. (2009) for unconditional quantiles. As previously discussed, the reason to conduct non-linear decompositions is that it allows for a less restrictive assumption regarding changes in covariates and their relation to student outcomes: it may well be the case that the changes in the composition of socioeconomic status have been different across the distribution of student outcomes, hence having a differential impact on the performance distribution. Following the classical Oaxaca-Blinder decomposition, I start by looking at a linear model in Equation 2.1:

$$Y_{ij} = \alpha + X_i^{'}\beta + P_j^{'}\theta + S_j^{'}\gamma + \epsilon_i \tag{2.1}$$

where Y_{ij} represents the competency score in the domain of interest of student i in school j, $X_i^{'}$ is a vector of individual student characteristics, $P_j^{'}$ is a vector of school characteristics, whereas $S_j^{'}$ is a vector of school j characteristics determined at the system level and ϵ_i is an error term. When comparing two groups of students at different years (year t_0 and year t_1), the difference of the PISA performance is computed as:

$$\begin{aligned} Y_{ij,t_1} - Y_{ij,t_0} = (\alpha_{t_1} - \alpha_{t_0}) + (X'_{i,t_1}\beta_{t_1} - X'_{i,t_0}\beta_{t_0}) + (P'_{j,t_1}\theta_{t_1} - P'_{j,t_0}\theta_{t_0}) + \\ (S'_{j,t_1}\gamma_{t_1} - S'_{j,t_0}\gamma_{t_0}) + (\epsilon_{i,t_1} - \epsilon_{i,t_0}) \end{aligned} \tag{2.2}$$

Adding and subtracting $X'_{i,t_0}\beta_{t_1}$, $P'_{j,t_0}\theta_{t_1}$, and $S'_{j,t_0}\gamma_{t_1}$ on the previous Equation 2.2 and rearranging terms gives the expression:

$$\begin{aligned} Y_{i,t_1} - Y_{i,t_0} = \{(X'_{i,t_1} - X'_{i,t_0})\beta_{t_1} + (P'_{j,t_1} - P'_{j,t_0})\theta_{t_1} + (S'_{j,t_1} - S'_{j,t_0})\gamma_{t_1})\} + \\ \{X'_{i,t_0}(\beta_{t_1} - \beta_{t_0}) + P'_{j,t_0}(\theta_{t_1} - \theta_{t_0}) + S'_{j,t_0}(\gamma_{t_1} - \gamma_{t_0}) + (\alpha_{t_1} - \alpha_{t_0}) + (\epsilon_{i,t_1} - \epsilon_{i,t_0})\} \end{aligned} \tag{2.3}$$

The previous equation represents the two-fold Oaxaca decomposition. It shows a first term in brackets which is the explained part of the differences: i.e. the difference due to difference in observed covariates across years. The second term in brackets is considered the unexplained part, and consists of the differences in returns to inputs (differences in the learning outcome premium of each of the covariates) and an error term representing unobserved characteristics which the model cannot capture. Following Lounkaew (2013), I do not report the results of the estimates of returns to covariates for each year (say, for example, β_{t_1} and β_{t_0}) because of their difficult interpretation. Hence, the interpretation of the model is only related to changes in student and school characteristics across years, but not to changes in returns to learning (i.e. the relation between these characteristics and student learning). To control for unobserved heterogeneity, I add control variables of student and school characteristics that may be correlated with the error term in order to minimize the impact of omitted variables in the model.

This model considers the returns to student and school characteristics in year t_1 to be the baseline returns, but alternative specifications of the Oaxaca-Blinder model and the organization of coefficient weights can be used. In a more general form, the model considers a linear combination of the two vectors of coefficients (β_{t_1} and β_{t_0}), so that the explained part of Equation 2.3 is multiplied by the following term b^*:

$$b^* = W\beta_{t_1} + (I - W)\beta_{t_0} \tag{2.4}$$

where I is an identity matrix and W is a matrix of weights of two groups of year t_1 and year t_0. The Equation 2.3 is re-written (given time vectors of covariates X_{t_1} and X_{t_0}) so that the difference in scores across years R is:

$$R = (X_{t_1} - X_{t_0})'\{W\beta_{t_1} + (I - W)\beta_{t_0}\} + \{X'_{t_1}(I - W) + X'_{t_0}W\}(\beta_{t_1} - \beta_{t_0}) \tag{2.5}$$

In the context of the OLS regression, I follow the Neumark (1988) approach with the coefficient $W = (X'_{t_1}X_{t_1} + X'_{t_0}X_{t_0})^{-1}(X'_{t_1}X_{t_1})$, which basically displays the parameters resulting of computing a two-fold decomposition pooled model excluding the year variable out of the model.

As previously discussed, the second step of the empirical strategy is to extend the OLS analysis to other distributional statistics. I do this by using a RIF-regression approach, proposed by Firpo et al. (2009). The problem about quantile regression is that the marginal effect of a specific variable cannot be generalized to an unconditional interpretation, fixing the rest of covariates constant. While this is possible in linear models for the mean due to the Law of Iterated Expectations, it does not apply for non-linear statistics. The proposal of Firpo, Fortin and Lemieux (FFL) is to replace the score function by a linearized function which approximates a statistic of interest (quantile, inequality statistic, etc...) and hence allows marginal effects to be interpreted for a specific quantile as unconditional effects without the need of constructing conditional effect estimate functions. For a quantile of interest q_τ, the RIF function is written as:

$$RIF(I; q_\tau) = q_\tau + \frac{\tau - D(I \leq q_\tau)}{f_I(q_\tau))} \tag{2.6}$$

where D is an indicator function, $f_I(.)$ is the density of the marginal distribution of scores. In practice, we do not observe $RIF(I; q_\tau)$, but just its sample version, which can be written as:

$$RIF(I; \hat{q_\tau}) = \hat{q_\tau} + \frac{\tau - D(I \leq \hat{q_\tau})}{f_I\hat{(}q_\tau)} \tag{2.7}$$

where $\hat{q_\tau}$ is the sample quantile and $\widehat{f_I(q_\tau)}$ is the density estimator kernel function. Hence, as part of the second step of the empirical strategy, I compute regular regressions using RIF functions as dependent variables for a range of quantiles for years t_0 and t_1. Similar to the linear approximation, I then compute the Oaxaca-Blinder decomposition by years for each quantile of interest.

2.5.2 Descriptive Statistics

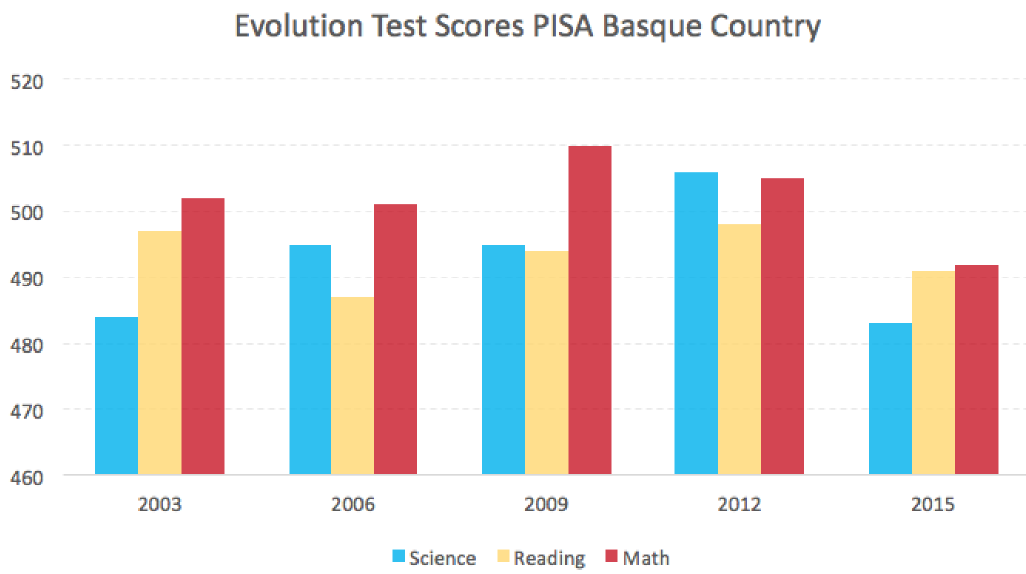
The decline observed in learning outcomes in the Basque Country in 2015 came after a gradual increase of test scores between 2006 and 2012 (see Figure 2.1). Between 2012 and 2015, the Basque Country decreased its performance in mathematics by 13 PISA points and especially science, by 23 PISA points. Beyond the statistical significance, this decline is large, given recent estimates that equate one year of schooling to 35 PISA points (OECD (2016b)). This decline represents an exception with respect to other regions in Spain, with most of them staying at the 2012 levels or even experiencing significant improvements[23]. Between 2012 and 2015, the average score in Spain increased significantly in reading, while it did not change in mathematics and science. The Spanish mild improvements contrast with an overall decline in many OECD countries. In fact, between 2012 and 2015 the OECD-35[24] average scores dropped from 501 to 493 points for science, from 496 to 493 points in reading, and from 494 to 490 points in mathematics. These declines were especially large in countries with historically good results such as Austria, Australia, Finland, South Korea, Netherlands, Poland or Sweden. Overall, these global trends show that providing quality education in the most developed systems is becoming more and more difficult, suggesting the rise of important challenges in the coming years.

Beyond computing the evolution of mean scores, I report the changes in the distribution of scores across years for the three domains. Figure 2.2 shows the performance distribution of students in the Basque education system for each of the five rounds in which the Basque Country has participated. The green line, representing the results of 2015, shows a shift to the left in all domains with respect to 2012 (yellow) and 2009 (black). However, the decline is larger (especially in mathematics and reading) for the group of highest achieving students, which have the lowest possible scores of all 5 rounds. With respect to the lowest achieving students, compared to the mean decline, the decline in scores was large in science, but smaller in mathematics and rather constant in reading: this results in a decline of the standard deviation in math, a constant evolution in reading and an increase in science. Overall, this does not go hand in hand with the evolution of socioeconomic status as seen in Section 2.4.

[23] It is to be noted that 500 points are anchored to the OECD average in 2000 (reading), math (2003) and science (2006) and hence stands as a fixed level of competencies for each of the three domains.

[24] OECD average-35: Arithmetic mean across all OECD countries.

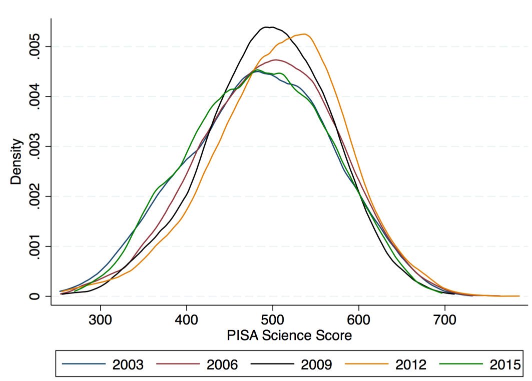
FIGURE 2.1: Changes in PISA Student Performance in the Basque Country



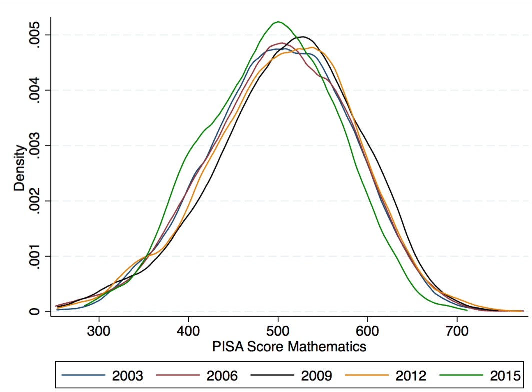
Notes: Source: OECD results for PISA 2003-2015 in the Basque Country.

FIGURE 2.2: Distribution of PISA Performance in the Basque Country across Domains and Years.

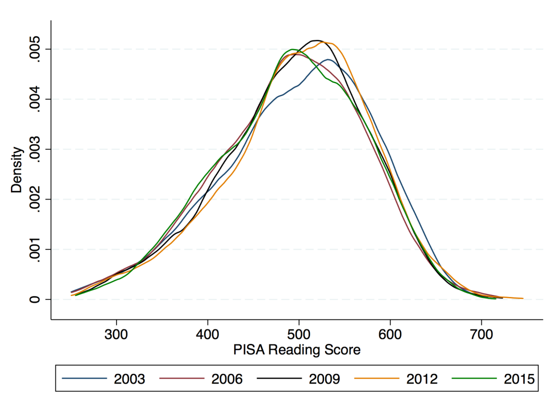
(A) Science.



(B) Mathematics.



(C) Reading.



I compute the average scores of the variables of interest described in the previous section by year. Results in Table 2.4 show no changes in the proportion of girls or the average age. Conversely, the proportion of students who take the test in a language similar from the one at home decreased by 9% (from 91% to 82%) between 2012 and 2015. The proportion of immigrants has increased in recent years, especially between 2003 and 2012, as it went from 2% to around 9%, remaining stable between 2012 and 2015. As was previously shown in Table 2.1, the re-calibrated Index of Social, Economic and Cultural status (ESCS) shows a progressive increase between 2003 and 2012, and a stagnation between 2012 and 2015, where there was not much change. More importantly, the proportion of 15-year old students enrolled in public schools has increased in the sample, which coincides with the official numbers provided by EUSTAT[25].

In terms of linguistic models of instruction, it can be seen that the joint (public and semi-public) proportion of students in the A model has decreased significantly between 2003 and 2015, especially in semi-public schools, where it felt from 25% to 8% of the total student population. In the other two language models of semi-public schools, this led to an increase of students in both the B model (from 17% in 2003 to 23% in 2015 out of the total students in the system) and the D model (from 18% in 2003 to 24% in 2015 out of the total students in the system). At the same time, the decrease in enrollment in A and B models in public schools, together with the slight increase of total enrollment in public schools has implied a large increase of enrollment in the D model in public schools from 23% in 2003 to 38% in 2015 out of the total student population.

Regarding other key variables, the rate of repeaters decreased progressively between 2003 and 2012 (from 24% to 20%), but it increased again between 2012 and 2015 up to 24% together with a shift in the proportion of students who had repeated once by age 15 to those which had repeated twice by that same age (from 4% to 6%). Regarding the harmonized indexes, there was a slight significant increase [26] in the Index of Instructional Improvement and Professional Development promoted by the principal as well as in the Index of Teacher Participation in Leadership. At the same time, the indexes of teacher and student behavior display divergent patterns: such indexes take large positive values when reflecting negative behavioral patterns at school, and negative values when reflecting positive behavioral patterns. This means that between 2012

[25] According to EUSTAT website, the proportion of students enrolled in public schools in lower secondary grades was 42% in the school year 2002/2003, 44% in 2005/2006, 45% in 2008/2009, 46% in 2011/2012, and 47% in 2014/2015.

[26] T-tests for the mean differences across years were conducted and found significant change in values.

and 2015, there was a deterioration of student behavior at school and an improvement in teacher behavior at school[27]. The indexes of school responsibility on both resources and pedagogy have decreased in 2015 with respect to previous years. Other relevant findings include a slight increase in student-teacher ratio (from 10.8 in 2009 to 11.5 in 2015), consistent with a decrease of public budget for education in the Basque Country.

[27] Again, t-tests for the mean differences were conducted and found very small standard errors of the estimates, resulting in statistically significant changes.

TABLE 2.4: Descriptive statistics of key Variables in the Basque Country, by year.

	2003		2006		2009		2012		2015	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender (Female)	0.50	0.50	0.50	0.50	0.42	0.49	0.50	0.50	0.50	0.50
Age	15.89	0.28	15.86	0.29	15.89	0.28	15.88	0.28	15.88	0.28
Language of test same as home	0.90	0.29	0.92	0.28	0.91	0.29	0.91	0.29	0.82	0.38
Immigrant status (Native)	0.98	0.13	0.96	0.19	0.95	0.21	0.91	0.28	0.91	0.28
ESCS (Harmonized)	-0.29	1.02	-0.08	0.99	0.09	0.90	0.14	0.98	0.15	1.03
Repeater (1 year)	0.22	0.41			0.17	0.37	0.16	0.37	0.18	0.38
Repeater (2 years)	0.02	0.12			0.05	0.22	0.04	0.21	0.06	0.24
LEADTCH (Harmonized)							-0.03	0.95	0.01	1.02
LEADPD (Harmonized)							-0.06	0.90	0.09	1.03
LEADINST (Harmonized)							-0.10	0.94	0.10	0.93
STUBEHA (Harmonized)	-0.03	1.04			-0.15	0.91	0.02	0.96	0.15	0.95
TEACHBEHA (Harmonized)	0.13	1.04			0.07	0.90	0.18	1.00	0.00	1.01
BELONG (Harmonized)	-0.04	0.87					0.01	1.02	0.03	1.10
JOYSCIE (Harmonized)			-0.10	0.91					0.11	1.07
SCIEEFF (Harmonized)			0.03	0.92					-0.04	1.08
RESPRES (Harmonized)			0.24	1.17	0.09	0.98	0.03	0.93	-0.22	0.86
RESPCURR (Harmonized)			0.12	0.93	0.07	0.99	0.18	0.94	-0.28	1.04
Public (Model A)	0.07	0.26	0.06	0.24	0.04	0.20	0.05	0.22	0.04	0.19
Public (Model B)	0.09	0.29	0.07	0.25	0.06	0.23	0.05	0.22	0.04	0.19
Public (Model D)	0.23	0.42	0.28	0.45	0.32	0.47	0.35	0.48	0.38	0.48
Semi-Public (Model A)	0.25	0.43	0.21	0.41	0.15	0.36	0.11	0.31	0.08	0.27
Semi-Public (Model B)	0.17	0.38	0.16	0.37	0.20	0.40	0.22	0.41	0.23	0.42
Semi-Public (Model D)	0.18	0.39	0.22	0.41	0.23	0.42	0.23	0.42	0.24	0.43
School Size	738.22	515.24	770.48	517.27	707.81	470.80	731.14	472.39	740.55	509.20
Student-teacher ratio	12.79	5.64	11.80	4.81	10.84	4.33	11.23	4.85	11.47	4.41

Notes: Values are computed using plausible values and replicate weights.

2.5.3 Results

I first display results of the linear regression model for 2015, with the covariates included in Table 2.4. Table 2.5 shows the OLS regression results for all domains in 2015. First, the lack of observations on specific variables (leadership indexes or school resources) reduces the sample by more than 30% of observations[28]. The results show a positive effect for boys in mathematics and science, while there is a positive effect for girls in reading, all of which are statistically significant. Moreover, taking the test in the same language as the main language at home is associated with large positive and significant increases of 10 to 22 PISA points, especially in reading. Combined with the decrease in the proportion of students who take the test in a language similar to the one they regularly speak at home observed in Table 2.4, this already suggests that this factor is going to be a key driver behind the decline in performance observed in 2015.

More importantly, the gradient of socioeconomic characteristics is small, and smaller compared to other countries (OECD (2016a)). This is important, as it lowers the expectations on the discussion on changes in socioeconomic conditions and its potential impact on learning. I find a negative relation of immigrant students and learning, even after controlling for socioeconomic status, which is more intense in mathematics and science. Finally, students who have repeated a grade are far behind those who have not even after controlling for socioeconomic characteristics (70 points for one-year repeaters, and around 100 points of 2-year repeaters). These numbers are large and consistent with García-Pérez et al. (2014) for the whole Spanish case.

Beyond student characteristics, the Index of Student Behavior is significantly associated with student performance. A one standard deviation positive increase in the index (hence showing behavior problems at the schools) is associated with a decrease of 6 to 7 PISA points in all three domains. In the science domain, the scientific learning indexes show a positive and significant relation with learning outcomes, especially the Index of Science Enjoyment (which seems of course to hide an endogenous relation with learning outcomes). Finally, it is to be noted that at the system level, there are no significant differences between school networks or linguistic models. Similarly, school size, student-teacher ratio and school autonomy variables do not display significant relations with student outcomes when controlling for key student characteristics.

[28] I consider alternative specifications by removing these variables, but the results are essentially similar for the key variables of interest.

TABLE 2.5: OLS results in 2015.

	Science	Math	Reading
Gender (Female)	-13.35***	-20.05***	8.98**
Age	1.48	2.32	5.19
Language of test same as home	17.11***	9.53	22.26***
Immigrant status (Native)	17.02***	22.62**	11.34
ESCS (Harmonized)	7.04***	10.90***	10.45***
Repeater (1 year)	-67.36***	-79.35***	-75.66***
Repeater (2 year)	-92.55***	-99.45***	-102.76***
ESCS at School (Harmonized)	-0.48	-0.55	-2.52
LEADTCH (Harmonized)	-4.61	1.27	-4.69
LEADPD (Harmonized))	1.44	-5.12	1.97
LEADINST (Harmonized)	-1.00	-0.20	-2.89
STUBEHA (Harmonized)	-7.71**	-6.63*	-7.02*
TEACHBEHA (Harmonized)	1.55	0.81	-0.73
BELONG (Harmonized)	1.77	1.34	4.95**
JOYSCIE (Harmonized)	21.42***		
SCIEEFF (Harmonized)	7.43***		
RESPRES (Harmonized)	-0.41	0.68	-0.95
RESPCURR (Harmonized)	4.50*	3.81	3.02
Public (Model B)	6.46	14.13	7.26
Public (Model D)	-11.31	-3.00	-15.12
Semi-Public (Model A)	-1.61	-0.77	2.45
Semi-Public (Model B)	-12.18	-9.10	-12.09
Semi-Public (Model D)	-4.88	-2.91	-11.14
School Size	0.00	0.00	0.00
Student-teacher ratio	-0.17	0.20	0.26
R-2	0.37	0.32	0.28
Observations	2,140	2,302	2,302

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Values are computed using plausible values and replicate weights.

In Table 2.6, I show the results of the Oaxaca-Blinder decomposition for the mean differences between 2015 and 2012, and between 2015 and 2009, for all three domains[29]. Results for 2012 show that the model explains around 35% of the differences in science, 40% in maths, and 90% in reading. In particular, I find that among the variables of the model, the increase in the share of students who take the test in a language different from the one at home (which is part of the individual characteristics group of variable) and the increase in the share of repeaters are statistically significant and account for most of the explained part. These two factors jointly explain around 5 points of the decline between 2012 and 2015 for all three domains. Although grade repetition (which takes place prior to the test) is highly endogenous to the PISA scores, it can be argued that the decision of taking the test in a specific language is rather exogenous and depends mostly on the school preferences. Hence, the effect measured is derived from a genuine variation of student and school behavior which in principle should not be related to unobserved endogenous factors.

Overall, these results show that the explanatory power of the model is limited for the mathematics and science drop in outcomes, and unfortunately does not capture an important share of the variation in scores. The fact that the explained part in reading is high is possibly due to the model construction, which measures differences in covariates which are repeated in the three models, except for science, for which more data is available to find additional sources of variation in covariates between 2012 and 2015. Given that the observed differences in reading between 2012 and 2015 are lower, the model is able to capture a larger share of the decline, although the absolute drop captured is similar to the one in mathematics and science.

For the case of differences between 2009 and 2015, results show that the model explains almost 50% of the differences in science, almost 40% in maths, and does not bring any significant results for reading, where the observed differences are very small. The increase of proportion of students who take the test in a difference language than the one at home and the increase in the share of immigrants are the two factors explaining

[29] In Tables 2.9 and 2.10 of the Data Appendix, results for 2003-2015 and 2006-2015 are showed in summarized way by grouping variables. What can be seen is that first, the differences in performance are not as large as with respect to 2009 and 2012. Second, the improvements in terms of socioeconomic backgrounds translate into counter-factual improvements in learning outcomes. This means that the explained part of the model accounts for positive increases in performance (although it is not statistically significant on aggregate terms), and hence, the unexplained part in the case of science and mathematics is negative, larger and statistically significant.

the incidence of individual characteristics in the decline, although the size is small: together, they account for an average of 2 PISA points, depending on the domain. Moreover, the worsening of student behavior indicators between 2009 and 2015 accounts for an important share of the explained part (especially for mathematics scores), and is statistically significant. The other variables observed in the model do not display relevant changes with respect to 2009 that incur in score differences that are statistically significant.

TABLE 2.6: Oaxaca-Blinder mean detailed decomposition in three domains between 2009-2012 and 2015.

	2012-2015			2009-2015		
	Science	Math	Reading	Science	Math	Reading
PISA 2015 Score	487.70	495.87	496.91	487.60	495.79	496.46
PISA 2012 Score	511.13	511.48	504.47	496.63	511.65	496.52
Difference	-23.4***	-15.6***	-7.56	-9.02**	-15.8***	-0.06
Explained	-8.66**	-6.93*	-6.79**	-4.71	-5.92*	-2.73
Unexplained	-14.8***	-8.68***	-0.76	-4.31	-9.93***	2.67
Explained variables						
Gender	-0.14	-0.18	0.13	-0.39	-0.91	0.50
Age	0.02	-0.01	0.06	0.01	0.02	0.01
Language of test same as home	-1.15**	-0.78*	-1.25**	-0.98**	-0.69**	-1.24**
Immigrant status	-0.04	-0.05	-0.03	-0.59*	-1.06**	-0.58**
ESCS (Harmonized)	-0.14	-0.13	-0.13	0.42	0.40	0.34
Repeater (1 year)	-1.66*	-1.78*	-1.74*	0.30	0.36	0.48
Repeater (2 years)	-1.90*	-1.90*	-1.97*	-0.26	-0.29	-0.28
ESCS School (Index)	-0.13	-0.11	-0.09	-0.71	-0.83	-0.80
LEADTCH (Harmonized)	0.31	0.33	0.17			
LEADPD (Harmonized)	-0.67	-0.97	-0.10			
LEADINST (Harmonized)	-0.58	-0.19	-0.46			
STUBEHA (Harmonized)	-0.95	-0.74	-0.81	-1.87*	-2.46*	-1.74*
TEACHBEHA (Harmonized)	-0.29	-0.16	-0.31	-0.40	-0.47	-0.26
BELONG (Harmonized)	0.25	0.17	0.42			
RESPRES (Harmonized)	0.12	-0.01	0.15	-0.01	-0.12	-0.06
RESPCURR (Harmonized)	-0.17	-0.07	-0.10	-2.09	-0.08	-1.70
Public (Model B)	0.02	-0.14	-0.02	0.83	-0.12	1.02
Public (Model D)	-1.06	0.48	-0.79	-1.46	-0.33	-0.83
Semi-Public (Model A)	-0.14	-0.12	-0.16	1.50	0.27	1.51
Semi-Public (Model B)	-0.08	0.06	0.07	0.07	0.21	0.05
Semi-Public (Model D)	0.31	-0.45	0.23	-0.02	-0.02	-0.01
School Size	-0.12	-0.01	-0.05	-0.11	-0.09	-0.07
Student-teacher ratio	-0.48	-0.15	0.01	1.04	0.30	0.93
Observations 2015	2,302	2,302	2,302	2,370	2,370	2,370
Observations 2009-2012	2,584	2,584	2,584	4,213	4,213	4,213

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Estimates are computed using plausible values and replicate weights. A Neumark model specification is considered. Standard Errors are clustered at the School level.

Finally, in Table 2.7, I decompose the differences in results across the performance distribution following the RIF regression approach. I do this by comparing both 2009 and 2015, as well as 2012 and 2015. For the sake of simplicity, results are reported in groups of variables. Moreover, it is to be noted that the standard errors are usually smaller compared to those computed for the mean in Table 2.5 and Table 2.6: this is because the model does not allow for clustering standard errors. The purpose of this estimation, nevertheless, is to take a deep look at point estimates, rather than its statistical significance, which was already estimated previously in the linear specification. For better interpretation, given that 2012 and 2009 scores are higher than 2015 for most of the performance distribution, results are reported by subtracting the 2015 scores from those of 2009 and 2012[30].

I use three different quantiles of the distribution, *p*20, *p*50, and *p*80. Regarding the comparison for 2012, I find that the explained differences are larger in the low percentile of the distribution (*p*20), which is related to the fact that the distributional statistic is related to a larger proportion of repeaters, and in particular, two-grade repeaters. Similarly, there are more students in that quantile who take the test in a language different from the language spoken at home, and hence the model is able to account for more observed differences. With respect of the largest percentiles (*p*80), results show that in mathematics and science, there is a significant (but small) improvement related to public schools in the linguistic D model and a decline of performance related to semi-public schools in A and D linguistic models. The aggregate effect of the school type differences is neutral and hence cannot be seen in aggregate terms in Table 2.7[31]. For the case of the comparison with respect to 2009, the explained differences are larger in the low-achieving students (*p*20). This is due to a larger increase in problems of student behavior with students in the lowest quantile, which has implied a larger observed decrease in scores with respect to 2009. The larger share of repeaters (two-year repeaters) and the increase in the proportion of students who take the test in a language different than the one they use at home are also behind this larger explained gap in *p*20.

[30]Just for the case of the reading percentile 20 comparing 2009 and 2015, the sample scores of 2015 are higher than those of 2009.

[31]For more details, please contact the corresponding author.

TABLE 2.7: Oaxaca-Blinder quantile decomposition in three domains between 2009-2012 and 2015.

	2012-2015			2009-2015		
	P20	P50	P80	P20	P50	P80
Science						
PISA 2015	416.7	490.06	558.29	416.66	490.29	559.71
PISA 2012	451.37	514.89	573.7	436.65	498.52	558.27
Difference (2015-2012/09)	34.67***	24.83***	15.41***	19.99***	8.23***	1.44
Explained	10.02***	9.77***	7.51***	8.04***	4.81***	1.37
Unexplained	24.66***	15.06***	7.9***	11.94***	15.06***	7.9***
Individual Characteristics	1.62***	3.32***	0.94**	1.96***	2.3***	0.07
ESCS (Harmonized)	0.1	0.44	0.18	-0.38	-0.49	-0.54
Repetition	5.05***	3.62***	2.01***	1.39	0.96	0.52
ESCS School (Harmonized)	0.02	0.23	0.2	-0.12	-0.34	-0.44
School Leadership	0.78*	1.78*	1.23***			
School Climate	1.17*	1.95***	1.56***	4.29***	1.95***	1.13**
Student Wellbeing	-0.43	-1.86	-0.18			
School Autonomy	-0.02	-0.11	-0.01	1.16	1.22	0.81
School stratum	0.96*	1.8*	0.92*	0.08	0.08	0.25
Resources	0.76*	2.02**	0.67***	-0.33	-0.86*	-1.71***
Mathematics						
PISA 2015	431.7	499.64	558.82	431.72	499.64	558.66
PISA 2012	445.35	514.89	579.15	446.96	517.18	582.28
Difference (2015-2012/09)	13.65***	15.25***	20.33***	15.24***	17.54***	23.62***
Explained	8.19***	7.55***	5.52***	7.88***	5.9***	3.67***
Unexplained	5.46**	7.7***	14.81***	7.35***	11.65***	19.96***
Individual Characteristics	1.26**	1.14**	0.94**	3.33***	2.68***	1.9***
ESCS (Harmonized)	0.08	0.15	0.18	-0.31	-0.44	-0.53
Repetition	5.79***	3.78***	1.95***	1.74	1.05	0.58
ESCS School (Harmonized)	0.16	0.11	0.14	-0.31	-0.29*	-0.35*
School Leadership	0.74	0.76**	1.04*			
School Climate	0.93	1.24**	0.65	3.78***	2.35***	1.8***
Student Wellbeing	-0.2	-0.31**	-0.15			
School Autonomy	-0.07	0.05	0.16	0.23	0.46	0.6
School stratum	-0.45	0.24	0.36	-0.4	0.11	-0.06
Resources	-0.06	0.38	0.26	-0.18	-0.03	-0.27
Reading						
PISA 2015	430.41	501.1	565.15	430.1	500.68	564.84
PISA 2012	440.79	512.2	569.99	429.57	503.67	565.39
Difference (2015-2012/09)	10.38**	11.10***	4.85*	-0.53	2.99	0.55
Explained	8.58	7.83***	4.86***	4.61*	3.29**	0.56
Unexplained	1.8	3.27	-0.01	-5.14*	-0.31**	-0.01
Individual Characteristics	1.90***	1.34***	0.62**	2.16***	1.74***	0.76**
ESCS (Harmonized)	0.11	0.14	0.17	-0.24	-0.39	-0.49
Repetition	6.29***	3.45***	1.72***	1.75	1	0.54
ESCS School (Harmonized)	-0.01	0.11	0.09	-0.58**	-0.46**	-0.54**
School Leadership	-0.28	0.29	1.03			
School Climate	0.96	1.67***	0.99	3.38***	1.34**	0.6
Student Wellbeing	-0.56**	-0.45*	0.23			
School Autonomy	-0.06	0.01	0.01	-0.83	0.23	0.75
School stratum	0.29	0.78	0.49	-0.13	0.18	-0.03
Resources	-0.06	0.5	-0.05	0.91**	0.34	-1.04**

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. A RIF-regression model was applied following the RIFREG command in Stata with Neumark model specification for the Oaxaca decomposition. The RIFREG command does not allow a computation using plausible values and replicate weights and clustering errors at the school level.Individual characteristics include gender, age and immigrant status. Repetition includes dummies for one and two-year repeaters. School Climate variables include the Index of teacher and student behavior. School autonomy variables are the two indexes RESPCURR and RESPRES. School stratum include the 5 streams of school ownership and linguistic models (with Public schools in Model A as omitted). Resources include student-teacher ratio and number of students at the school.

2.6 Discussion and Conclusions

This paper describes and documents the recent trends in secondary education student outcomes in the Basque Country (Spain) as measured by PISA data. The "PISA Shock" experienced by the region in 2015 was large and significant in mathematics and especially science. I explain only part of this decline (around 30% to 50%), which has to do with three unrelated factors.

First, repetition rates slightly increased in 2015, both among 1-year repeaters and 2-year repeaters. Repetition is a policy implemented at the school level, and although it is endogenously related to the learning potential of students, previous research has shown that an important part is closely linked to cultural factors, teacher biases and misaligned internal assessments. Long-term trends show that repetition has been decreasing very slowly in the last years, and hence, this abrupt increase may have to do with sample selection of students or a temporary stagnation in the trend. According to the recent repetition trends in primary and lower secondary[32], one would expect a short decrease of repeaters at the end of lower-secondary (age 15) in the coming years, and hence a slight increase of learning outcomes measured by the PISA sample (which includes non-repeaters and repeaters) in the nearby future.

Second, I discuss the effect of the language of instruction and language of testing: mother-tongue instruction and test administration are positively related to student outcomes, at least in the short term. The fact that more students are taking the test in a language different from the one at home is a consequence of school decision and the school linguistic policy, not to student factors or key education policy decisions. In particular, the change in 2015 has to do with more students who speak Spanish at home now taking the test in Basque. This increase is concentrated mostly in schools which provide instruction in the linguistic model D, where students are receiving their education predominantly in the Basque language. This makes sense from a pedagogic and policy point of view, and given the low-stakes nature of the test, there are no real losses in terms of learning and opportunities by taking the test in one language or another. Overall, these are cosmetic changes that do not reflect changes in system reforms, school policies or pedagogic practices, but rather the externalities of a multi-lingual school system such as the one in the Basque country.

[32]See MECD (2017).

However, results regarding student behavior are more worrying. The index of student behavior has progressively deteriorated between 2009 and 2015. When I look at its components, I see a deterioration of three items: student truancy, students skipping classes and students lacking respect towards teachers. Whether these are real problems taking place in Basque schools or whether there is an growing dissatisfaction of principals and teachers due to the increasingly diverse student population or the social effects of the economic crisis remains unknown. Nevertheless, clearly additional information will be required required through civic engagement and school climate assessments in the nearby future.

What remains clear is that there is no apparent overall effect of the economic crisis on student learning, at least as regards what can be measured with PISA background questionnaires. First, socioeconomic conditions of students have on average not worsened since 2009, although low-ESCS students have experienced socioeconomic declines during the economic crisis. Nevertheless, this decline for disadvantaged students does not translate into worse student outcomes, probably due to the low ESCS gradient on learning observed in the Basque Country in comparative international assessments. It is to be noted that the PISA ESCS Index does not consider employment status of parents, hence overlooking the large variation in that respect in the Basque economy, which usually experiences large employment fluctuations. At the same time, student-teacher ratio has slightly increased in PISA 2015, consistent with the budgetary situation of schools, but I do not observe a relevant association with the decline of student outcomes.

In fact, although given the previous factors discussed account to explain a relevant share of the "PISA Shock", a large proportion of the decline remains unexplained. The PISA decline in the Basque Country goes hand-in-hand with the declines in the regional external assessments conducted by the Department of Education of the Basque Government. Beyond the aforementioned contingent factors, this points to deeper structural changes in the Basque education system. From a policy perspective, this generalized phenomenon raises at least are two questions that I cannot test, but which will require further research.

The structural decline is also taking place in high performing education systems, such as Finland, South Korea or the Netherlands, as has been shown in the PISA 2015 results. Staying on top of performance is becoming increasingly difficult. For example, some authors have argued that the penetration of technology is generating a competition with schools as a knowledge institution (see Välijaärvi and Alasuutari (2017) for Finland). Others have argued that the demand for education may be changing due to the perceived low returns to schooling after the economic crisis, hence generating the perception that the education social contract may be broken.

The other reason behind the abrupt decline in performance may be the surge of anti-assessment cultures in the Basque (and Spanish) education system given the fact that the PISA test is of low-stakes nature. Zamarro et al. (2016) find that cross-country differences in international low-stakes tests like PISA are explained to a relevant extent (between 32% and 38%) by socio-emotional skill measures, such as survey and test effort. In particular, in Souther European countries like Spain -and all its regions including the Basque Country, as reported by Balart and Cabrales (2017)-, these measures of student effort are usually lower than in Asian or Nordic countries. After the new Spanish Education Act-LOMCE was approved in 2013 (which implied a large reform in the national assessments) there has been increasing anecdotal evidence of reporting problems and test-taking issues, including episodes of family and school boycotting of testing. If Spain (and the Basque Country) is more prone to being affected by the testing context (as shown by Zamarro et al. (2016)), such events may be having a non-negligible impact on student outcomes. Further research should address this in the future, in order to disentangle cultural factors around low-stakes testing with the real cognitive skill dynamics of students due to the evolution of education policies, school practices and socioeconomic factors.

2.7 Data Appendix

TABLE 2.8: Index Items and Years.

Sub-Dimension	Item
Wealth Index (WEALTH)	A room of your own
	A link to the internet
	A DVD-Video
	A desk to study at
Cultural Possessions Index (CULTPOSS)	Classic literature
	Books of poetry
	Work of art
Home Educational Resource Index (REDRES)	A quiet place to study
	A computer you can use for school work
	Books to help with your school work
	A dictionary

TABLE 2.9: Oaxaca-Blinder mean decomposition in three domains between 2006 and 2015.

Variable	Science	Math	Reading
PISA 2015 Score	492.20	495.70	496.36
PISA 2006 Score	497.72	503.08	492.12
Difference	-5.53	-7.38	4.24
Explained	5.87*	5.88	5.35
Unexplained	-11.40***	-13.26***	-1.10
Explained Variables			
Individual Characteristics	-2.42***	-2.53**	-2.22**
ESCS (Harmonized)	2.59***	3.91***	3.82***
ESCS School (Harmonized)	3.14**	3.99**	3.73**
Science Learning (Harmonized)	3.88**		
School Autonomy	-1.10	-0.69	-0.21
School stratum	0.26	1.93	1.11
Resources	-0.47	-0.73	-0.88
Observations 2015	2,302	2,302	2,302
Observations 2006	2,584	2,584	2,584

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Estimates are computed using plausible values and replicate weights. A Neumark model specification is considered. Standard Errors are clustered at the School level. Individual characteristics include gender, age and immigrant status. Scientific learning variables include the two indexes JOYSCIE and SCIEEFF. School autonomy variables are the two indexes RESPCURR and RESPRES. School stratum include the 5 streams of school ownership and linguistic models (with Public schools in Model A as omitted). Resources include student-teacher ratio and number of students at the school.

TABLE 2.10: Oaxaca-Blinder mean decomposition in three domains between 2003 and 2015.

Variable	Science	Math	Reading
PISA 2015 Score	487.94	496.08	496.86
PISA 2003 Score	488.81	504.74	501.94
Difference	-0.86	-8.66**	-5.08
Explained	0.58	-0.53	-2.07
Unexplained	-1.45	-8.13***	-3.01
Explained Variables			
Individual Characteristics	-1.13**	-1.45*	-1.70**
ESCS (Harmonized)	5.86***	4.48***	5.24***
Repetition	-1.51	-1.58	-1.69
ESCS School (Harmonized)	2.73*	2.18	2.00
Student wellbeing	0.01	0.08	
School Climate	-1.06	-1.12	-0.86
School stratum	-4.32***	-3.40***	-5.34***
Resources	0.01	0.29	0.00
Observations 2015	2,302	2,302	2,302
Observations 2006	2,584	2,584	2,584

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Estimates are computed using plausible values and replicate weights. A Neumark model specification is considered. Standard Errors are clustered at the School level. Individual characteristics include gender, age and immigrant status. Repetition includes dummies for one and two year repeaters. Student wellbeing includes the sense of belonging index. School Climate variables include the Indexes of Teacher and Student Behavior. School School autonomy variables are the two indexes RESPCURR and RESPRES. School stratum include the 5 streams of school ownership and linguistic models (with Public schools in Model A as omitted). Resources include student-teacher ratio and number of students at the school.

FIGURE 2.3: Description of Index Items

Index	Question	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8
BELONG	Thinking about your school: to what extent do you agree with the following statements?	I make friends easily at school	I feel I belong to school	Other students seem to like me					
JOYSCIE	How much do you disagree with the statements about yourself?	I generally have fun when I am learning broad science subjects	I like reading about science	I am happy working on science topics	I enjoy acquiring new knowledge in science	I am interested in learning about science			
SCIEEFF	How easy do you think it would be for you to perform the following tasks on your own?	Recognize te science question that underlies a newspaper report on a health issue	Explain why earthquakes occur more frequently in some areas than in others	Describe the role of antibiotics in the treatment of disease	Identify the science question associated with the disposal of garbage	Predict how changes to an environment will affect the survival of certain species	Interpret the scientific information provided on the labelling of food items.	Discuss how new evidence can lead you to change your understanding abou the possibility of life on mars	Identify the better of two explanations for the information of acid rain.
RESPRES	Regarding your school, who has a considerable responsibility for the following tasks? (Principal, Teachers, School Governing Board, Regional or local authority, national authority)	Selecting teachers for hire	Firing teachers	Establishing teachers' starting salaries	Determining teachers' salary increases	Formulating the school budget	Deciding on budget allocations within the school		
RESPCURR	Regarding your school, who has a considerable responsibility for the following tasks? (Principal, Teachers, School Governing Board, Regional or local authority, national authority)	Establishing student assessment policies	Choosing which textbooks are used	Determining course content	Deciding which courses are offered				
LEADINST	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during the last year.	I promote teaching practices based on recent educational research	I praise teachers whose students are actively participating in learning	I draw teachers' attention to the importance of pupils' development of critical and social capacities					
LEADPD	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during the last year.	When a teacher has problems in his/her classroom, I take the initiative to discuss matters	I pay attention to disruptive behaviour in classrooms	When a teacher brings up a classroom problem, we solve the problem together					
LEADTCH	Below are statements about your management of this school. Please indicate the frequency of the following activities and behaviours in your school during the last year.	I provide staff with opportunities to participate in school decision-making	I engage teachers to help build a school culture of continuous improvement	I ask teachers to participate in reviewing management practices					
STUBEHA	In your school, to what extent is the learning of students hindered by the following phenomena?	Student truancy	Student skipping classes	Students lacking respect for teachers	Student use of alcohol or illegal drugs	Students intimidating or bullying other students			
TEACHBEHA	In your school, to what extent is the learning of students hindered by the following phenomena?	Teachers not meeting individual students' needs	Teacher absenteeism	Staff resisting change	Teachers being too strict with students	Teachers not being well prepared for classes			

Chapter 3

School Choice, Student Mobility and School Segregation in Madrid

3.1 Introduction

School choice is one of the main pillars of a country's educational framework. The school decision is key for children, since it determines a large portion of the student's future social and academic development, as well as labor market outcomes. At the government level, it is a public policy concern due to the fact that it implies making decisions about who is granted priority to a certain school when available places are limited. At the family level it is a fundamental decision, insofar as taking the system as given, the objective is to identify the strategy which maximizes the options of being admitted to their preferred school.

School choice has been at the center of a lively debate during the last decades. Friedman (1955) first advocated for market-based education policies. Proponents of large school choice settings claim that, through student vouchers, a market-based school system increases the competition between schools for students who would lead to a rise in the effectiveness of schools for better student outcomes for all. Competition derived from parental choice can result in a rise of learning through more effective schools. For example, Hoxby (2000) finds more effective schools in districts with higher levels of choice. Conversely, opponents of school choice claim that the policy does not clearly present positive results, while it tends to exacerbate inequities through student sorting (Levin (1998); Epple et al. (2017)). Research has documented the diverse demand patterns of families when more choice can be exerted. While choice can allow for better opportunities for some households which live in disadvantaged neighborhoods, reforms granting excessive choice to parents overall appear to be exacerbating educational inequality through larger school segregation (sorting across schools) of students by their social or immigrant background (Musset (2012)).

This paper investigates a large-scale inter-district school choice reform in the city of Madrid on student geographic mobility and its consequent impact on within-school social and immigrant segregation. In a nutshell, we find that the reform increases the level of geographic mobility for most families, which goes from a non-response from households with immigrant children to a larger response being exerted by higher educated households. The increase in the level of mobility is associated to a sharp increase of the within school immigrant segregation and a small decrease of social segregation, although heterogeneous results are found depending on the average performance of the school of interest.

During the last three decades, there has been a clear pattern of educational authorities having increased the degree of school choice in their educational systems (Musset (2012)). In the US, many school choice reforms were complemented by busing programmes (e.g. Seattle in 1999 or North Carolina in 2002). In particular, school choice reforms involve, among others, zoning and de-zoning policies, changes in admission criteria, and changes in the system of assignment of students and schools.

In Spain, the national legislation in place in 2013 regulates school admission through the principles of equality of opportunity of learning and freedom of choice for families. The region of Madrid provides education to more than 1 million students in the city of Madrid (our main focus of study) and the 178 municipalities of the region, which taken as a whole is the third largest metropolitan area in the European Union. Regarding school choice criteria, one of the most important priority criteria conceding several points in case of over-demand for school was proximity to the school up until 2012. The city of Madrid has 21 school districts, which pre-2013 coincided with administrative areas regarding such proximity criteria. Between 2012 and 2013, the school system of the region of Madrid implemented an inter-district school choice reform in primary and secondary schools of the city of Madrid (and other municipalities), virtually granting no importance to the proximity criteria for schools within the same school district of the household. This largely expanded the supply of potential school choice for families: in the case of the city of Madrid, families shifted from an average number of 20 schools per district (the catchment area) to all the municipal network of more than 500 schools.

The relation between school choice reforms and student segregation has generated significant policy interest, although the lack of identification in large-scale reforms is a regular caveat. Our paper constitutes a novel piece of evidence on the effects of a large-scale school choice reform on household mobility and school segregation. We exploit a unique rich administrative dataset of the universe of student applications and the final assignment to primary schools receiving public funds in the region of Madrid from 2010 to 2015.[1] Data allow us to geographically locate households and schools in order to test how parents change their preferences in terms of commuting distance and willingness to take their children to another district of the city. We also use student and census data to better understand the social and immigrant dynamics after the reform. Moreover, we use measures of school quality derived from the "Indispensable Knowledge and Skills" (CDI in its Spanish acronym) external assessment of students, implemented yearly for all students at the end of primary school in sixth grade.

We find that parents reacted positively in terms of commuting distance and foreign-district applications to the inter-district school choice reform. Highest educated parents are more prone to commute, and the reform exacerbated this: while the increase in inter-district mobility for the lowest education quintile group was of 1.3 percentage points, that of the most educated household went up to 5.7 percentage points. At the same time, parents of non-immigrant children reacted strongly to the reform (an increase of 3.9 percentage points in inter-district applications), while parents of immigrant children did not react at all to the reform in terms of inter-district mobility. We distinguish between intensive margin (within district) and extensive margin (between districts) effects. We find that the increase in distance traveled was mostly due to extensive margin effects, i.e. parents commuting to districts different from where they reside. This suggests, on the one hand, that the reform results that we identify are well related to the policy implemented (at the district level). On the other, although to a lesser extent, some distance mobility (an average 4% increase) was also exerted within districts as a consequence of the reform, probably due to the fact that outer-district applications

[1] Privately funded schools are not subject to the admission policy, representing around 15% of the total school network in the region. We do not observe these students in our analysis.

leave new places in schools within the same district.

Given such heterogeneous responses, we evaluate the consequent effects of mobility on student segregation across schools from a social (parental education) and immigrant status perspective. Controlling for school effects and residential segregation dynamics, we find that student social segregation experienced a small decrease. However, we find heterogeneous results when we look at the performance of the school of interest: schools at the bottom of the score distribution seem to report a decrease (increase) in the level of heterogeneity (segregation), whereas schools in the 3^{rd}, 4^{th} and 5^{th} quintile of the distribution experience a larger increase in terms of social diversity (decrease in segregation). We find a significant increase of immigrant segregation, as there is a polarization of the share of immigrants at schools: while there was an increase of schools with a low share of immigrants, the share of schools with a very large share of immigrants increased significantly. From an overall perspective, computing segregation indexes shows non-significant changes of social segregation and a large significant increase of immigrant segregation, jointly due to the policy effect of the reform and the residential dynamics of immigrant population during the reform.

The results are driven by cream-skimming mechanisms. In particular, we find that student mobility responses were mediated by the willingness to attend to semi-public schools (a sort of charter school network in Spain which accounts for 35% of students in Madrid). Semi-public schools are characterized by serving a larger share of students from higher socioeconomic backgrounds and native-born status compared to public schools. The small increase in social diversity of students took place in semi-public schools and schools with better grades in the standardized tests, whereas social diversity in public schools and schools with lower grades did not experience relevant changes. Conversely, the increase of immigrant segregation was driven by an increase in the level of segregation in schools placed in low-income districts, the divergence of public and semi-public schools, and schools in the 3^{rd} and 4^{th} quintiles of the test score school distribution.

The policy context of the reform cannot be ignored. More specifically, we consider two key aspects of the system which need to be mentioned. First, school segregation due to socioeconomic characteristics is already very large in Madrid, while immigrant segregation is rather low, as has been shown by Murillo et al. (2017) and Murillo and

Garrido (2018)). Second, in Madrid, as in many other school systems, school assignment through choice in based on the so-called Boston Mechanism, which has shown limitations to capture truthful families' preferences (Abdulkadiroglu and Sönmez (2003)). In this scheme, families' first option of school usually becomes (around 85% for the case of Madrid) the final school assignment. Together with other contextual factors of the policy scheme, it is not surprising that the magnitude of the results of the reform are not as large as one would expect, with an average increase of 3 percentage points on outer district applications. Looking at final school assignment rather than the first choice of school as the outcome of interest does not qualitatively alter the magnitude of the results: if any, compared to first school option, final school assignment (through the Boston Mechanism) is associated with a larger gradient of parental socioeconomic and immigrant characteristics for mobility outcomes.

We tackle the regular identification problem of large-scale choice schemes (Epple et al. (2017)) by exploiting variation in implementation of the reform in municipalities beyond the city of Madrid. Our identification strategy aims to detect the immediate response[2] of households with new entrants to the education system (at age 3), which potentially reduces bias concerns. Our approach follows a before-and-after analysis with a control group of municipalities by looking at the outcomes of interest through three different school years.

While the policy was implemented in the city of Madrid for the school year 2013/2014, it was announced in March 2012, allowing for parents to anticipate the reform in terms of residence. If highly educated parents moved to cheaper districts, the increase in mobility would be mediated by an artificial decrease of residential segregation, rather than by the true preferences of parents given their place of residence. Nevertheless, controlling for the implementation in other municipalities allows us to address this: we look at parental responses of households living in other municipalities in the region which also implemented the inter-district school choice, some in 2012 (the smaller ones) and others in 2013 (the larger ones). We find a similar response in terms of mobility for both type of municipalities for the year when the proximity criteria was implemented, suggesting that for the case of those in 2012, the anticipation to the reform may have played a small role in confounding the response to the reform. Moreover, there were

[2] Testing family reactions in subsequent years after the reform (after 2013) would potentially raise concerns in the identification of family strategic choices (e.g. changing their residence as a consequence of the reform) plus other confounding factors that could also be taking place during that period.

other forces operating at the same time along with the inter-district school choice reform, which blurred the marginal impact of the school-choice reform on mobility and segregation, such as the implementation of the bilingual program for primary schools: we find that the expansion of this program did not seem to be a crucial confounding factor for the school choice reform effects, in terms of mobility and segregation.

Another issue in our identification strategy is that the city of Madrid, together with the rest of municipalities in the region, implemented other relevant changes in terms of the priority criteria during the year 2012. In particular, the low-income priority bonus points became more restrictive, whereas a new criterion prioritizing students with relatives that were former students of the school was introduced (allowing for additional social sorting in highly demanded schools). However, compared to the small change in mobility observed in Madrid in 2012 (with respect to 2011), we find that the change in 2013 was independent of these two new criteria, which makes us think that the true preferences for mobility were mediated by the inter-district school choice criteria and not by other factors.

Finally, it is puzzling to find that the mobility response driven by highly educated parents does not link well with the mild decrease in social segregation. It is worth mentioning that the lack of accurate data in terms of parental education suggests that we may be over (under)-estimating the increase (decrease) in within-school (segregation) heterogeneity of parental education if the reform correlates with unobserved measurement error. Considering the fact that parental education and immigration status correlate negatively in Madrid, we may be computing a lower-bound estimate of social segregation in 2012 and 2013, so that social segregation may have actually stayed constant or even increased, as was the case of immigrant segregation.

The rest of the paper is organized as follows. Section 3.2 deals with the literature on school choice and school segregation. Section 3.3 describes and contextualizes the Madrid school choice reform. Data are detailed in Section 3.4. The results are shown in Section 3.6 and the mechanisms driving the reform are explored in Section 3.7. Section 3.8 addresses several robustness checks and finally Section 3.9 concludes.

3.2 Related Literature

This research is related to the literature analyzing the effects of school choice policies on parental preferences, the heterogeneity of those preferences, and the indirect effect on school segregation. Finally, we discuss the relation of student segregation and student outcomes.

3.2.1 Geography and Demand for Schooling

The debate on school choice policies often lacks appropriate context to better understand trade-offs faced by households, especially in urban areas. Without specific background information on the geographical distribution and density of households in a school district, region or country, it is hard to disentangle results from key theoretical and empirical nuances. In particular, the degree of specific Tiebout choice around the household deeply influences the real incentives faced by families. For example, in large urban areas with usually large residential segregation patterns, excessively small school zones (with the extreme case of an ex-ante assignment of each student to a unique nearby school) can end up perpetuating existing geographic segregation patterns and even promoting strong housing selection incentives which end up exacerbating urban segregation (Burgess et al. (2007); Gibbons et al. (2006)). The opposite case would be to extend all the school districts to a unique one, whch was what happened in the Madrid reform. In that case, the weight of social and immigrant-origin differences in the demand for schooling and the commuting capacity may gain importance, leading to a potential effect on the degree of school sorting. In the Madrid reform, the Regional Government enlarged the choice zone to the municipal level, granting a larger set of options to all households. The reform moved from around 2,000 within-municipality catchment areas to 179 single municipal zones.

Parents typically choose the school with the highest possible performance taking into account other school characteristics such as proximity and accessibility. The school performance depends on two main dimensions: school effectiveness and student composition. Even for the most educated families, ranking schools in terms of the school effectiveness is a difficult and imprecise task. In particular, the literature describes that parents tend to mostly value peer composition of the school and, only to a lesser extent, the effectiveness of the school in the learning progress of students given their socioeconomic characteristics (Rothstein (2006); Mizala and Urquiola (2013)). Nevertheless, provision of information matters: Hastings and Weinstein (2008) show with the help of a natural field experiment that low-socioeconomic parents receiving information about the school performance increase their likelihood of choosing a high-scoring school.

Additionally, preferences for schools are different depending on family socioeconomic background since preferences for different dimensions of education vary across types (see Anderson et al. (1992) and Burgess et al. (2015)). Hastings et al. (2009) find that while high-income families care mainly about test scores, poorer and minority families must trade off preferences for high-performing schools against preferences for a predominantly minority nearby schools: the authors argue that the difference in choice responses leads to a more stratified school system, as the impact of school choice policies will be determined eventually by parents' preferences on education. Beyond income factors, the sociology and psychology literature have identified several mechanisms through which school choice is shaped by own aspirations, behaviors, social capital and networks. For example, Teske and Schneider (2001) discuss parental involvement and motivation as drivers of differences in school choice.

Moreover, parental preferences can take even more explicit forms as a reaction to the dynamics of school enrollment. For example, there is evidence that those families who make the choice to opt out of the assigned school, either to non-public schools or, when possible, to alternative public schools, are more advantaged (see Levin (1998) for a review), and place higher value on the academic achievement of schools (Hastings et al. (2009)). There is also evidence of *white flight* and of so-called tipping points for schools, meaning that white students opt out if the minority share increases above a certain threshold (Card et al. (2010)). In summary, it seems that socioeconomic conditions, immigrant origin, aspirations and motivation, as well as school dynamics shape parental preferences in both implicit and explicit forms when choosing a school for their siblings.

3.2.2 School Choice and Segregation

The implementation of reforms that have granted greater choice to families seems to be positively related to an increase in social and immigrant segregation. For example, Sweden implemented two key reforms in 1992 that offered families more choice options when choosing schools. After the 1992 reform, Böhlmark and Lindahl (2007) find evidence of students being sorted by immigrant origin and parental background shortly after the reform was implemented, whereas Böhlmark et al. (2016) find a smaller impact in the long term. The 2000 reform, which sought to grant more choice in secondary schools, has also been documented to increase segregation at schools (Söderström and Uusitalo (2010); Yang Hansen and Gustafsson (2016)). In the case of Chile, Hsieh and Urquiola (2006) find that school choice did not increase average quality of education, but it actually increased segregation. Moreover, the New Zealand reforms implemented in the 1990s have also shown to increase social and immigrant segregation of schools (Ladd and Fiske (2001)).

However, other factors interacting with choice settings may play a hidden role regarding the real effects, such as how schools are able to implement explicit or implicit forms of discrimination. For example, Burgess and Briggs (2010) investigate the effect of school choice on social mobility in secondary education in England. They find that children from poor families are less likely to get places in good schools and this probability is unaffected by the degree of school choice. This suggests that there must be other additional features belonging (or related) to the educational system affecting student mobility beyond the degree of school choice.

3.2.3 Segregation and Outcomes

In the case that certain choice settings may lead to more segregation, it is equally important to justify why segregation may become an issue for advanced education systems. Excessive school segregation is becoming a growing public concern in policy and public debates. International organizations are starting to warn education systems about the risks of chronifying the cycle student disadvantage through school segregation. In the economics of education literature, the debate on segregation usually pivots around peer effects (by ability grouping) in the classroom and its impact on academic outcomes, which seems to be mixed. While Duflo et al. (2011) have found positive effects of sorting in Kenya basic schools, there are other important consequences of segregation that take the place both within and outside the classroom and the school. Hastings et al. (2012) find that the demand-side differences observed among households generate divergent pressures on high and low-scoring schools, resulting in a wider, rather than narrower, gap in achievement. A recent work by Chetty et al. (2016) analyzes the long-term impact of the Moving to Opportunity (MVO) housing project, which granted low-income families housing vouchers to move out to better neighborhoods. The authors observe that younger participants moving to a lower-poverty neighborhood were more likely to attend college and earn a higher wage.

On other topics, Billings et al. (2014) show that the end of de-segregation policies (through busing programmes) in the USA widened racial inequality again through more crime and poorer academic results, despite large efforts through compensatory resource allocation. However, there is also a general concern that social cohesion is adversely affected in excessively segregated education system. While, for example, the 2000 reform in Sweden did not seem to imply a decline in civic attitudes and social cohesion as a consequence of greater school segregation (Shafiq and Myers (2014)). Levin (2002) highlights the importance of social cohesion as one of the key dimensions in a democratic and free society when thinking about school choice settings. In particular, he refers to common elements of schooling for diverse student population, with regard to curriculum, social values, language, and political institutions. Another example is presented by Cullen et al. (2006), who study the outcomes for students who randomly get the place at their most preferred school and those who do not. While no academic effects are found, they find that students experience improvements on nontraditional measures (e.g. self-reported disciplinary incidents and arrest rates). While most students benefit from better peers, clearly the one important question is whether those potential gains come at the expense of losses of high-performing students in the classroom and school.

3.3 The School choice reform in Madrid

The Spanish Education System. The Spanish education system consists of 10 years of compulsory education, which start at age 6 and includes six years of primary school (up to age 12) and four years of lower secondary education (up to age 16). Even though compulsory primary education starts at the age of 6, students are offered free universal access to the public education system from age 3 onwards. Since most of publicly-funded schools offer preschool and primary education together, age 3 is typically the time when families enroll their children to school.[3].

Regarding the access to schools, the 1978 Spanish constitution grants the coexistence of the right to education and the freedom to educate children, an equilibrium of rights stemming from a political pact between progressive and conservative forces. In terms of education policy, the second principle was translated into the 1985 education act (LODE) which explicitly regulated the freedom of families to choose their children's school[4]. In the following years, this was accompanied by a decentralization process through which educational policies started to be jointly determined at the national, regional and municipal level[5]. Since then, the central government is responsible for establishing the organic laws (*Organic Laws*) and the royal decrees that the regional governments are allowed to further develop as long as they do not contradict the organic laws. With respect to the Spanish school choice policies in the years around the reform, the national organic law in place at the time of the reform (LOE[6]) established the general regulatory principles to be followed by the regional governments in order to determine the priority criteria of students in over-demanded schools.

In Spain, the vast majority of the school network is publicly funded. Such network includes public and semi-public schools (we follow Calsamiglia and Güell (2018), who name the network of privately managed schools as semi-public schools). Public schools are fully funded by the government and managed by civil servants and local school boards. Semi-public schools (*centros concertados*) are privately run but mostly financed through public funds. Although tuition fees are not allowed in semi-public schools, in practice parents pay small quasi-compulsory symbolic donations for basic educational services that can act as barrier of entry to disadvantaged families. With respect to admissions, all the schools of the public system (public and semi-public) are expected to unconditionally accept all students assigned by the centralized school choice mechanism, provided demand does not exceed supply.

School Choice in Madrid. In the region of Madrid, the majority of schools (around 85%) are part of the publicly funded network of schools by the Regional Government.

[3]Preschool education is fully publicly funded from age 3 to 6. This right was recognized in the Organic Law 1/1990 (*Ley Orgánica de Ordenación General del Sistema Educativo, LOGSE*).

[4]Organic Law 8/1985 (*Ley Orgánica reguladora del Derecho a la Educación.*).

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

[6]Ley Orgánica de Educación 2/2006.

Such system includes publicly managed schools (which cover around 50% of all students) and semi-public schools (privately managed, which cover around 35% of all students). Some authors have argued that preferences for education in Spain are mediated by the existence of the semi-public network (Arellano and Zamarro (2007); Mancebón et al. (2012)). These schools tend to be located in urban areas are larger in size, and serve more upper-middle income and non-immigrant households. In terms of the social composition of students in secondary education schools, the region of Madrid is the most socially segregated (where social means social, economic and cultural characteristics of the family) among all autonomous communities in Spain and neighboring countries (Murillo and Garrido (2018)), while its immigrant-origin segregation is rather low in terms of other Spanish regions (Murillo et al. (2017)).

The school choice system has been further updated and regulated after the 2006 LOE education act was passed. A centralized assignment mechanism is used to allocate students to schools in the publicly funded system (both public and semi-public) for pre-primary (starting at age 3) primary, and lower secondary, as well as for special education. Families are requested to submit a ranked-order list of schools up to a total number of choices, and their children are allocated by the centralized and algorithm-based automatic allocation procedure, the so-called Boston Mechanism (see Abdulkadiroglu and Sönmez (2003)).

The application timing works as follows. Before the school year starts in September (between end of April and early May), every participating family is requested to submit the ranked-order list of schools to their first choice. Applicants are assigned into a school by means of the Boston Mechanism (BM hereafter), a centralized school choice system which works as follows. First, students are allocated to their first choice school. For schools where there is an over-demand of students, students are granted priority points (according to several criteria which depend on student characteristics and location of the household or parental job) which provide them with a rank number that assigns places to students until all available places are filled. Ties are broken conditional on priority bonus points obtained[7]. In the second step, students who are rejected from their first choice are proposed to their second submitted school in the rank list provided there is room available after the first step. If there are more applicants than available places, students are allocated in the same way as before with the

[7] See Table 3.12 in the Appendix for further details.

priority points granted in the first choice. In the third step, those students who are rejected from their second choice are proposed to their third choice, and the mechanism continues until all students are assigned a place in a school of the public system.

The final assignment is made public in June, and enrollment must take place at the end of June (for pre-primary and primary education) or July to September (for lower secondary education). A special feature of the system is that student's priority points that are used for tie-breaking at all stages are based on the ones obtained in the first choice.[8].

Until recently, the BM has been very influential in practice (beyond Spanish regions, US school districts which used this mechanism included Boston, Cambridge, Denver, Minneapolis and Seattle, among others, as well as other cities such as Beijing, Amsterdam or Frankfurt). One of the special features of this assignment system is that the choice of the first ranked school is highly important, since the assignment in each round is final. The probability of a student of being admitted in the second round relative to the first is substantially reduced, and the chances decrease even more in further rounds.[9] Moreover, recent empirical evidence supports the theory. In the case of Barcelona, Calsamiglia and Güell (2018) highlight the fact that more than 85 percent of the assignments are resolved in the first round, and this is persistent across different cities worldwide[10]. In Madrid, around 86 percent of children were assigned to the school they ranked first, as will be seen in Section 3.4.

Reform of the priority criterion to school access. As highlighted above, in case of over-demand at a certain school, students are assigned to schools based on a government-determined priority criterion, which grants points to students according to their characteristics and their home residence or parental job location. For school choice, the region

[8]As Calsamiglia (2014) states, the main reason why the government uses this procedure is because it is computationally easier. Alternative assignment mechanisms require computational power that currently the education administration cannot deal with.

[9] Abdulkadiroglu and Sönmez (2003) highlighted that one of the major difficulties of the BM is the fact that it is not strategy-proof. A student may be highly prioritize to enter a school s, but if she does not list it as her first choice, she loses her seat in favor of students who have listed s as their top choice. BM provides incentives to families to misrepresent their preferences by ranking first those schools where they have higher priorities.

[10]See also Abdulkadiroglu et al. (2006) for Boston; Hastings et al. (2009) for Charlotte; Lavy (2010) for Tel Aviv; and De Haan et al. (2015) for Amsterdam.

of Madrid is divided into 179 municipalities, with the medium and large-size municipalities being subsequently divided into school choice catchment areas[11]. In particular, the city of Madrid (the largest municipality and our main unit of analysis) is divided into 21 districts, which coincide with such choice catchment areas. Table 3.1 exhibits the score scale used in the city of Madrid before and after the reform[12]. Before the reform (school year 2011/2012) and regarding proximity, children living or any parent/guardians working in the (a boundary) district of the school ranked first received 4 points (2 points). Regarding student individual characteristics, students were awarded 2 points if their per capita household income was under the IPREM Index (7,236.60 euros),[13] and got 1 point if their per capita household income was between 100% and 200% of the IPREM (between 7,236.60 and 14,473.20 euros). Families which ranked a school where there was a sibling first got 4 points, plus an additional 3 for every sibling enrolled at this school. Students received extra points if they have a family member with a disability (1.5 points), and if they belonged to a large family (1.5 if general - 3 children - and 2.5 if special- 4 or more children-). In addition, a specific point (1 point) was decentralized to the school decision, which must be decided according to objective criteria that are made public.

In March 2012, the Regional Government announced a reform that aimed at strengthening the principle of school choice by households with children entering pre-primary, primary and lower secondary schools[14]. The Regional Government founded its arguments on the constitutional right that parents have to educate their children based on their convictions. The goals of the government policy were to increase families participation, improve the availability of information on schools (through the results on the standardized test scores, school's educational program, school resources and services), simplify the admission process, promote school competition, and enhance free school choice. In particular, the reform modified the computation of student's priorities in each school, as can be seen in Table 3.1 for the city of Madrid. The changes in priority points and school districts were implemented in two consecutive years:

1. In 2012/2013:

[11] Called as *zonas de influencia*.

[12] The Table also applies to the rest of the municipalities regarding the individual characteristics.

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

[14] Order 2939/2012 of March 9 of the Regional Government of Madrid.

- The criterion to obtain bonus points awarded to low-income students was reduced and changed.

- A new priority bonus was awarded when a student's family member was a former student of the first listed school.[15]

2. In 2013/2014: The proximity to school criterion was relaxed: the Regional Government of Madrid updated the regulatory framework with a regional decree[16] which regulated the single school choice for all the municipalities of the region. For the city of Madrid, this implied turning from 21 districts (with around 20 schools per district) as choice catchment areas to a single municipal with more than 500 schools [17].

In 2012/2013, a sharp decline in the bonus for low-income students was introduced. Children were awarded 2 points if the family received the Minimum Income for Insertion Subsidy (*Renta Mínima de Inserción*), a social program which is granted to a small proportion of low-income households with no earnings (0.9% of the total population in the region of Madrid). The number of recipients of this subsidy are much less numerous (around 30,000 households in a region of more than 6 million population) than the families with a per capita household income under the 100% of the IPREM (around 15% of the population[18]) and even less than those between 100% and 200% of IPREM. Finally, an additional 1.5 points were awarded for students in a school were any family member had been a former student, an issue that may potentially limit equality of opportunity of students to access certain schools, given the weight given to socioeconomic background of parents[19].

In 2013/2014, family incentives to apply for a school inside their residence district were shifted with the implementation of the inter-district school choice (hereafter the

[15]Additionally, more points were awarded to families with siblings enrolled at school. Following Calsamiglia and Güell (2018), we consider this change irrelevant for the analysis, given that students' choice are previously conditioned by their older siblings' choice, and we do not include students with older siblings in our analysis.

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

[17]Relative to 2012/2013, 2 more points were awarded to families with siblings enrolled at school. We do not consider this change relevant for our analysis.

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

[19]Tie-break criteria were slightly modified, as can be seen in Table 3.12 of Appendix.

Madrid reform is referred as the inter-district school choice reform). A student living or with parents working and applying for a school in the same school district was awarded with 0.5 points, plus 4 points (2 points) if the school was in the same municipality (other municipality of the region) of the household or the parental workplace[20][21]. Overall, the inter-district choice reform was operationalized through a large drop in the importance of the proximity criteria for over-demanded schools.

TABLE 3.1: Priority Points in case of over-demand of Schools in the City of Madrid.

BONUS	CRITERIA	NUMBER OF POINTS		
		Before 2012/2013	**2012/2013**	**2013/2014**
Proximity -Madrid city-	Family house or parents' work in:			
	School district	4	4	
	Boundary school district	2	2	
	Family house or parents' work in:			
	Same municipality			4
	School district			0.5
	Region of Madrid			2
Low-income	Income<=IPREM	2		
	IPREM <Income <=2IPREM	1		
	Minimum Insertion Subsidy		2	2
Siblings	First one 4pts, and additional 3pts	4		
	One or more		8	10
Disability	Parents, students or siblings	1.5	1.5	1.5
Large Family	General	1.5	1.5	1.5
	Special	2.5	2.5	2.5
Former student	Family member is former student		1.5	1.5
School discretionary		1	1	1

Notes: The changes beyond the proximity criteria were applied together across all medium and large municipalities. IPREM is the acronym in Spanish for the Multiple Effects Income Public Index, which was €7,455.14 in the period of study. The Minimum Insertion Subsidy (*Renta Mínima de Insercin*) is a special provision granted for people with lower income than IPREM. School discretionary is a point that the schools have freedom to assign by "public and objective" criteria.

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

[21] Tie-break criteria were also slightly modified, as can be seen in Table 3.12 of Appendix.

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

3.4 Data and Descriptive Statistics

3.4.1 Data

We use a combination of four administrative datasets that provide rich and unique information on the universe of primary school applications of each household in the region of Madrid, the characteristics of publicly funded schools in the region of Madrid, the education of households at the census block level in the city of Madrid, and the standardized test scores at the school level (*CDI*). Data on student applications, schools and school test scores were provided by the Education Ministry (*Consejería de Educación*) of the Madrid Regional Government, and data on parental education was obtained from the Madrid Census. Information is available for every year from 2010/11 (2010 hereafter) to 2015/2016 (2015 hereafter).

Primary education applications. Our primary source of analysis is a unique and rich dataset containing information on the universe of all students who applied for a new school in primary education in the region of Madrid. For each applicant, the dataset contains the priority list of schools ranked from one to fourteen, the basic student information regarding family characteristics, home address, priority points obtained based

[22]Table 3.13 in the Appendix provides a summary of the municipalities that joined the single-zone school choice system across years.

on such characteristics and final school of assignment. Regarding family information, student home residence is given using geographic coordinates which we link (with the help of geolocation software) to different geographical areas (neighborhoods and districts). In addition, the application contains information about the student's country of birth, which we use to construct a variable of immigrant background status.

School database. The Madrid Regional Government also provided a list with the universe of schools and some of their features: the geographic coordinates of each school, the school ownership (public, semi-public or private), whether the school offers bilingual education (Spanish and English), and the levels of education offered.

Household socioeconomic characteristics. We use information from the Census Office of the city of Madrid which provides the distribution of education levels of the population by geographical levels of the city in 2014. The data is accessible for the three geographical levels of disaggregation: districts, neighborhoods and blocks. The most disaggregated units are the census blocks (*Sección censal*), which are constructed for local, regional and national election purposes (assigning each block to one voting center), and usually contain no more than 2,500 people[23]. We have access to information on the proportion of population in each education category by age groups at the census block level. We use this to translate the corresponding level of education to an equivalent number of years of schooling[24], which allows us to compute the average number of years of schooling for each block[25]. We assign to each family the corresponding value at the block where the family resides. This proxy for parental education is potentially obtained with measurement error. Limitations and potential unobserved heterogeneity issues are discussed in Section 3.8.

CDI Tests. In order to proxy school quality, we use a standardized exam administered for all Year 6 students in the region of Madrid between the 2004/2005 and 2014/2015 school years. The exam, known as the Essential Knowledge and Skills test

[23]Figure 3.5 in the Appendix includes an example of a block of the Central district of Madrid.

[24]The construction of this variable is detailed in the Appendix.

[25]Ideally, one would use data from 25-49 age group to obtain a more accurate proxy for the parental education, as this is the most relevant level of education for parents with preschool and primary school children. Nevertheless, we do not use this measure as the main one because the database corresponds to 2017, too far from the years of the reform.

(CDI-*Conocimientos y Destrezas Indispensables*), was designed for education policy measures, not for specific academic consequences on students. In particular, the goal of the test was to provide information for policymakers, schools and families about the school's average performance. The test focused mostly on curriculum content knowledge in the areas of reading and mathematics. The results were publicized every year with the purpose of facilitating school choice for families with new students entering the system.

3.4.2 Sample Restrictions

Table 3.2 presents the restrictions for the analysis sample. Depending on the year of interest, there are about 34,000 to 37,000 students who apply for a school of the public education system in the city of Madrid. Our population of interest consists of those students who live in the city of Madrid, apply to schools in Madrid, and have no siblings in the first-ranked school. Following Calsamiglia and Güell (2018), we focus on students with no older siblings enrolled in the school of first choice due to the fact that such families have different incentives and behavior compared to the rest of applicants: their present choice is previously conditioned by their past choice. The amount of bonus points that families obtain when applying for a school where a sibling is already enrolled is the highest, and the admission to those schools is almost automatically guaranteed. These families may therefore react differently to policy changes since they have distinct preferences and incentives.

Despite the fact that the administrative dataset of applications is unique and exhaustive, adding other sources of information for the analysis implies some observations (detailed in Table 3.2) are missing. For example, our analysis focuses on all students who apply for a school where information on the test scores (our measure of school quality) is available. We are able to identify the school test score measure for about 90% to 95% of the observations. Finally, given the importance of identifying geographical mobility patterns, we discard students whose address information is missing or not valid. This may be because school application forms are manually submitted to the school or the central administration which then introduce the information into the digital centralized system (Calsamiglia and Güell (2018)). We are able to identify the address of about 99% percent of the final population of interest. The analysis is based on a population of about 26,000 to 29,000 observations, depending on the school year.

TABLE 3.2: Sample Restrictions: School Applicants in the city of Madrid over 2010-2015.

Number of Applicants	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Total unique applications	37,146	39,986	37,300	36,273	35,727	34,124
Students with no siblings	32,009	30,594	30,930	30,022	29,341	27,762
School test scores	29,977	28,884	29,488	28,714	27,810	26,487
Valid Address	29,666	28,720	29,338	28,577	27,807	26,491
Total Final Sample	29,736	28,700	29,205	28,478	27,579	26,271

Notes: Each year corresponds to the year of application and the school year starting in September of that year.

3.4.3 Descriptive Statistics of Applicants

Table 3.3 displays summary statistics of applicants. First, the majority of applicants (more 60 percent) are applying to start in preschool education at age 3, which is the key stage of enrollment at schools. Following Calsamiglia and Güell (2018), we keep this as our group of interest in the analysis, given that assignment after this age in over-demanded schools becomes extremely difficult due to the lack of available places. Beyond this, a large fraction of applicants are native students (around 85 percent), whereas the rest represents the foreign-born student population. In 2013, there was a significant rise in the proportion of immigrant students, from 12.9% to 16.6, although this is consistent with the city demographics[26]. In our empirical strategy, we account for these changes by tracking residential dynamics of the immigrant population. The proportion of boys and girls remains constant over the period. Moreover, students awarded places with each of the two new priority bonus created in the 2012 reform only account for a small part: between 3% and 4% for the new low-income bonus and around 6% for former student relatives at the school. We do not drop these students from the main sample analysis due to the fact that we do not have this information for years before the reform. Discarding these observations may lead to a sample selection bias between the period before and after the reform.

[26] According to official municipal data in the city of Madrid, there was a sharp decrease of births in 2009 (who were age 3 in 2012) and 2010 (who were 3 in 2013) with respect to those born in 2008. In particular the number of births in the city of Madrid was 36,663 in 2008, 35,147 in 2009 and 33,987 in 2008. Conversely, the number of immigrants increased between 2010 and 2013. Hence, the observed drops in the census are consistent with those observed in Table 3.2.

TABLE 3.3: Summary Statistics: School Applicants in the City of Madrid over 2010-2015.

Variable	2010	2011	2012	2013	2014	2015
A. Grades						
Preschool Age 3	18,391	18,289	18,006	16,970	16,323	16,266
	[0,62]	[0,64]	[0,62]	[0,6]	[0,59]	[0,62]
Preschool Age 4	2,738	2,571	2,850	2,746	2,556	2,268
	[0,09]	[0,09]	[0,1]	[0,1]	[0,09]	[0,09]
Preschool Age 5	2,087	1,925	2,156	2,306	2,109	1,769
	[0,07]	[0,07]	[0,07]	[0,08]	[0,08]	[0,07]
Primary 1th grade	4,254	3,946	3,973	4,121	4,315	3,899
	[0,14]	[0,14]	[0,14]	[0,14]	[0,16]	[0,15]
Primary 2 to 5th grade	1,602	1,373	1,539	1,618	1,559	1,451
	[0,05]	[0,05]	[0,05]	[0,06]	[0,06]	[0,06]
Primary 6th grade	664	596	681	717	717	618
	[0,02]	[0,02]	[0,02]	[0,03]	[0,03]	[0,02]
B. Students characteristics - Preschool Age 3						
Total Students	18,391	18,289	18,006	16,970	16,323	16,266
B.1. Gender						
Boys	9,351	9,290	9,316	8,671	8,142	8,291
Girls	9,040	8,999	8,690	8,299	8,181	7,975
B.2. Immigrant Status						
Native	16,221	15,923	15,023	14,335	13,969	14,046
Immigrant	2,170	2,366	2,983	2,635	2,354	2,220
C. Bonus information						
Bonus for Former student	N/A	N/A	909	821	863	2,125
Bonus for RMI	N/A	N/A	638	601	645	615

Notes: Each year corresponds to the year of application and the school year starting in September of that year. Data on bonus information for former students and RMI was not available before 2012 given the reform was implemented that year.

3.5 Empirical Strategy

Families elaborate their submitted list of schools depending on their preferences and school priorities. This, together with the response of other families, determines the probability of being assigned to a given school. In this paper, we attempt to first estimate the impact of changes in priorities on families choices and, second, their subsequent outcomes in terms of student composition at schools according to their social (i.e. parental education) and immigrant (estimated through the country of birth) characteristics. Hence, we investigate whether the choice reform has a direct impact on student mobility, and as a secondary effect on student sorting and segregation.

We exploit two successive changes in the school priority points (described in Section 3.3): (i) In 2012, bonus points for low-income households were reduced, whereas bonus points for new criteria of former school students were incorporated; (ii) In 2013, bonus points for proximity were relaxed, leaving an almost *de facto* inter-district school choice area for the city of Madrid. These two changes modified the set of schools where families could have a high degree priority to attend. We identify the effect of these changes on family preferences comparing school choices immediately before and after these changes. We focus our main analysis on families that applied for a school in the city of Madrid, with no siblings enrolled in the first listed school, and with children starting school at age 3. We use the school ranked in the first position as the main proxy for household preferences.

3.5.1 Geographic Mobility

We look at two different indicators to investigate changes in mobility after the reform in the city of Madrid, which are complementary to each other. First, we compute the commuting distance between household and the first ranked school. We do so by using the Open Source Routing Machine (OSRM) routine, which returns the travel distance using the latitude and longitude coordinates of the household and the school [27]. Second, we look at whether parents aim to send their children to a school located in a different district from the one of the household. Both variables provide different information: while

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

outer district application may be seen as a pure “extensive margin” measure of mobility, the average distance to first choice school may represent a combination of “intensive” and “extensive margin” variable of mobility, as the distance traveled may include within and between district mobility. An important issue in this case and, to a lesser extent, in the computation of the commuting distance variable, is that we cannot identify parental workplace. Given that the reform reduces the importance of the district of both the household and workplace location, we need to assume that parental mobility to the workplace is constant during the years of the reform. Otherwise, changes in student mobility to schools could be driven by greater parent mobility to the workplace during the years of the reform.

Baseline regression. To identify the effect of change in priorities on the different measures of families preferences for geographic mobility, we use the following reduced-form equation:

$$D_{ijt} = \alpha + \beta Year2012_{ij} + \gamma Year2013_{ij} + X_n + \epsilon_{ijt} \tag{3.1}$$

where D_{ijt} is either the travel distance from the residence of the student i to the first choice school j in year t in her application form, or a dummy that takes value 1 if student i applies for a school j located in a different district where she resides in year t, and 0 otherwise. $Year2012_{ij}$ is a dummy that takes value 1 if student i applies for a school j in the academic year 2012/2013. $Year2013_{ij}$ is a dummy that takes value 1 if student i applies for a school j in the academic year 2013/2014. The coefficient α captures the mean outcome variable for the year prior to the reform, 2011 (academic year 2011/2012). We also control for the population density of the neighborhood n (the next geographical unit inside districts) where the student's household i lives (X_n) since it is a determinant factor on our outcome variables. We focus on the year immediately after the reform in order to capture the response to change in priorities. Looking at the family reactions in consequent years after the reform (after 2013) would potentially increase the bias concern in families choices, due to the fact that families had time to learn from the reform and change their reactions to the policy (e.g. changing their residence as a consequence of the reform) plus other confounding factors that could also be taking place during that period. Our identification strategy aims to detect the immediate response of households (β for 2012 and γ for 2013) with entrants to the education system (at age 3)[28]. Our empirical strategy analyzes the response of different cohorts of new students across years: hence we need to assume that the distribution of preferences remain constant over time. This seems to be highly plausible at least for the specific years of the reform.

Heterogeneous Responses. We also attempt to identify heterogeneous responses by family characteristics as a prior step to understand the student segregation and sorting dynamics. We look at the family background of students in two dimensions: parental education and student nationality (immigration status). This allows the social and immigrant-origin dynamics in the responses to the reform to be understood.

In order to identify the heterogeneous effects on parental education we use:

$$\begin{aligned} D_{ijt} =& \alpha + \beta Year2012_{ij} + \gamma Year2013_{ij} + \pi_c \sum_c EDU_{i,c} + \kappa_c \sum_c EDU_{i,c} * Year2012_{ij} + \\ &+ \eta_c \sum_c EDU_{i,c} * Year2013_{ij} + X_n + \epsilon_{ijt} \end{aligned} \tag{3.2}$$

where $EDU_{i,c}$ is a set of dummies accounting for the family educational background c for the student i. These dummies represent quintiles of years of adult education of the census block in which the student lives in.

We use the following specification to estimate heterogeneous effects on student immigrant status:

$$\begin{aligned} D_{ijt} =& \alpha + \beta Year2012_{ij} + \gamma Year2013_{ij} + \lambda Inm_{ijt} + \theta Inm_{ijt} * Year2012_{ij} + \\ & \sigma Inm_{ijt} * Year2013_{ij} + X_n + \epsilon_{ijt} \end{aligned} \tag{3.3}$$

where Inm_{ijt} is a dummy that takes value 1 if student i is an immigrant who applies for a school j in year t, and 0 otherwise. π_c and λ represent the average differences in the outcome variable for each category of student background status (parental education or

[28] Using students who are already in the system and apply for a school change may bias the results. First, after the first year of entrance to the system (at age 3), students have priority to remain at the same school (if they plan to continue in primary education at first grade). Students who enter the system after age 3 do not face the same set of feasible schools than students that enter the system at age 3 as they depend on school available slots due to current students leaving the school, under-demand of capacity, etc. Second, these families may have different preferences for schools. Third, these households may act even more strategically due to the fact that they potentially know better how the system works since they had previously applied. Hence, students who enter the system for the first time (at age 3) may have different preferences, priorities and behavior than students that aim to change the school later in the system, which may make it more difficult to compare both groups.

immigrant status) before the reform. Our parameters of interest are κ_c and η_c (θ and σ) for parental education (immigrant status), which capture the immediate heterogeneous responses of families choices after the reform.

3.5.2 School Segregation

The second goal of this paper is to estimate the effect of the change in priorities on the level of sorting and segregation of new students entering the school system. We test whether the flow of new entrants to the school system is now more or less socially and immigrant-origin concentrated within schools due to the policy reform. We refer to social diversity/heterogeneity (the opposite of social segregation) for parental education and immigrant segregation for immigrant-status segregation. We identify the reform effects net of other important factors such as school effects or dynamics in residential segregation of students (e.g. due to the increase in immigrant arrivals concentrated in some neighborhoods). The analysis on mobility outcomes may drive heterogeneous differences in responses to the reform that may lead to an impact on later stages of the education cycle (primary, secondary and post-secondary schooling) and affect opportunities for students to attend schools with students from diverse social and immigrant backgrounds.

Social Heterogeneity. We use a two-stage method to compute the change in within-school heterogeneity regarding student's parental education. Hence, what we are computing is an inverse measure of social segregation of schools. In the first step, we regress student's parental education on school dummies and neighborhood fixed effects of the school in a joint regression for 2011, 2012 and 2013. This can be seen as:

$$EDU_{ijt} = \alpha + \phi_j + \delta_n + \epsilon_{ijt} \tag{3.4}$$

where EDU_{ijt} are the number of parental years of education of student i applying for a school j in year t, ϕ_j are school fixed effects, and δ_n are neighborhood fixed effects. We take the absolute value of the student residual ($|\hat{\epsilon}_{ij}|$), and for a better interpretation, we use the logarithm of the residual in the second step. This provides a measure of the student parental education variation that cannot be explained by the school that the student is applying for and the neighborhood heterogeneity in parental education where the student resides.

In the second step, we regress this value on a year dummy to measure the average effect of the reform on school heterogeneity. In other words, to establish whether the unexplained variance of student parental education has changed due to the reform. Hence, the estimation can be described as follows:

$$log|\hat{\epsilon}_{ijt}| = \alpha + \theta Year2012_{ij} + \kappa Year2013_{ij} + \varepsilon_{ij} \tag{3.5}$$

where the coefficients θ and κ are the average change in the student social diversity or heterogeneity in 2012 and 2013. A positive value of the estimates would be interpreted as an increase in social heterogeneity (decrease in segregation), and a negative one as decrease in the level of social heterogeneity (raise in social segregation).

The residual $|\hat{\epsilon}_{ij}|$ of student i for school j reflects the level of student heterogeneity that may not be explained by the school fixed effect and the residential dynamics of the Madrid households between 2011 and 2013. The second estimation step regresses this measure of student heterogeneity on the post-reform year dummies net of other contextual factors such as the school and the neighborhood.

Immigrant-origin Segregation. A slightly different process is followed regarding the estimation of changes in the composition of immigrant population across the school network in the city of Madrid after the reform. While parental education (measured by years of schooling) is a continuous variable, immigrant status of the student is a dummy variable which is identified at the student level. This implies that, although the variable is very precise at the student level, we cannot compute a measure of immigrant diversity at the school level, but simply a measure of incidence of immigrant population for each school. Hence, we first consider the following linear model at the level of the school:

$$Im_{jt} = \alpha + \phi_j + \delta N_{jt} + \theta Year2012_j + \kappa Year2013_j + \epsilon_{jt} \tag{3.6}$$

where Im_{jt} is the proportion of immigrants among students applying for school j. ϕ_j are school fixed effects, and N_{jt} is the average proportion of immigrants in the neighborhood where school j is located in year t, which allows to control for residential dynamics of the immigrant population across neighborhoods. Moreover, θ and κ are the average change in the immigrant school composition in 2012 and 2013. To understand the differential effects among the distribution of immigrants across schools, we report the equivalent quantile regression of Equation 3.6 to measure the differential effect of 2012 and 2013 with respect to 2011 for each of the nine deciles of the distribution of the proportion of immigrants at each school.

3.5.3 Identification Challenges

A potential identification challenge may be that we cannot entirely capture the response in family preferences since we only observe the school ranked first in the submitted list. In fact, the school choice allocation mechanism plays a key role. A special characteristic of the BM is the fact that first choice is crucial for final allocation. Abdulkadiroglu and Sönmez (2003) show that this mechanism is not strategy-proof, implying that families have incentives to misreport their true preference list by ranking first a school where they have higher priorities, due to the fact that rounds are final in this algorithm.[29] Furthermore Abdulkadiroğlu et al. (2011) highlight that one of the features of the BM is that it better takes the students' cardinal preferences into account, compared to other algorithms with better ordinal desirable properties (e.g. Gale Shapley or Top Trading Cycles). Moreover, a particularity of Madrid is the fact that student's priority points are always measured for the school ranked first, which provides an additional incentive to families to think carefully about their first choice and potentially to apply for the school where they have highest chances. Table 3.4 shows the percentage of students who were assigned to the school they ranked first, those who went to a school they did not rank first, and those not assigned to any of their listed schools. It shows that about 86 percent of the students were assigned to the school they ranked first, and around 3 percent were not assigned to any of the schools families listed. These numbers are consistent with the results of Calsamiglia and Güell (2018) and Lavy (2010), where under the same allocation mechanism, find that between 85 and 90 percent of the students were assigned to the school they ranked first in Barcelona and Tel Aviv respectively. Hence, despite not being able to capture the entire information of family incentives, the school ranked first is crucial under the BM to understand preferences and provides a vital amount of information about them (while not so much under a truth-telling allocation mechanism). In Section 3.8, we consider an alternative specification of results with data from school assignment as opposed to data from the first choice of school.

[29] This allocation mechanism is not efficient, and it might create situations, called justified envies, when a student is rejected from a school while another one with lower priority is accepted.

TABLE 3.4: School Allocation in Madrid over 2010-2015.

Variable	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Assigned to first choice of school	15,640	15,286	15,703	14,845	14,253	1,4049
	[0.85]	[0.84]	[0.87]	[0.87]	[0.87]	[0.87]
Assigned to other school	2,226	2,203	1,961	1,693	1,604	1,539
	[0.12]	[0.12]	[0.11]	[0.1]	[0.1]	[0.09]
Not assigned to any ranked school	525	800	342	432	466	678
	[0.03]	[0.04]	[0.02]	[0.02]	[0.03]	[0.04]
Total Students	18,391	18,289	18,006	16,970	16,323	16,266

Notes: Each year corresponds to the year of application and the school year starting in September of that year.

A second potential identification threat would be the announcement of the reforms. The inter-district school choice reform was a proposal in the electoral program of the political party that won the elections in the region of Madrid in May 2011[30]. The information on the 2012 (and 2013) priority changes was disclosed to the press in February 2012, and announced formally through an administrative order in March 2012. A potential identification problem would arise if families reacted before the implementation of the reform (e.g. by changing their main household of residence). Hastings et al. (2009) find that the preferences for high ranked schools increase with proximity, and that parents with higher income are more willing to commute farther away for such schools. This increases the demand for districts with higher average public school performance that raise housing prices in those districts (see e.g. Fack and Grenet (2010)). A relaxation of the proximity priority criteria may reduce these incentives, and potentially stimulate families to change their main residence to districts with lower housing prices (typically positively correlated to lower school performance), since living in the school district of the desired school is not as decisive as previously in terms of admission probabilities. In this case, a potential positive significant effect on out of district application (and distance to school) may be correlated to this dispersion effect, leading to an upward bias in our estimates that would reflect an upper bound of the "true" effect. Applications must be handled between April and May of the academic year, implying that families did not have much time from March to May of 2012 to distort their decisions. However, they had about one year between the reform was formally

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

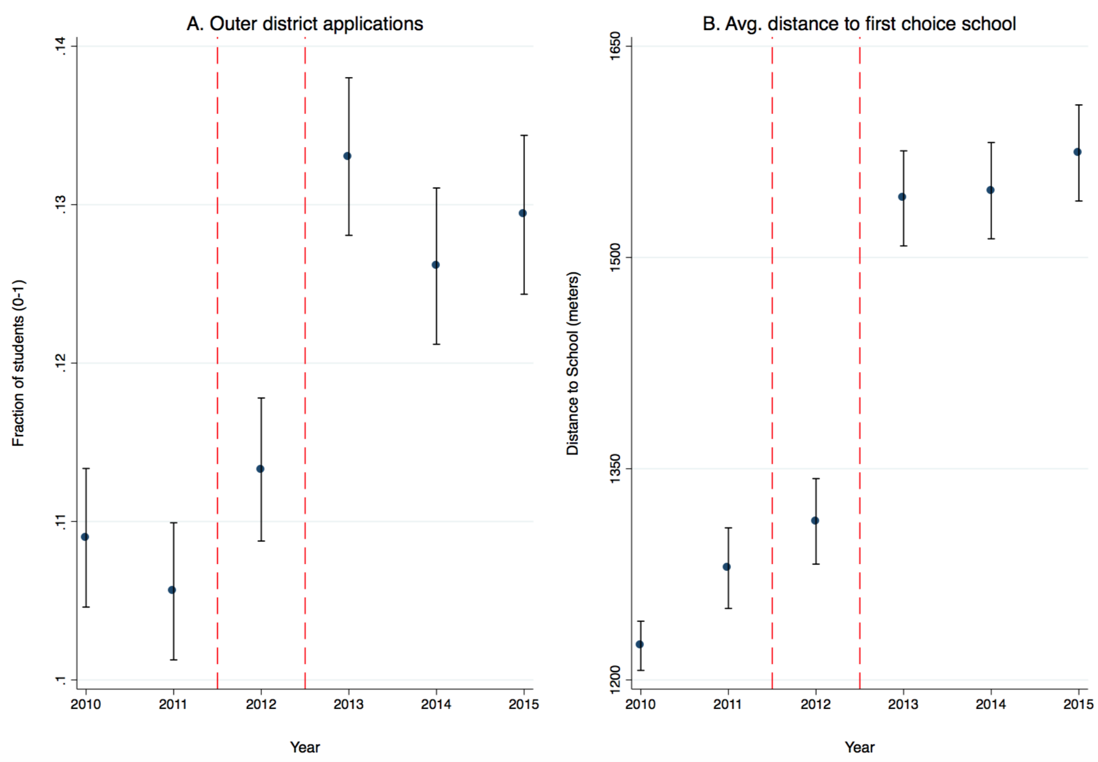
announced and the 2013 applications, potentially raising concerns about their anticipation.

3.6 Results

This section analyzes the changes in student mobility and student segregation induced by the school choice reform in the city of Madrid. First, we start by describing the changes in geographical mobility. Second, we investigate the effects of the reform on the within-school social and immigrant segregation. Third, we look at overall equilibrium effects by computing an index of segregation widely used in the literature.

3.6.1 Geographic Mobility

FIGURE 3.1: Measures of Student mobility to school before and after the reform.



Notes: Blue dots are point estimates. Horizontal lines are confidence intervals at 95%.

Figure 3.1 plots the average and confidence intervals of the estimates for two measures of geographic mobility (outer district application and distance traveled) between 2010 and 2015. The red stripes reflect the two consecutive changes in the choice priority system in 2012 and 2013. Panel A and B of the Figure show a significant increase in the level of mobility after the reform. In particular, the measure of intra-district mobility shows that almost 11% of parents had already aimed to bring their children to districts other than the ones they lived in (probably reflecting parental job mobility prior to the reform) by 2010. After the reform, this rose above 13% of applications, an increase of more than 2 percentage points (18%). Moreover, the measure of distance traveled to school increased from an average of 1,300 meters in 2010 and 2011 to almost 1,540 meters in 2013 and the subsequent years, a rise of 18%. For both indicators, the change seems to be fairly stable for all years after the reform. The most important moment when the change took place in 2013 and although 2012 already marked a change in the trend, it does not seem to be significant with respect to 2011.

Table 3.5 presents the results of the baseline regression and heterogeneous models described previously in Equations 3.1, 3.2 and 3.3. We present the results of the 2012 and 2013 effect using 2011 as the baseline year in all equations. Panel A shows that the reform increased the fraction of students applying to a school outside their district of residence by 1.2 and 3.3 percentage points in 2012 and 2013, respectively, at the 0.1 percent level of confidence. In addition, the policy raised the willingness to travel a greater distance by 11.6% in 2013. Panel B displays heterogeneous effects by parental education. It seems that families with more educated parents are more willing to both apply for an outer district school and commute a longer distance, relative to lower educated households. The effect of the reform in outer district applications was highly heterogeneous. While only the highest educated families (5^{th} quintile) responded positively to the changes in 2012, all education groups responded positively by 2013 but with different magnitudes: while the effect for the 1^{st} and 2^{nd} education quintile was a 1.3 percentage point increase, the magnitudes of the effects are increasing in the quintiles of the parental education distribution, with additional 2.2, 2.4 to 4.4 percentage points for 3^{rd}, 4^{th} and 5^{th} quintiles respectively. In contrast, the reform does not seem to affect the willingness to commute similarly: while all groups respond positively, the second and fifth education quintiles are more willing to commute after 2013 (by about an additional 7% to 8%) than the rest. Panel C also shows heterogeneous effects by immigrant status. Parents of immigrant students were less prone to apply for a school out of their district of residence (a difference of 4.2 percentage points compared to native students) and to commute (18.8% less relative to native students) prior to the reform. The reform affected positively both natives and immigrants in 2012 (1.6 percentage points) for outer district applications, and did not have a significant effect on commuting distance. In 2013, however, immigrants did not changed their behavior compared to 2011, whereas native students reacted positively to the reform (in about 4 percentage points) for outer district applications. In terms of the distance commuted, households with native students responded responded by a 12.9% greater willingness to commute, whereas immigrant students responded 6.1% less than their native peers, so that their aggregate response to the reform is positive and significant (6.8%), but smaller relative to native students.

Overall, there are two important findings. First, on average, the reform implied a positive response in terms of mobility for most households, both in distance traveled and outer-district commutes. Second, the effect was not homogeneous across social groups and native-immigrant population. In particular, students from the most educated households and native students where the ones that reacted positively to the reform, whereas students from less educated households or students from immigrant origins either displayed mild or negligible response to the reform. These results are consistent with previous evidence of other country reforms (e.g. Sweden[31]), showing that less advantaged and immigrant households tend to exert less (or simply do not strain any) choice than the rest of families when choosing for schools (which may reflect differences in preferences, information, etc.). This could have distributional implications that are further analyzed in the next section. Finally, the magnitude of the change in terms of the total student population is modest from a broader perspective. It is probably due to the weight of other key contextual factors, such as the 0.5 points that families still get when they reside in the district of first choice application, or/and the disutility that commuting provides to household's preferences.

[31] See Böhlmark and Lindahl (2007).

TABLE 3.5: Geographic Mobility Coefficient Effects

	Dependent variable					
	Outer district applications (Fraction)			Travel Distance (Log Meters)		
	2011	2012	2013	2011	2012	2013
		(1)			(2)	
A. Baseline Estimates						
Baseline Effect (vs 2011)	-	0.012***	0.0332***	-	0.0085	0.116***
	-	(0.034)	(0.0034)	-	(0.011)	(0.011)
B. Heterogeneous Effects: Parental Education						
		(3)			(4)	
Baseline Effect (vs 2011)	-	0.0083	0.0126*	-	0.014	0.0825***
	-	(0.0074)	(0.0076)	-	(0.007)	(0.0251)
2^{nd} quintile EDU	0.0143*	0.0004	0.0125	0.095***	0.0682*	0.0367
	(0.0075)	(0.0106)	(0.0108)	(0.0245)	(0.0351)	(0.0357)
3^{rd} quintile EDU	0.065***	0.0031	0.0218**	0.186***	0.0002	0.0008
	(0.0075)	(0.0106)	(0.0108)	(0.0245)	(0.0351)	(0.0357)
4^{th} quintile EDU	0.076***	-0.0083	0.0245**	0.267***	0.0265	0.050
	(0.0075)	(0.0106)	(0.0108)	(0.0245)	(0.0352)	(0.0357)
5^{th} quintile EDU	0.0877***	0.023**	0.044***	0.340***	0.0165	0.0795**
	(0.0075)	(0.0106)	(0.0107)	(0.0248)	(0.0352)	(0.0357)
C. Heterogeneous Effects: Immigrant Status						
		(5)			(6)	
Baseline Effect (vs 2011)	-	0.0160***	0.0393***	-	0.0173	0.129***
	-	(0.034)	(0.0037)	-	(0.0121)	(0.0122)
Immigrant student	-0.042***	-0.0148	-0.0320***	-0.188***	-0.0100	-0.0607*
	(0.0071)	(0.0096)	(0.0098)	(0.023)	(0.032)	(0.032)
Number of Obs.		53,220			53,220	

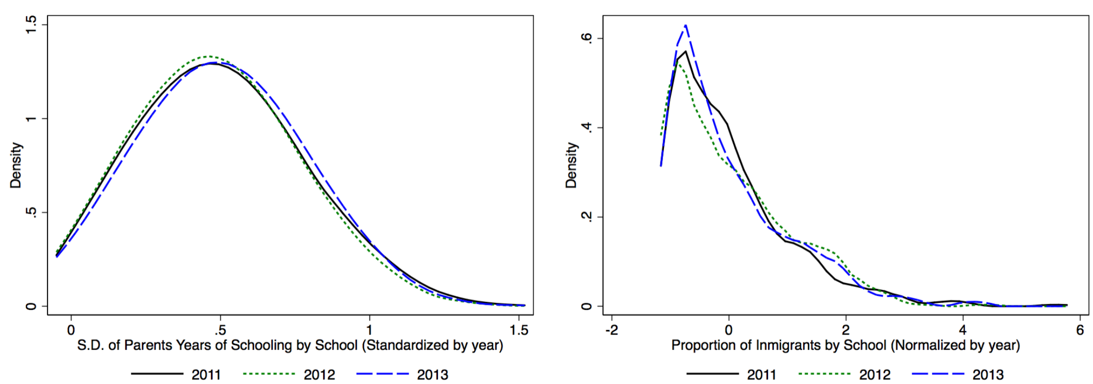
Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Each panel represents regressions from Equation (1), (2) and (3) respectively. In Equations (2) and (3) heterogeneous effects need to be interpreted as an addition to the baseline effect. We control for the population density in the student neighborhood. Clustering errors at the district or neighborhood levels does not alter the significance level of the results.

3.6.2 Estimating student segregation

The second goal of this paper is to identify the marginal impact that the reforms had on the level of segregation of new students entering the school system, net of the dynamics of school characteristics and residential segregation. This is relevant, especially in immigrant segregation, given the increase observed in immigrant population in 2012 and 2013. We consider segregation in two dimensions: social (estimated through parental education) and immigrant-origin segregation. The previous analysis shows that after the implementation of the inter-district school choice reform, there was an overall increase in the preference for mobility of families, but this change was heterogeneous and more intense for middle-educated families as well as those with non-immigrant children. A heterogeneous change in choice may make that the flow of new entrants to the school system more socially and immigrant-origin segregated, which may impact on the opportunities of students to attend schools with students with different backgrounds.

FIGURE 3.2: Densities of within-school socioeconomic variation and immigrant share, by year.

(A) Parental Education.　　　(B) Immigrant Origin.



We first present descriptive evidence of the dynamics in student composition at schools. Figure 3.2 displays the density of within-school variation of student background for 2011, 2012 and 2013. The solid line represents the density for 2011, while the short-dashed and long-dashed lines refer to 2012 and 2013 respectively. The left-hand side graph (Figure 3.2a) shows the density of within-school standard deviation of parents' years of education. The curve shifts mildly to the left in 2012 (with respect to 2011), whereas the 2013 density curve shifts to the right in the low and middle part of the distribution indicating a mild increase in within-school student heterogeneity in terms of parental education. The right-hand side graph (Figure 3.2b) plots the distribution of the proportion of immigrants within school normalized by the total share of immigrants in the city.[32] This Figure shows a polarization of the distribution in 2012 and 2013: while the share of schools with very low proportion of immigrants slightly increases, the proportion of schools with a median proportion of immigrants decreases in 2012 and 2013. At the same time, there was an increase in the share of schools with higher proportion of immigrants (usually more than 30% of immigrants). Whereas the evidence in terms of social segregation is mild, the second graph describes a process of immigrant segregation across schools at the time of the reform.

TABLE 3.6: Change in within-school student heterogeneity by parental education.

Dependent variable: $log\vert\hat{\epsilon}_{i,j}\vert$	(2012)	(2013)
Specification 1	0.0059	0.0517***
	(0.0127)	(0.0129)
Specification 2	-0.134	0.043***
	(0.0128)	(0.013)
Specification 3	0.0059	0.0517***
	(0.0127)	(0.0129)

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively.

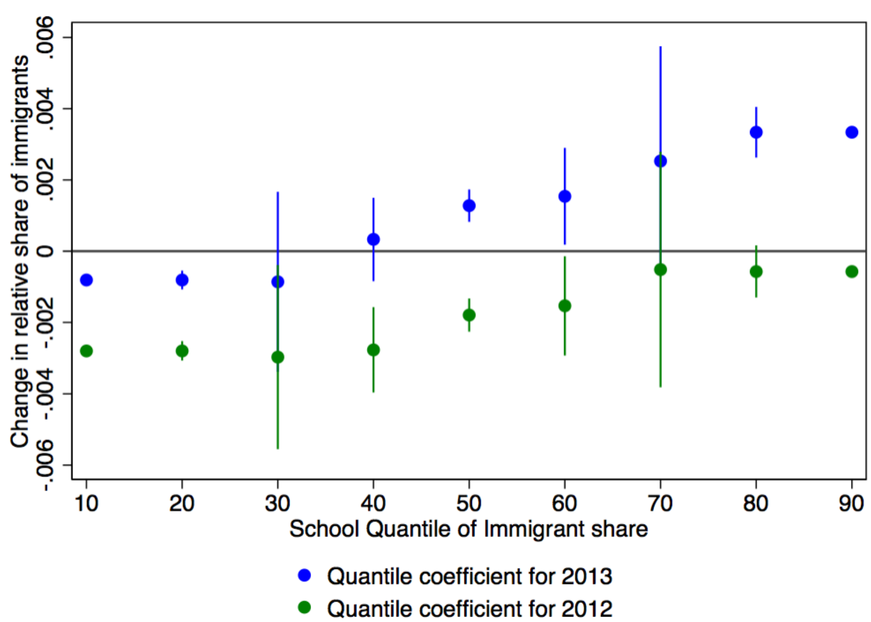
Social Heterogeneity (by parental education). Table 3.6 presents the results for the estimations of the model of Equation 3.5 model. We estimate changes in the within-school heterogeneity of parental education of students. It compares the marginal effect of 2012 and 2013 relative to 2011, controlling for school fixed effects and neighborhood dynamics of residential segregation. We conduct three different specifications: (i) we compute the residuals in a joint regression for the three years, controlling for neighborhood heterogeneity of parental education; (ii) we calculate residuals in separate year regressions, controlling for neighborhood heterogeneity of parental education; (iii) we measure residuals in a joint regression for all three years, controlling for district heterogeneity of parental education. The Table shows that the changes in within-school heterogeneity of parental education did not change in 2012, but increased significantly in 2013 with respect to 2011. The magnitude of the effect is between 4% and 5%. In other

[32] We standardize the standard deviation of parental education by dividing by the total yearly standard deviation of education. Conversely, we normalize the proportion of immigrants by subtracting the yearly mean and dividing by the average mean of immigrants in the system. This is because the socioeconomic characteristics of parents may not be constant in the 2011-2013 period.

words, it seems that the reform increased diversity within school by 4 to 5%. Nevertheless, the measurement error of the parental education variable may be correlated with the direction of the reform, inducing a positive (negative) bias in our heterogeneity (segregation) estimates that we discuss in Section 3.8.

Immigrant Segregation. Figure 3.3 displays the parameter of interest θ and κ in Equation 3.6 for each decile of the distribution of quantiles. The Figure shows that, conditional on school fixed effects and immigrant share in the school neighborhood, the relative proportion of immigrants in 2013 seems to increase for schools with a larger share of immigrants, while it decreased for schools with a lower share of immigrants. In 2012, the relative proportion of immigrants decreased in schools with a smaller share of immigrants, while those with the highest share of immigrants did not experience relevant changes. Nevertheless, the size of estimates is relatively small, mostly because accounting for school and neighborhood effects reduces the magnitude of the coefficients[33], an issue which suggests that geographical segregation of immigrants was also an important factor during the period of the reform.

FIGURE 3.3: Post-Reform effects on share of immigrant students, by quantiles.



[33]Estimates without controlling for school or neighborhood effects were larger in magnitude for all quantiles, both for positive values (at schools with higher share of immigrants) and negative values (at schools with lower share of immigrants). Results can be provided upon request.

Computing Overall Segregation Indexes. Although we cannot provide causal evidence on general equilibrium effects of the reform (due to other confounding factors related to school effects and geographic mobility taking place at the same time), we aim to assess the potential effects of the reform from a broader policy perspective.[34] To measure the overall change in social and immigrant segregation, we compute the Dissimilarity Index (Duncan and Duncan (1955)), one of the most widely used measures in school segregation literature. For a two-type groups of students a_i and b_i in the school system, the "D" Index is defined as:

$$D = \frac{1}{2} \sum_{i=1}^{I} \left| \frac{a_i}{A} - \frac{b_i}{B} \right| \tag{3.7}$$

where there are I schools in the city, and school i has a_i students of the first category and b_i of the second category. The value of D represents the proportion of a group that would need to move in order to create a uniform distribution of population across schools. Thus, this index will go from 0 to 1. The index is equal to 0 when the educational system is fully heterogeneous across schools and 1 if the system is completely segregated. This Index can be generalized to be a multi-group index, following Reardon and Firebaugh (2002) and can be interpreted as the percentage of all students who would have to transfer between schools to equalize the group proportion across schools, divided by the percentage of students who would have to transfer if the school system was fully segregated. For parental education, we divide it into five quintiles and compute the multi-group version of the index.[35].

TABLE 3.7: Evolution of school segregation during the reform in the city of Madrid.

	Social Segregation			Immigrant Segregation		
	2011	2012	2013	2011	2012	2013
Dissimilarity Index Estimate	0.495	0.501	0.488	0.421	0.449*	0.467*
	(0.0028)	(0.0029)	(0.0018)	(0.0058)	(0.0084)	(0.0072)
Number of Obs.	18,289	18,006	16,970	18,289	18,006	16,970

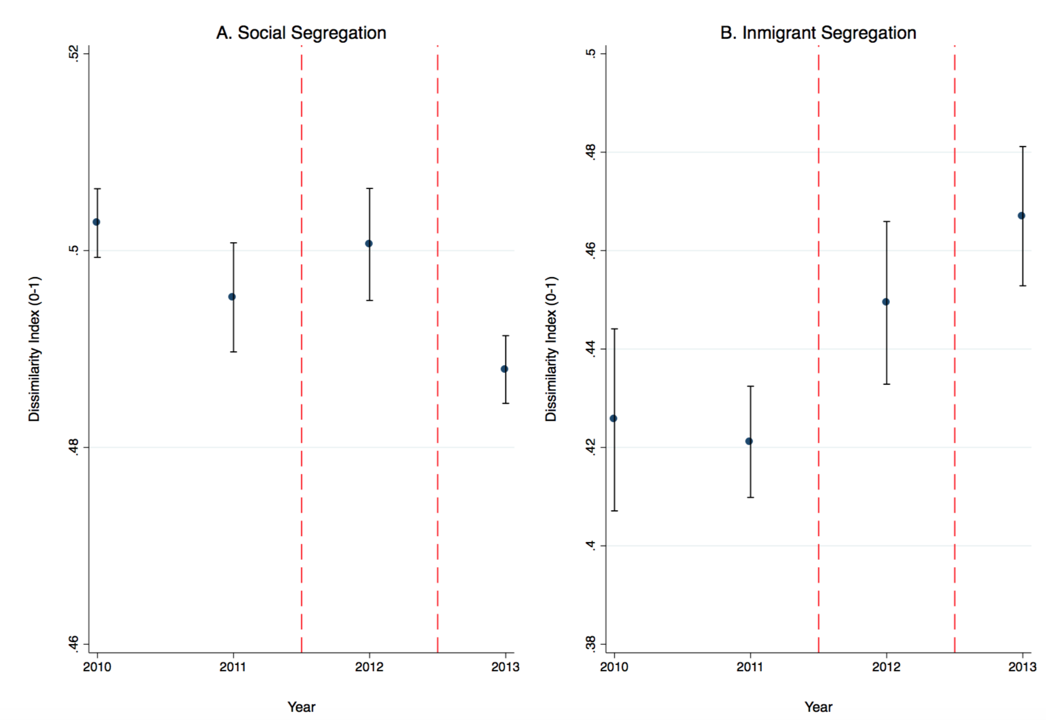
Notes: Standard errors in parentheses are computed by bootstrapping individual data. We perform 20 bootstrap replications for each estimate. * indicate significance change at the at 5 percent level between 2011 and 2013.

[34]The validity of the conclusions draw in this subsection has to be considered in an appropriate context.

[35]To calculate standard errors, we compute this measures for years 2011, 2012 and 2013 using the STATA commands *localseg* for a multi-group model (parental education quintile groups) and *dicseg* for a two-group model (immigrant status of students).

Table 3.7 presents the values of the index for both measures of student background before and after the reform. Results show that social segregation seems to be constant across the years of the reform, given that there is a mild decrease in the point estimate which is not statistically significant. Nonetheless, there is a considerable increase of immigrant segregation (about 11%), which is significant both in 2012 and 2013 relative to 2011. This can be better seen in Figure 3.4 below. From a broader perspective, these results imply a large increase in the overall immigrant segregation and a rather stable dynamics of social segregation. The driving mechanisms may go beyond the impact of the school choice reform (e.g. changes in residential segregation patterns), and hence we cannot conclude that there is a causality link between the policy reform and the aggregate effect.

FIGURE 3.4: Computing overall Segregation (Social and Immigrant).



Notes: Blue dots represent point estimates. Horizontal lines depict upper and lower confidence intervals at the 95% of confidence.

3.7 Mechanisms

This section discusses the mechanisms that may be driving the results. First, we estimate the impact of the reform on mobility between and within district in order to capture direct and indirect effects of the reform. Second, we investigate heterogeneity of the results by school characteristics, such as school average score, income of school district, or school ownership (public vs semi-public). Finally, we examine to which extent the implementation of the bilingual education programme was a confounding factor for our baseline results.

3.7.1 Extensive and Intensive Margin of Geographic Mobility

The inter-district school choice reform in Madrid has been shown to improve the access to schools outside household's districts, as the willingness to commute out of the district increases. However, it is not clear whether all the changes in distance commuted are exclusively driven by this or, in addition, the willingness to commute to another district is also accompanied by changes within district. If the total distance commuted increases, it could be the joint effect of more mobility within and between districts. In particular, if households aim to move towards other districts, this may create an emerging supply of places in their district and a subsequent internal demand for them. Therefore, the inter-district school choice may be jointly operating in the extensive (between districts) and intensive (within districts) margin of mobility. In order to disentangle these two levels of the mechanism through which mobility is generated, we compute estimates of distance mobility and school segregation considering within and between district applications separately.

Table 3.8 shows the baseline estimates of willingness to commute for households who apply for an outer or inner district school. Results are mainly driven by households that are willing to move out of their district. Households moving to another district increase their willingness to commute by 31% after 2013, as opposed to the average of 11.6% seen in Table 3.5. A small effect remains in the intensive margin, since the effect on those who are willing to stay in their own district increases in a magnitude of 4.6%. This could be the consequence of the potential new supply of places at schools in their same districts (i.e. indirect effect), so that families are willing to go to these schools that are farther away from their homes. The extensive margin effect is present in all educational groups, although the most educated group of households displays a very large effect on commuting distance: an additional increase of 25.6% relative to other households and hence a total of 47% with respect to the beginning of the reform. No differences regarding immigrant and native status of the student are found for mobility of students applying to schools inside and outside their own district.

TABLE 3.8: Geographic Mobility Coefficient Effects - Intensive and Extensive Margin

	Dependent variable: Travel Distance (Log Meters)					
	Inner district applications			**Outer district applications**		
	2011	2012	2013	2011	2012	2013
		(1)			(2)	
A. Baseline Estimates						
Baseline Effect (vs 2011)	-	-0.007	0.0459***	-	0.0263	0.315***
	-	(0.012)	(0.012)	-	(0.023)	(0.023)
B. Heterogeneous Effects: Parental Education						
		(3)			(4)	
Baseline Effect (vs 2011)	-	-0.0286	0.0554**	-	0.0141	0.211***
	-	(0.025)	(0.025)	-	(0.075)	(0.074)
2^{nd} quintile EDU	0.084***	0.072*	0.0235	-0.136*	0.0705	0.0238
	(0.025)	(0.036)	(0.036)	(0.071)	(0.09)	(0.096)
3^{rd} quintile EDU	0.139***	-0.0038	-0.0337	-0.143**	-0.042	0.0006
	(0.025)	(0.036)	(0.036)	(0.0645)	(0.089)	(0.088)
4^{th} quintile EDU	0.209***	-0.053	-0.0044	-0.0477	-0.0753	0.124
	(0.025)	(0.036)	(0.037)	(0.064)	(0.089)	(0.087)
5^{th} quintile EDU	0.286***	-0.0112	-0.0296	0.0339	0.0165	0.256***
	(0.025)	(0.0363)	(0.037)	(0.064)	(0.086)	(0.086)
C. Heterogeneous Effects: Immigrant Status						
		(5)			(6)	
Baseline Effect (vs 2011)	-	-0.00016	0.051***	-	0.023	0.322***
	-	(0.012)	(0.013)	-	(0.0245)	(0.025)
Immigrant student	-0.154*	-0.00089	-0.0017	-0.0424	-0.0424	-0.0703
	(0.023)	(0.032)	(0.033)	(0.06)	(0.08)	(0.081)
Number of Obs.		46,992			6,095	

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Each panel represents regressions from Equation (1), (2) and (3) respectively. In Equations (2) and (3) heterogeneous effects need to be interpreted as an addition to the baseline effect. We control for the population density in the student neighborhood. Clustering errors at the district or neighborhood levels does not alter the significance level of the results.

We then compute the differential effect of the extensive versus intensive margin on student segregation at schools. We follow a similar approach as the one used in the previous section, but with separate segregation estimates depending whether their own residence is in the district of choice or in a different one. For example, when looking at social heterogeneity (by parental education of students) as in Table 3.6, the statistical significance of the increase observed does not hold for students who move to another district, as the p-value is larger than 0.1 (although the point estimate is positive and larger - around 6%-). With respect to students staying in the same district, the effect remains significant at 90%, but the magnitude goes down up to 2.3%[36]. All in all, it seems that the mild increase in student diversity is rather associated to the intensive margin effects of the reform. With respect to immigrant segregation, we run Equation 3.6, but for each school we control for the share of students who come from a different district: the results are qualitatively the same compared to those in Figure 3.3[37]. Overall, the results in terms of mobility and segregation are mostly driven by the extensive margin of student mobility to other districts, rather than the intensive margin.

3.7.2 Type of School and District

In this subsection, we estimate differences experienced in mobility and school social and immigrant segregation over key dimensions of the school system, such as the characteristics of the district where the school is located, the school performance in the standardized test, the socioeconomic characteristics (parental years of education and immigration status) of students entering the school previously, and the type of school in terms of school ownership (i.e. public versus semi-public schools). This is fundamental in order to better understand what contextual features of the school system are driving the policy responses of the households.

In terms of the student mobility, given the importance shown of the extensive margin in the final effects, we only examine the outer district applications. We update Equations 3.1, 3.2 and 3.3 in order to allow for interactions with school characteristics. We

[36]Results can be provided by the authors upon request.
[37]Results can be provided by the authors upon request.

use the following specification:

$$
\begin{aligned}
D_{ijt} =& \alpha + \beta Year2012_{ij} + \gamma Year2013_{ij} + \sum_k \gamma_j Z_{ijt} + \theta_k \sum_k Z_{ijt} * Year2012_{ij} + \\
& \kappa_k \sum_k Z_{ijt} * Year2013_{ij} + X_n + \varepsilon_{ijt}
\end{aligned}
\tag{3.8}
$$

where Z_{ijk} is a vector of school observable characteristics for the first choice school j of student i in year t. We investigate five school variables (with values over k): (i) average income of the school district, dividing districts into three groups of income; (ii) school score in the previous year's standardized test of the region of Madrid normalized for each year, information that was publicly available to parents during their choice process and which takes values from 0 to 10[38]; (iii) school average years of parental education of students entering the school in the previous year; (iv) school proportion of immigrant students entering the school in the previous year; (v) school ownership (public vs semi-public).

We implement a similar specification regarding the estimation of student segregation at schools net of other factors, both in terms of social segregation (Equation 3.11) and immigrant segregation (Equation 3.10):

$$
\begin{aligned}
log|\hat{\epsilon}_{ijt}| =& \alpha + \theta Year2012_{ij} + \kappa Year2013_{ij} + \sum_k \gamma_k Z_k + \theta_k \sum_k Z_k * Year2012_{ij} + \\
& \kappa_k \sum_k Z_k * Year2013_{ij} + \varepsilon_{ij}
\end{aligned}
\tag{3.9}
$$

$$
\begin{aligned}
Im_j =& \alpha_\tau + \theta_\tau Year2012_{ij} + \kappa_\tau Year2013_{ij} + \theta_{k,\tau} \sum_k Z_k * Year2012_j + \kappa_{k,\tau} \sum_k Z_k * Year2013_j \\
& + \theta_{j,\tau} + \delta_\tau N_j + \epsilon_{j,\tau}
\end{aligned}
\tag{3.10}
$$

Geographic Mobility. The results in Table 3.9 show the existence of cream-skimming mechanisms in the baseline estimates of outer district applications by different school characteristics. Panel A displays heterogeneous effects by dividing districts in terciles according to their average income distribution[39]. Families are more willing to leave

[38] For instance, we use CDI results of 2012/2013 for the school year 2013/2014 applications.

[39] For a detailed description of district classification according to income, see Figure 3.14 in the Appendix.

their district for schools located in richer districts (particularly for schools in districts within the top income tercile) relative to poorer (in 3.2% for middle income and 15.6% for high income districts). The reform increased the proportion of applications to a school outside their district, but the willingness to commute to a school placed in a high or middle relative to a low-income district did not change. Panel B presents heterogeneous effects by the average school performance in *CDI*. Households seem to prefer to leave their district for schools with higher performance score (i.e. for an increase of 1 standard deviation in the school exam, parents are more willing to move from their district in 3.3 percentage points). This preference is heterogeneous across school performance score quintiles: family willingness to change their district is larger for schools for which the performance score quintile is higher, although the reform did increase this preference only for schools with scores in the 4^{th} quintile [40]. Panel C analyzes heterogeneous effects by school ownership (public vs semi-public): while families seem to be 5.8% more willing to leave their districts for a semi-public school, the reform exacerbated these differences by an additional 3.4 and 3.1 percentage points in 2012 and 2013 respectively.

Panel D and E investigate the student socioeconomic characteristics at the school (parental years of education and immigration status) for those students who applied in the previous academic year. Although this information is not public to parents, it serves as a proxy for student socioeconomic characteristics entering the school, an issue which the literature has documented to be important in parental preferences (Mizala and Urquiola (2013)). In particular, Panel D looks at whether there are heterogeneous effects by the average years of parental education at the school: families seem to be more willing to leave their district for schools where the average parental years of education of peers is higher (i.e. for an increase of 1 year of school average parental education, families are more willing to leave the district by 2.4% percentage points). What is more important, the reform increases the willingness to apply for schools with higher average parental education of peers by an additional 0.4 percentage points per year of education. Similarly, Panel E examines heterogeneous effects by average proportion of immigrants at the school. For each decrease of a percentage point of immigrants at the school, parents are 0.23% more willing to leave the district to apply for such a school:

[40] See Table 3.15 in the Appendix. We observe a differential effect for the 4^{th} quintile of schools by 2.9%. Before the reform, the willingness to commute is higher by 4.7%, 7.5%, 7.6% and 9.3% for for 2^{nd}, 3^{rd}, 4^{th} and 5^{th} quintiles with respect to the 1^{st}.

the reform increased even more such preferences by an additional 0.09% for each percentage point of immigrants lower at a school. Overall, these two findings reveal that what parents reacting to the reform are looking for is a combination of high average scores and high socioeconomic status of peers or semi-public schools.

TABLE 3.9: Mechanisms for Geographic Mobility, by School Characteristics

Dependent variable: Outer District Applications (Fraction)

A. District Income	
Constant	0.046***
	(0.004)
Year 2012	0.012**
	(0.005)
Year 2013	0.024***
	(0.005)
Middle Income Dist.	0.032***
	(0.005)
High Income Dist.	0.156***
	(0.006)
2012*Middle Income Dist.	0.000
	(0.008)
2012*High Income Dist.	-0.004
	(0.008)
2013*Middle Income Dist.	0.012
	(0.008)
2013*High Income Dist.	0.009
	(0.008)
Number of Obs.	53,22

B. School Grade	
Constant	0.065***
	(0.003)
Year 2012	0.013***
	(0.003)
Year 2013	0.033***
	(0.003)
Sch. Grade (t-1)	0.033***
	(0.003)
2012*Sch. Grade (t-1)	0.003
	(0.004)
2013*Sch. Grade (t-1)	0.004
	(0.004)
Number of Obs.	53,22

C. School ownership	
Constant	0.040***
	(0.004)
Year 2012	-0.005
	(0.005)
Year 2013	0.017***
	(0.005)
Semi-Public	0.058***
	(0.005)
2012*Semi-Public	0.034***
	(0.007)
2013*Semi-Public	0.031***
	(0.007)
Number of Obs.	53,220

D. Years of Education	
Constant	-0.206***
	(0.013)
Year 2012	-0.004
	(0.018)
Year 2013	-0.023
	(0.018)
Avg.Years Education (t-1)	0.024***
	(0.001)
2012*Avg.Years Education (t-1)	0.001
	(0.002)
2013*Avg.Years Education (t-1)	0.004***
	(0.002)
Number of Obs.	52,468

E. Proportion of Immigrants	
Constant	0.084***
	(0.004)
Year 2012	0.012**
	(0.005)
Year 2013	0.056***
	(0.005)
Avg. Frac. Inmig. (t-1)	-0.233***
	(0.019)
2012*Avg. Frac. Inmig. (t-1)	-0.018
	(0.027)
2013*Avg. Frac. Inmig. (t-1)	-0.093***
	(0.025)
Number of Obs.	52,468

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. We control for the population density in the student neighborhood. Clustering errors at the district or neighborhood levels does not alter the significance level of the results.

TABLE 3.10: Mechanisms for Social Heterogeneity, by school characteristics

Dependent variable: Social Segregation					
A. District Income		**B. School Grade**		**C. Type of School**	
Constant	-0.636***	Constant	-0.514***	Constant	-0.470***
	(0.014)		(0.009)		(0.013)
Year 2012	0.003	Year 2012	0.008	Year 2012	0.005
	(0.019)		(0.013)		(0.018)
Year 2013	0.031	Year 2013	0.045***	Year 2013	0.014
	(0.020)		(0.013)		(0.018)
Middle Income Dist.	0.337***	Sch. Grade (t-1)	0.097***	Semi-Public	-0.056***
	(0.021)		(0.009)		(0.018)
High Income Dist.	0.112***	2012*Sch. Grade (t-1)	0.026*	2012*Semi-Public	0.001
	(0.022)		(0.013)		(0.025)
2012*Middle Income Dist.	-0.000	2013*Sch. Grade (t-1)	0.054***	2013*Semi-Public	0.075***
	(0.030)		(0.014)		(0.026)
2012*High Income Dist.	0.010				
	(0.031)				
2013*Middle Income Dist.	0.029				
	(0.030)				
2013*High Income Dist.	0.024				
	(0.032)				
Number of Obs.	53,017	Number of Obs.	53,205	Number of Obs.	53,205
D. Years of Education		**E. Proportion of Immigrants**			
Constant	-1.236***	Constant	-0.321***		
	(0.048)		(0.012)		
Year 2012	-0.008	Year 2012	0.008		
	(0.068)		(0.018)		
Year 2013	0.023	Year 2013	0.129***		
	(0.069)		(0.018)		
Avg.Years Education (t-1)	0.063***	Avg. Frac. Inmig. (t-1)	-1.458***		
	(0.004)		(0.071)		
2012*Avg.Years Education (t-1)	0.001	2012*Avg. Frac. Inmig. (t-1)	0.003		
	(0.006)		(0.100)		
2013*Avg.Years Education (t-1)	0.002	2013*Avg. Frac. Inmig. (t-1)	-0.134		
	(0.006)		(0.092)		
Number of Obs.	52,454	Number of Obs.	52,454		

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively.

Social Heterogeneity (or Social Segregation). Table 3.10 presents the baseline estimates of changes in social heterogeneity by different school characteristics. Panel A shows that social diversity within school is higher in middle income (31%) and high income (25.5%) districts relative to low income districts, but the reform did not change the within-school heterogeneity of parental education differentially by income district. In Panel B, it can be seen that schools with a standard deviation higher in the standardized school score has around 10% more heterogeneity in parental education. This fact is heterogeneous across school performance score quintiles: the social diversity is larger as the school performance score quintile is higher (17.8%, 21.5%, 30.2% and 28.2% for for 2^{nd}, 3^{rd}, 4^{th} and 5^{th} quintiles respectively). The reform increased diversity for schools with higher standardized score by 5.4 percentage points for each standard deviation of tests. When looking at this more in depth, we observe a decrease in the social diversity (an increase in the level of segregation) for schools of the first quintile of the average performance score by 6.1 percentage points, while an increase in social diversity by 12.3, 15.8, and 17.5 percentage points for schools with performance in the 3^{rd}, 4^{th} and 5^{th} quintiles respectively[41]. Panel C shows that semi-public are 5.6% less diverse than public schools. The reform increased the within-school heterogeneity of parental education of semi-public schools relative to public schools by 7.5 percentage points. Panel D and E show that schools with a higher average years of parental education are more diverse, and schools with higher proportion of immigrants are less heterogeneous. Neither of these two relations changed with the reform. In any case, the measurement error may be correlated with the direction of the reform, inducing a bias in our estimates that is discussed in Section 3.8.

Immigrant Segregation. The increase in the immigrant segregation due to the reform is not related to the years of education or the proportion of immigrants at the school. However, there is a strong relationship between the increase in immigrant segregation and the school standardized score in the previous year's test: segregation decreased in schools with higher grades.[42]

Segregation Index. Finally, Table 3.17 looks at the overall evolution of segregation by looking at the Dissimilarity Index in different school characteristics. Although the

[41] See Table 3.15 in the Appendix.
[42] See Table 3.16 in the Appendix.

statistical significance is low (due to a smaller size of schools), point estimates of social segregation (the opposite measure of social diversity) increase for low scoring schools and decrease for high scoring schools. These are rather stable in terms of district income, and decrease for semi-public schools and bilingual and non-bilingual schools. In contrast, the increase observed in immigrant segregation is mostly concentrated and statistically significant in the 3rd quintile of the school test distribution, in low income districts, in both semi-public and public schools and in bilingual schools.

The role of semi-public schools. Overall, the previous results show that there is clear evidence pointing to the role of semi-public schools in driving the increase in the willingness to change the district to attend a semi-public school during the reform and this holds in explaining the rise in social diversity. However, it is not clear whether this is related to other school characteristics or to the genuine interest of families in semi-public schools. In order to better understand which factors may be behind the raise in the preference for semi-public schools, we look at the interactions of several school characteristics and semi-public schools (see Table 3.18 in the Appendix). Overall, we find that neither school test scores, nor school average years of education, nor the proportion of immigrants in the school in previous years seems to explain the increase in the willingness to travel to semi-public schools due to the reform. Hence, the greater preference for semi-public schools seems to be due to other semi-public schools' unobserved factors that we cannot identify (such as pedagogical approaches or other specific characteristics of semi-public schools).

3.7.3 Expansion of Bilingual Schools

Since 2004/2005, the Regional Government of Madrid implemented a new program of bilingual education for publicly funded schools where a substantial part of the subjects started to be taught in English language. Since then, the program experienced a large expansion. In the first year of implementation, 26 schools joined the program, and by 2014/2015 there were 298 bilingual public and 122 semi-public schools for primary grades for all the municipalities of the region. In particular, in our baseline sample for the city of Madrid, the number of bilingual primary schools expanded from 144 in 2011 to 182 in 2013: out of these new 38 schools 10 were public and 28 semi-public. Previous evidence has shown that the expansion of the bilingual network among public schools in the early years of the program was potentially associated with student sorting according to their socioeconomic status (Anghel et al. (2016)). Hence, we analyze whether the programme have confounding effects with the school choice reform on student mobility and student segregation across schools. We follow a similar approach as in the previous section in order to track interaction effects of bilingual schools on mobility and segregation outcomes.

Table 3.19 in the Appendix presents the results of the analysis for a differential effect of bilingual schools in terms of mobility and segregation. Families are willing to travel 13.2% more to attend a bilingual school, and these schools seem to be 21% more socially diverse than non-bilingual schools. Nevertheless, the reform did not seem to significantly affect the willingness to change the district or to commute further to attend bilingual schools. Although the effect significance is mild (10 percent confidence level), the reform appears to have increased the willingness to travel to bilingual schools by 4.1 percentage points. Overall, the implementation of the bilingual program does not seem to be a crucial confounding factor for the school choice reform. Moreover, the changes in within-school social and immigrant heterogeneity are not related to the expansion of the programme.

3.8 Robustness Checks

In this section, we conduct a series of robustness checks looking at other factors that could influence student's mobility and segregation effects during the reform. First, we analyze the fact that the choice reform in the city of Madrid was implemented in two steps (first criteria to be reformed in 2012, whereas the inter-district reform was implemented in 2013) to isolate the real effect of inter-district reform. Second, we consider other municipalities of the region as a control group to validate the results found in the city of Madrid. Third, we test whether results are robust for considering the final school assignment of schools instead of the first school choice as the main outcome. Finally, we discuss potential unobserved heterogeneity issues on parental education data at the census block level.

3.8.1 Reform Implementation in Two Years: Low-income and Former Student Bonus

A potential issue in our identification strategy is that changes in the priority criteria introduced in the city of Madrid in 2012 may be driving part of the effect of the 2013's inter-district reform. As mentioned in Section 3.3, the reform introduced changes in 2012 in two key criteria (i.e. low-income status and bonus points for former students at the school). Table 3.5 shows that higher educated parents and parents of native-born students already reacted to the 2012 reform. However, the overall changes in mobility mostly took place in 2013 compared to 2012. Hence, we investigate whether the changes observed in 2013 with respect to 2012 in terms of mobility in the extensive and intensive margin are related to the criteria introduced in 2012.

Table 3.20 in the Appendix presents the change in mobility of new students in 2013 with respect to those of 2012. We look at the interaction term of the new criteria introduced in 2012. We find that changes in mobility observed in 2013 are independent of the new 2012 criteria. Hence, the introduction of the new 2012 bonus criteria is unrelated to the large effects observed in 2013 in terms of mobility, which again suggests that the mobility response is closely linked to the proximity criteria reform in 2013. Results are robust to both measures of mobility.

3.8.2 Control Group of Other Municipalities in the Region

As described in Section 3.3, the school choice reform was also implemented in other municipalities beyond the city of Madrid. Table 3.1 reflects the implementation of most of the new criteria in 2012 for over-demanded schools in the city of Madrid: these new criteria were also applied in the rest of municipalities. The only criteria that changed in two separate school years depending on the municipality was the inter-district criterion: for some municipalities (usually of medium size), the proximity inter-district reform was implemented in 2012, whereas for the larger ones (including Madrid), it took place in 2013[43]. We exploit this heterogeneity in joining the inter-district school choice to analyze the effects on other municipalities as a robustness check. In particular, we pay attention to the fact that households in municipalities that implemented the

[43]Table 3.13 in the Appendix describes which municipalities joined the inter-district criterion in 2012 and 2013.

inter-district reform in 2012 did not have time to anticipate the reform implementation.

For the control groups, we consider municipalities that have similar geographical characteristics compared to the city of Madrid, as these condition mobility patterns of households. We use three criteria: population density, average distance commuted by households prior to the reform, and total number of year applications. In particular, the city of Madrid is highly dense (4,835 inhabitants per km^2) and the average distance commuted before the reform was fairly low (1,500 meters). With respect to the total number of applications, the rest of municipalities have much lower populations than Madrid (with a population between 70,000 and 205,000, whereas Madrid's population is around 3.1 million): this implies that the total number of applications is much lower compared to that of the city of Madrid. Among such a group, we use municipalities that have at least 800 yearly applications for more than 25 schools in each municipality. Applying these filters, the final sample of municipalities which implemented the inter-district school choice reform includes two in 2012 (which we call as Group 1) and eight in 2013 (Group 2):

- **Group 1 municipalities (reform in 2012/2013)**: Alcobendas and Valdemoro.
- **Group 2 municipalities (reform in 2013/2014)**: Alcalá de Henares, Alcorcón, Fuenlabrada, Getafe, Leganés, Móstoles, Parla and Torrejón de Ardoz.
- **Group 3 municipality (reform in 2013/2014)**: Madrid.

In terms of mobility outcomes, we can only measure distance traveled given that we do not have access to geographical data on the catchment areas in other municipalities prior to the reform[44]. Regarding our empirical strategy, it has to be noted that two reforms were gradually introduced: (i) low-income and former student bonus criteria (in 2012 for all municipalities); (ii) inter-district school choice. For municipalities of Group 1, we are then measuring an effect which is a combination of the two reforms. For municipalities of Groups 2 and 3, we are measuring the effects of the reforms being implemented progressively[45]. Hence, we consider an alternative version of Equation 3.1, so that municipality k (with population density X_k) in Group g with school j is included

[44] In Madrid, the choice catchment areas coincide with the city district. However, in other municipalities, the catchment areas were defined through other criteria.

[45] This means that for Group 2 and 3, for 2012 we measure the first changes in criteria in 2012, whereas for 2013 we measure the differential effect of the inter-district choice reform with respect to the new other criteria previously introduced.

in the category $Group_{kgj}$. Our specification is as follows:

$$D_{ijt} = \alpha + \beta * Year2012_{ij} + \delta * Year2013_{ij} + \pi_g \sum_g Group_{kgj} + \theta_g \sum_g Group_{kgj} * Year2012_{ij} + \kappa_g \sum_g Group_{kgj} * Year2013_{ij} + \gamma X_k + \varepsilon_{ijt} \tag{3.11}$$

Table 3.12 displays the results of this regression. The baseline effect for 2012 is significant and the size is a 13% increase in distance for municipalities in Group 1 (*Alcobendas* and *Valdemoro*), but it goes down (by 7.1%) to 4.3% for those municipalities in Group 2, and more importantly, down to non-existing effect for the case of Madrid. Conversely, the baseline effect for 2013 is on average a 11.6% increase in distance, with no differences among the three groups of municipalities. As a robustness check, Column 2 considers a joint regression where Madrid is included with the rest of municipalities from Group 2: as can be seen, for municipalities of Group 1 there was an increase of 13% in the distance traveled that same year, in 2012. This change cannot be extended to the rest of municipalities in Group 2 and 3, as the interaction term of these with the year 2012 is negative and of similar size (although this is probably due to the larger number of observations in Madrid, Group 3). In 2013, however, the average commuting distance effect slightly increases up to 16.4%, and it holds for both type municipalities, which means that those of Group 2 and 3 experienced an overall change similar to that experienced by the municipalities one year earlier.

TABLE 3.11: Geographic Mobility (Distance traveled) for Madrid and its Control Groups.

A. Group 2 and 3 separate		B. Groups 2 and 3 together	
Constant	6.899***	Constant	6.798***
	(0.028)		(0.028)
Year 2012	0.130***	Year 2012	0.130***
	(0.038)		(0.039)
Year 2013	0.166***	Year 2013	0.164***
	(0.037)		(0.037)
Group 2	-0.103***	Group 2 and 3	-0.209***
	(0.031)		(0.030)
Group 3	0.144***	2012*Groups 2 and 3	-0.096**
	(0.034)		(0.040)
2012*Group 2	-0.071*	2013*Groups 2 and 3	-0.033
	(0.041)		(0.038)
2012*Group 3	-0.115***		
	(0.040)		
2013*Group 2	-0.029		
	(0.040)		
2013*Group 3	-0.039		
	(0.039)		
Number of Obs.	92,903	Number of Obs.	92,903

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Each panel represents regressions from equation (11) with two specifications. The dependent variable used is the log-distance traveled. We control for the population density in the student municipality. Clustering errors at the district or neighborhood levels does not alter the significance level of the results.

Overall, we observe that municipalities of Group 1 register a similar shift to a greater distance commuted in the same year as they implemented the inter-district reform, with parents in those municipalities having no time to anticipate the implementation of the reform. This would potentially rule out the hypothesis, at least partially, that changes in distance commuted by parents (and outer district applications) have to do with parents moving to cheaper areas in their municipality. It seems that the reform genuinely shifted the interest of households to more mobility for their siblings' schools.

Finally, the results show that the change was already significant in 2012 for municipalities in Group 2 and accounted for one third of the total change experience one year later. This could be related to the fact that other priority criteria (changes in low-income and former students criteria) may have driven those changes: it would be partly consistent with what is observed in Table 3.5, where part of the change for the city of Madrid was already taking place in 2012. We cannot directly test whether this is related to these changes in 2012. Nevertheless, similarly to what is done for the case of the city of Madrid in Table 3.19, we find that Group 2 municipalities that reformed the criteria in two consecutive years experimented an increase in distance traveled in 2013 with respect to 2012 is independent of the priority criteria changes introduced in 2012[46].

3.8.3 Preferences, Assignment and Strategic Behavior

Section 3.5 discusses how the Boston Mechanism of assignment drives parents' incentives to be conservative and list their first choice for schools where they have greater chances. This is the reason why a large proportion of students (around 85%) end up being assigned to their first option, as was shown in Table 3.4. Nevertheless, as a robustness check, we consider an alternative specification to estimate the results of the reform by using the school final assignment in the outcomes of interest.

We first look at the strategic behavior of parents (and schools) by looking at the mobility outcomes in terms of final school assignment vis a vis the first choice analyzed in Table 3.5. Results are displayed in Table 3.21 of the Appendix. First, it is to be noted that the baseline effect in 2013 in terms of mobility is slightly lower in the final school assignment compared to the first choice option: for example, we find 2.5% more students attending a school in a different district in 2013 as opposed to 3.3% in the first choice option. This may imply that the reform generated a true interest to attend a farther school which outbalanced the strategic component of the Boston mechanism, so that the final assignment mechanisms lowered the final mobility response. Second, when looking at heterogeneous responses in mobility, compared to the first choice option, the school final assignment exacerbates the differences according to social and immigrant characteristics. The reaction to the reform by parental education in the terms of outer-district school assignment is exclusively concentrated among parents in the top education groups, whereas no effect is found for the rest of education groups. Regarding distance traveled, the effect is positive for all households, but the differential increase observed for most educated parents is larger compared to the first choice option specification seen in Table 3.5. Regarding immigrant status, the results are similar for native students, but with respect to immigrant students, the reaction to the 2013 effect is completely canceled. This increases the gap between native and immigrants in the final school assignment compared to the first choice option.

[46]Results can be provided by the authors upon request.

With respect to the impact on social and immigrant segregation, we follow a similar approach as in Section 3.6: we find an increase in social heterogeneity (segregation) of 3.9 percentage points, a slightly lower number than the ones reported in Table 3.6, probably related to the larger mobility of highly educated parents seen in Table 3.21[47]. Finally, with respect to immigrant segregation, we present the results in Figure 3.6 in the Appendix and the results are very similar to those found in Figure 3.3, but the segregation already starts in 2012 and remains in 2013[48].

3.8.4 Unobserved heterogeneity in Census Blocks

Results from Table 3.5 describe that high-educated parents react more to the reform in terms of mobility. Conversely, estimates of social segregation in Table 3.6 proxy a mild reduction of student segregation by parental education, and, similarly, the overall evolution of segregation seen in Table 3.7 shows that social segregation (net of residential dynamics and school fixed effects) did not increase during the time of the reform in Madrid. The contradiction of these two findings is puzzling, especially when the results with respect to immigrant segregation (that we can identify at the individual level) present higher consistency between the mobility and segregation results.

A challenge of this paper is to reconcile the findings presented. The previous literature has documented an increase in social segregation which goes hand in hand with immigrant segregation (Böhlmark et al. (2016)). If, for the case of Madrid, the mobility response is positively related to parental education of students, one would expect an increase in social segregation too. Even more so when there is a high correlation between immigrant status and student socioeconomic characteristics[49].

A potential explanation of the puzzle could be the unobserved heterogeneity caused by the measurement error of parental education. While we are able to measure student immigrant status with high accuracy, the data from parental background is assumed

[47] Additional results on this can be provided by the authors upon request.

[48] Due to computation issues, the model computed for immigrant segregation only controls for the share of immigrants at the neighborhood but does not consider school effects.

[49] In the city of Madrid, immigration status and social background are negatively correlated: in 2013, the correlation of immigrant status and parental years of education was -0.21. Moreover, only 5.9% of immigrant students were part of the top parental education quintile group, while almost 60% of immigrant students had parents with education levels among the bottom two education quintiles.

from census blocks covering a population of about 2,000 inhabitants. Our measure incurs in an error which reverses to the mean tails in each block. For example, in low and middle-income blocks, highly educated parents are identified with much lower education levels (and not identified through education quintiles, as their observed educational level is confounded with those living in their same census block). A similar phenomenon occurs for low-educated households in highly educated blocks. Moreover, it is reasonable to assume that the higher reaction to the reform of high-educated parents can be extrapolated to parents within census blocks: hence, the most regular bias in our measure is caused by the highest educated parents in blocks with low and middle levels of education. This would mean that our data of parental education at the census block incurs in a measurement error that is correlated with the effect of the inter-district school choice reform. If that is the case, our estimates of changes in student mobility by parental education is suffering from a negative bias, so that the additional real gaps in mobility may be even larger after the reform[50]. Conversely, our measure of social heterogeneity (through parental education) suffers a positive bias, especially in low-educated blocks, so that the true gap between parents that react to the reform and their preferred school is not that large. This leads to an under-estimation of the gaps in mobility and an over-estimation (under-estimation) in the change of social diversity (social segregation) after the reform. This bias more the qualitative response in the estimation of social segregation, as the empirical approach we follow builds student heterogeneity measures with the average years of education of the census block[51]. Hence, although the change in social segregation (parental education) derived from the reform may be higher as the one estimate, this remains unknown and stays for future research.

3.9 Conclusion

In this paper we use novel administrative data to analyze the impact of a large-scale inter-district school choice reform in the region of Madrid on household mobility and school segregation. We find that the reform increased the level of geographic mobility

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

[51] With respect to mobility, the fact that we use education groups is enough to show the qualitative response.

of students, with a larger response for higher educated families and families of non-immigrant children. We show that this reform mildly increased the average level of social diversity, but with heterogeneous results: while schools in the bottom of the average performance distribution experienced a decrease (increase) in diversity (segregation), schools in the 3^{rd}, 4^{th} and 5^{th} quintile of the distribution showed a decreased in the level of sorting. Moreover, we find solid evidence of a large increase of immigrant segregation with the implementation of this policy. It is also noticeable that most of the changes are driven by semi-public schools and schools with low and high scores. The puzzling results in terms of segregation (small decrease of social segregation while a large increase in immigrant segregation) will require further research.

It is worth mentioning that the levels of segregation prior to the reform matter. In 2015, the region of Madrid reported very high levels of social segregation in secondary schools in comparison with other Spanish and European regions. Although our paper addresses the mobility and segregation dynamics of new entrants in primary schools, and hence not being completely comparable, the small changes in terms of social diversity (and segregation) may be explained by the fact that the original level of sorting was potentially near to a maximum. In contrast, the opposite effects may be expected in terms of immigrant segregation: since the region of Madrid showed relatively low levels of immigrant segregation, the large increase on the level of segregation may be partially explained by their potential to increase towards higher levels. Our paper provides empirical evidence of the existence of a trade-off between choice and equity in the access to schools. Regarding the future policy agenda, the results raise worrying concerns of an unbalanced equilibrium of the two principles behind (access to quality education and freedom of choice). Further research will need to identify at what geographical levels of choice this trade-off becomes stronger, and, given a level of choice, what are the more efficient education policies that contribute to mitigate the heterogeneous impact of choice by social and immigrant characteristics.

3.10 Appendix

3.10.1 Description of Assignment Criteria

TABLE 3.12: Tie-break Criteria in the Region of Madrid.

	Ties are broken in favor of the student who got higher points on		
	Before 2012/2013	**2012/2013**	**2013/2014**
1	Siblings	Siblings	Siblings
2	Proximity	Proximity	Proximity
3	Disability	Disability	Disability
4	Large family	Former student	Former student
5	Low annual income per capita	Large family	Large family
6	Random lottery	Low-income	Low-income
7		Random lottery	School discretionary
8			Random lottery

Notes: IPREM is the acronym in Spanish for the Multiple Effects Income Public Index, which was 7,455.14 euros in the period of study. The Minimum Insertion Subsidy (*Renta Mínima de Inserción*) is a special subsidy granted to people with lower income than IPREM. School discretionary is a point that the schools have freedom to assign by "public and objective" criteria.

TABLE 3.13: Municipalities with the Single-zone School Choice Setting.

Academic year	**2011/2012**	**2012/2013**	**2013/2014**
Municipalities with single-zone school choice	142 small size municipalities	Alcobendas Algete Colmenar Viejo Tres Cantos Aranjuez Arroyomolinos Brunete Humanes Navalcarnero Pinto San Martín de Valdeiglesias Valdemoro Villaviciosa de Odn Arganda Daganzo Mejorada del Campo San Fernando de Henares San Martín de la Vega Villalbilla Las Rozas de Madrid Moralzarzal Torrelodones	Madrid San Sebastin de los Reyes Alcorcón Fuenlabrada Getafe Leganés Móstoles Parla Alcalá de Henares Coslada Rivas-Vaciamadrid Torrejón de Ardoz Boadilla del Monte Collado-Villalba Galapagar
# Municipalities with inter-district school choice	142	164	179
#Municipalities in the region of Madrid	179	179	179

3.10.2 Section Example

Figure 3.5 shows an example of the layout of the sections in the Madrid districts of "Centro" and "Retiro". The larger the sections the less is the proportion of residential area there:

FIGURE 3.5: Census Blocks in the City of Madrid.

3.10.3 Years of Schooling

Census education categories are the followings:

1. Cannot read, cannot write
2. No studies
3. Incomplete Primary
4. Middle school, Primary or Compulsory Secondary Education
5. Vocational Training (Elemental)
6. Vocational Training (Advanced)
7. High school
8. Other Intermediate Graduates
9. University School Graduates
10. Technical Engineer
11. College Graduated
12. No-university Graduated
13. PhD and other Post-graduates

We group all these 13 categories into 6 new categories. Categories 1 and 2 are grouped as No studies. Category 3 remains as Incomplete Primary Education. Categories 4 and 5 form the Lower Secondary Education. Categories 6, 7 and 8 are gathered together in Upper Secondary Education. Categories 9 and 10 form the Lower Tertiary Education. Groups 11, 12 and 13 are joined in category Post-Graduate education. In order to assign an equivalent amount of years of schooling, we make the assumption that parents were educated under the legal framework of the Education Act LGE (*Ley General de Educación*), which was in place for pupils born before 1985. We do this as the average maternity age for the first offspring was 29.5 years old in 2007 (therefore those students aged 3 in 2010) and 30.5 for 2013 (therefore those students aged 3 in 2016). This means that with a large probability, mothers of students born in 2007 and 2013 were at school as part of the LGE framework. The LGE framework consisted of 8 years of basic primary schooling, with 5 years of primary school and 3 years of lower secondary school. We assume category *No Studies* as only 3 years of primary and *Incomplete primary* as just 5 years of schooling. After that Basic schooling, there was vocational training with degrees of 2 to 4 years. We assume Categories 4,5 and 6 to be in between Basic schooling (8 years and some vocational training), averaging 9 years of schooling. Categories 7 and 8 pertain to High School and Other intermediate graduates, which corresponded to 12 years of schooling. Finally, university graduates and technical engineers are assumed to do 18 years of schooling, and post-graduate studies are given an average of 3 additional years of education. The equivalence of years of schooling for a specific census block is given by the following formula, where each percentage of census population is multiplied by the equivalent years of schooling.

$$\begin{aligned} YS_s =& Non-studies_s * 3 + Primary_s * 5 + LowSecondary_s * 9 + \\ & UpperSecondary_s * 12 + LowTertiary_s * 17 + College_s * 20 \end{aligned} \tag{3.12}$$

3.10.4 School Classification

TABLE 3.14: Districts ranked by income.

Low Income District	**Middle-Income District**	**High-Income District**
Puente de Vallecas	San Blas - Canillejas	Centro
Villaverde	Moratalaz	Barajas
Usera	Ciudad Lineal	Moncloa - Aravaca
Carabanchel	Tetan	Retiro
Latina	Fuencarral - El Pardo	Chamberí
Vicálvaro	Hortaleza	Chamartín
Villa de Vallecas	Arganzuela	Salamanca

Notes: Data from district average income is derived from Municipal census data in the city of Madrid.

3.10.5 Tables and Figures

TABLE 3.15: Geographic Mobility, Social Diversity and Immigrant Segregation by school grade quintile.

	Geographic Mobility (1)	**Social Diversity** (2)	**Immigrant Segregation** (3)
Constant	0.006	-0.707***	0.019
	(0.006)	(0.022)	(0.015)
Year 2012	0.024***	0.014	0.042***
	(0.008)	(0.031)	(0.016)
Year 2013	0.029***	-0.061*	0.029*
	(0.009)	(0.032)	(0.016)
2nd Quintile Sch. Grade (t-1)	0.047***	0.178***	0.007
	(0.008)	(0.031)	(0.017)
3rd Quintile Sch. Grade (t-1)	0.075***	0.215***	0.005
	(0.008)	(0.029)	(0.018)
4th Quintile Sch. Grade (t-1)	0.076***	0.302***	0.015
	(0.008)	(0.029)	(0.018)
5th Quintile Sch. Grade (t-1)	0.093***	0.282***	0.012
	(0.007)	(0.028)	(0.019)
Year 2012*2nd Quintile	-0.012	-0.122***	-0.015
	(0.011)	(0.043)	(0.023)
Year 2012*3rd Quintile	-0.034***	0.101**	-0.033
	(0.011)	(0.041)	(0.022)
Year 2012*4th Quintile	0.010	-0.093**	-0.040*
	(0.011)	(0.041)	(0.022)
Year 2012*5th Quintile	-0.019*	0.074*	-0.070***
	(0.011)	(0.040)	(0.021)
Year 2013*2nd Quintile	-0.012	0.042	-0.002
	(0.012)	(0.044)	(0.023)
Year 2013*3rd Quintile	0.002	0.123***	-0.030
	(0.011)	(0.043)	(0.023)
Year 2013*4th Quintile	0.029**	0.158***	-0.014
	(0.011)	(0.042)	(0.022)
Year 2013*5th Quintile	-0.004	0.175***	-0.048**
	(0.011)	(0.041)	(0.022)
Number of Obs.	53,220	53,017	1,492

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively.

TABLE 3.16: Mechanisms for Immigrant Segregation.

School Grade	(1)	Years of Education	(2)	Prop. of Immigrants	(3)
Constant	0.027**	Constant	-0.114	Constant	0.051***
	(0.011)		(0.110)		(0.014)
Year 2012	0.010	Year 2012	0.029	Year 2012	0.011
	(0.007)		(0.035)		(0.009)
Year 2013	0.010	Year 2013	0.032	Year 2013	0.021**
	(0.007)		(0.036)		(0.010)
Sch. Grade (t-1)	0.005	Avg.Years Education (t-1)	0.013	Avg. Frac. Inmig. (t-1)	-0.167***
	(0.006)		(0.010)		(0.042)
2012*Sch. Grade (t-1)	-0.022***	2012*Avg.Years Education (t-1)	-0.002	2012*Avg. Frac. Inmig. (t-1)	-0.012
	(0.007)		(0.003)		(0.044)
2013*Sch. Grade (t-1)	-0.012*	2013*Avg.Years Education (t-1)	-0.002	2013*Avg. Frac. Inmig. (t-1)	-0.016
	(0.007)		(0.003)		(0.042)
Number of Obs.	1,492	Number of Obs.	1,478	Number of Obs.	1,478

Notes:*, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. We run similar models as those of Equation 3.6, with interaction terms of school characteristics for the mean immigrant share. Because of this reason, we cannot interact school characteristics that are constant across years (school ownership or school district).

TABLE 3.17: Mechanisms: Segregation Index of Dissimilarity

Dissimilarity Index	**Social Segregation**			**Immigrant Segregation**		
	2011	2012	2013	2011	2012	2013
A. School Grade						
1st School Quantile Sch. Grade	0,435	0,443	0,455	0,331	0,368	0,361
	(0,011)	(0,010)	(0,007)	(0,012)	(0,015)	(0,018)
2nd School Quantile Sch. Grade	0,465	0,505	0,475	0,357	0,411	0,392
	(0,005)	(0,005)	(0,004)	(0,013)	(0,016)	(0,017)
3rd School Quantile Sch. Grade	0,454	0,462	0,468	0,349	0,392	0,447*
	(0,009)	(0,007)	(0,005)	(0,020)	(0,022)	(0,020)
4th School Quantile Sch. Grade	0,455	0,489	0,455	0,349	0,435	0,449*
	(0,005)	(0,006)	(0,006)	(0,015)	(0,014)	(0,027)
4th School Quantile Sch. Grade	0,432	0,442	0,423	0,452	0,456	0,427
	(0,007)	(0,003)	(0,006)	(0,019)	(0,023)	(0,019)
B. Income District						
Low Income District	0,361	0,364	0,358	0,373	0,382	0,410*
	(0,004)	(0,006)	(0,004)	(0,012)	(0,011)	(0,013)
Medium Income District	0,409	0,421	0,403	0,427	0,454	0,446
	(0,005)	(0,004)	(0,006)	(0,022)	(0,017)	(0,011)
High Income District	0,278	0,297	0,292	0,474	0,493	0,471
	(0,006)	(0,008)	(0,005)	(0,016)	(0,014)	(0,018)
C. Public vs Semi-Public Schools						
Public School	0,481	0,485	0,482	0,372	0,386	0,401*
	(0,004)	(0,005)	(0,004)	(0,010)	(0,007)	(0,007)
Semi-Public School	0,499	0,504	0,481	0,433	0,502	0,508*
	(0,004)	(0,002)	(0,004)	(0,009)	(0,011)	(0,013)
D. Bilingual vs Non-Bilingual						
Bilingual School	0,473	0,484	0,460	0,413	0,451	0,466
	(0,005)	(0,004)	(0,004)	(0,015)	(0,014	(0,012)
Non-Bilingual School	0,486	0,487	0,482	0,423	0,426	0,443*
	(0,003)	(0,003)	(0,005)	(0,009)	(0,009)	(0,009)

Notes: Standard Errors (in parentheses) are computed by bootstrapping individual data. We perform 20 bootstrap replications for each estimate. * indicate significance change at the at 5 percent level between 2011 and 2013.

TABLE 3.18: Mechanisms: Interactions with Semi-Public School.

School Grade	(1)	Years of Education	(2)	Prop. of Immigrants	(3)
Constant	0.046***	Constant	-0.074***	Constant	0.038***
	(0.004)		(0.019)		(0.006)
Year 2012	-0.005	Year 2012	0.013	Year 2012	-0.004
	(0.005)		(0.027)		(0.008)
Year 2013	0.016***	Year 2013	-0.048*	Year 2013	0.052***
	(0.005)		(0.027)		(0.008)
Avg. Grade	0.017***	Avg.Years Education (t-1)	0.011***	Avg. Frac. Inmig. (t-1)	-0.043*
	(0.003)		(0.002)		(0.026)
Semi-Public	0.035***	Semi-Public	-0.199***	Semi-Public	0.080***
	(0.005)		(0.026)		(0.007)
Sch. Grade (t-1)*Semi-Public	0.022***	Avg.Years Education (t-1)*Semi-Public	0.021***	Avg. Frac. Inmig. (t-1)*Semi-Public	-0.319***
	(0.006)		(0.002)		(0.043)
2012*Sch. Grade (t-1)	0.005	2012*Avg.Years Education (t-1)	-0.002	2012*Avg. Frac. Inmig. (t-1)	0.001
	(0.005)		(0.002)		(0.036)
2013*Sch. Grade (t-1)	0.005	2013*Avg.Years Education (t-1)	0.006**	2013*Avg. Frac. Inmig. (t-1)	-0.149***
	(0.005)		(0.002)		(0.034)
2012*Semi-Public	0.043***	2012*Semi-Public	0.001	2012*Semi-Public	0.027***
	(0.007)		(0.037)		(0.010)
2013*Semi-Public	0.039***	2013*Semi-Public	0.078**	2013*Semi-Public	0.003
	(0.007)		(0.037)		(0.010)
2012*Semi-Public*Sch. Grade (t-1)	-0.016**	2012*Semi-Public*Avg.Years Education (t-1)	0.002	2012*Semi-Public*Avg. Frac. Inmig. (t-1)	-0.005
	(0.008)		(0.003)		(0.062)
2013*Semi-Public*Sch. Grade (t-1)	-0.013	2013*Semi-Public*Avg.Years Education (t-1)	-0.005	2013*Semi-Public*Avg. Frac. Inmig. (t-1)	0.144***
	(0.008)		(0.003)		(0.055)
Number of Obs.	53,22	Number of Obs.	52,468	Number of Obs.	52,468

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Each panel represents regressions from Equation 3.11, interacting semi-public schools with other school characteristics. We control for the population density in the student neighborhood.

TABLE 3.19: Geographic Mobility, Social and Immigrant Segregation. Bilingual vs. non-bilingual schools

	A. Outer District Applications (Fraction)	B. Travel Distance (Meters)
Constant	0.064***	6.729***
	(0.004)	(0.013)
Year 2012	0.012***	-0.020
	(0.004)	(0.015)
Year 2013	0.028***	0.080***
	(0.005)	(0.015)
Bilingual School	0.006	0.132***
	(0.005)	(0.016)
2012*Bilingual School	-0.002	0.036
	(0.007)	(0.023)
2013*Bilingual School	0.007	0.041*
	(0.007)	(0.023)
Number of Obs.	53,017	53,017
	C. Social Segregation	**D. Immigrant Segregation**
Constant	-0.581***	0.018
	(0.011)	(0.013)
Year 2012	-0.010	0.014*
	(0.017)	(0.009)
Year 2013	0.016	0.014
	(0.017)	(0.008)
Bilingual School	0.211***	0.022
	(0.018)	(0.023)
2012*Bilingual School	-0.005	-0.018
	(0.025)	(0.014)
2013*Bilingual School	0.026	-0.016
	(0.026)	(0.014)
Number of Obs.	53,004	1,483

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively.

TABLE 3.20: Relationship between low-income and former student bonus in the effects of the 2013 reform on mobility for the city of Madrid.

	A. Outer District Application (Fraction)		B. Travel Distance (Meters)	
	Low-Income Bonus	Former Student Bonus	Low-Income Bonus	Former Student Bonus
	(1)	(2)	(3)	(4)
Constant	0.080***	0.083***	6.808***	6.814***
	(0.004)	(0.004)	(0.013)	(0.013)
Year 2013	0.021***	0.020***	0.108***	0.109***
	(0.004)	(0.004)	(0.012)	(0.012)
Bonus Points	-0.033**	-0.058***	-0.070	-0.139***
	(0.013)	(0.011)	(0.044)	(0.037)
2013*Bonus	0.005	0.018	-0.006	-0.030
	(0.019)	(0.016)	(0.063)	(0.054)
Number of Obs.	34,949	34,949	34,808	34,808

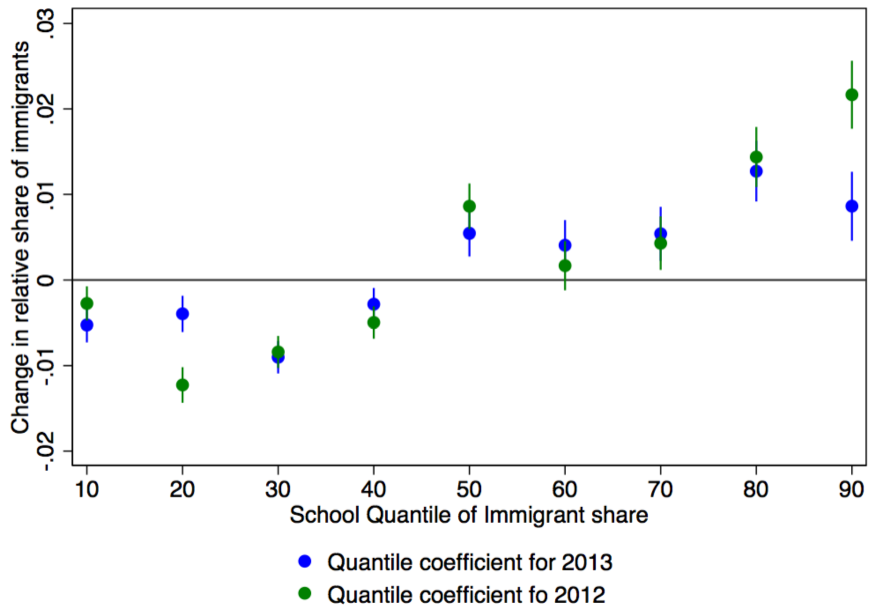
Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. We control for the population density in the student neighborhood.

TABLE 3.21: Geographic Mobility Coefficient Effects (Final School Assignment)

	Dependent variable					
	Outer district applications (Fraction)			Travel Distance (Log Meters)		
	2011	2012	2013	2011	2012	2013
		(1)			(2)	
A. Baseline Estimates						
Baseline Effect (vs 2011)	-	0.0124***	0.0247***	-	0.0078	0.0939***
	-	(0.035)	(0.0035)	-	(0.012)	(0.012)
B. Heterogeneous Effects: Parental Education						
		(3)			(4)	
Baseline Effect (vs 2011)	-	0.0042	0.0099	-	-0.0252	0.060***
	-	(0.0075)	(0.0075)	-	(0.025)	(0.0257)
2^{nd} quintile EDU	0.0149**	-0.0005	0.0035	0.111***	0.0597	0.0246
	(0.0076)	(0.0107)	(0.0109)	(0.025)	(0.039)	(0.0363)
3^{rd} quintile EDU	0.069***	0.0049	0.0101	0.196***	0.0170	-0.0101
	(0.0075)	(0.0106)	(0.0109)	(0.025)	(0.0354)	(0.0361)
4^{th} quintile EDU	0.073***	0.0042	0.0234**	0.287***	0.0121	0.0253
	(0.0075)	(0.0107)	(0.0109)	(0.025)	(0.0357)	(0.0363)
5^{th} quintile EDU	0.094***	0.031***	0.0324***	0.319***	0.065***	0.117***
	(0.008)	(0.011)	(0.011)	(0.026)	(0.037)	(0.0379)
C. Heterogeneous Effects: Immigrant Status						
		(5)			(6)	
Baseline Effect (vs 2011)	-	0.0168***	0.0305***	-	0.0152	0.106***
	-	(0.038)	(0.0038)	-	(0.0125)	(0.0127)
Immigrant student	-0.041***	-0.0175*	-0.0296***	-0.179***	-0.0056	-0.0452
	(0.0073)	(0.0097)	(0.10)	(0.024)	(0.033)	(0.033)
Number of Obs.		51,126			51,126	

Notes: Standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Each panel represents regressions from Equation (1), (2) and (3) respectively. In Equations (2) and (3) heterogeneous effects need to be interpreted as an addition to the baseline effect. We control for the population density in the student neighborhood. Clustering errors at the district or neighborhood levels does not alter the significance level of the results.

FIGURE 3.6: Post-Reform effects on share of immigrant students for final school assignment, by quantiles.



Chapter 4

Job Tasks and Wages across the world: Evidence from PIAAC

4.1 Introduction

The technological change phenomenon is leading to an automation of a range of low and medium-skill occupations, in particular those whose contents follow precise and well-understood procedures or routines. Workplace computerization leads to an automation of these routines and hence to a gradual change of contents/tasks demanded at the workplace. Such computerization is driven by a decline in the relative price of computer capital (identified with technological progress), which increases computer and technology adoption at firms and the workplace, hence altering the allocation of labor across different task inputs. Specifically, computer capital and labor are relative complements in carrying out non-routine/non-codifiable tasks, while computer capital and labor are substitutes in performing routine (codifiable) tasks. Indeed, Autor et al. (2003) document a change in the occupational distribution of employment employment polarization and assesses that computerization is indeed a key driver to understand such changes.

Understanding employer behavior with respect to their labor demands in the face of computerization requires a theoretical framework which specifically links skill requirements with technological change. The traditional Mincerian human capital model was very successful in explaining the relationship between wages and investment in human capital but given its supply based nature, it not helpful in understanding skill demand requirements. Autor et al. (2003)(ALM) and, more recently, Acemoglu and Autor (2011) developed a new theoretical task framework rooted in a fundamental difference between skills and tasks. In this framework, the basic units of production are job content/tasks, which are supplied by production factors (labor and capital) which combined can produce units of output. Assignment of factors to tasks is determined in equilibrium by comparative advantage and hence each task is done by the least-cost factor.

This task framework, however, still faces important challenges and measurement is undoubtedly one of the most important ones. The first empirical studies used detailed occupation data to approximate job tasks. However, tasks may share similarities across very different occupations and hence the occupation data approach does not allow such similarities to be identified in an objective way. A subsequent set of empirical studies used two datasets from surveys which offered job descriptors (contents) to each occupation: the Dictionary of Occupational Titles (DOT) and its successor, the Occupational Information Network (O*NET). Autor and Dorn (2013), Acemoglu and Autor (2011), Goos et al. (2014), among others, have recently used these datasets for the measurement of task contents. However, this has two major drawbacks even though this approach assigns job descriptors to occupations objectively (and hence can be validated by the agencies which supply them).

The first one is that these job descriptors DOT and O*NET are derived from US surveys. Whereas the occupational based job descriptors may be similar to the US ones for some OECD countries, such approximation for other countries may lead to large errors in job contents. The second limitation is that this approach overlooks job tasks heterogeneity within occupations. In fact, empirical research has found important heterogeneity in job content at the worker level within detailed occupations (see Spitz Oener (2006) and Autor and Handel (2013)). In particular, Spitz-Oener argues for the case of Germany that job content changes take place mostly within, rather than between occupations[1]. Hence, measuring job tasks at the workplace remains critical for an accurate understanding of tasks, relating them to occupations, human capital and other demographics, as well as for identifying its link with wages on the verge of a technological

[1]Spitz documents the case of Germany in the period between 1979 and 1999 and divides job contents in five categories: non-routine analytic, non-routine interactive routine cognitive, routine manual, and non-routine manual. Results from a shift-share analysis show that task changes within occupations account for 85%, 87%, 99%, 86% and 98% respectively of the total change in tasks between 1979 and 1999.

transformation.

The empirical studies that measure tasks at the worker level are, to date, single-country analysis (Spitz Oener (2006) and Autor and Handel (2013)). Our paper extends the empirical assessment of task measurement at a worker level to a broad sample of countries. Specifically, we compute differences in the tasks content of jobs across a harmonized and hence comparable sample of countries that includes European, Asian and American countries. There is evidence that the process of de-routinization has not followed identical paths across countries. Hardy et al. (2016) document an increase rather than a decrease in routine cognitive employment in the transition economies of Eastern and Central Europe. Gimpelson and Kapeliushnikov (2016) and Aedo et al. (2013) found similar results for Russia and Southern European countries, respectively. Hence, there is a need for an assessment of job tasks around the world from a comparable sample of countries at the worker level.

We therefore use the Programme for the International Assessment of Adult Competencies (PIAAC), a survey which provides harmonized information across countries. We use very precise information on job contents at the worker level, which allows for job task heterogeneity within occupations to be considered and hence the nuanced interplay between job demand, tasks, occupational sorting and earnings to be understood. At the same time, given the availability of worker level information on the intensity of computer use at work, we assess the empirical relationship between computerization at work and the task content of workers. This is indeed the first of the three contributions of our paper to the literature, i.e., to depict comparable cross-country differentials in the ALM task framework for 20 OECD and non-OECD countries.

The second contribution of this paper is to study the relationship between job contents and other demographics and job characteristics, in particular, the role played by worker sorting into occupations in determining job tasks, as well as the association between computerization at work and the type of tasks that workers perform. Such analysis is conducted by very detailed and unique measures of worker cognitive skills, which help to disentangle such relationship from potential confounding worker effects. Finally, our third contribution is to assess the predictive power of job tasks on worker wages: we do so under the task approach, which assumes that workers self-select into jobs according to their comparative advantage.

Our findings show that New Zealand and Great Britain display the largest values in non-routine cognitive and personal tasks. On the other side, routine (cognitive and manual) ones are found to be highest in Lithuania and Turkey. Second, computer use at work is revealed as a key factor in explaining differences in job contents, both between and within occupations. Third, job tasks exert a high predictive power to explain wage differences, even within occupations. Additionally, results are consistent with the assumption by the task approach that workers self-select into occupations based on their comparative advantage. Finally, our individual-based task measures are validated by comparing them with those which would be obtained when assigning job descriptors from O*NET dataset (which have been previously used in the literature), bringing further nuances that can add more value in future research.

The rest of the paper is organized as follows. Section 2 discusses the data sources and a discussion on data sources and the construction of task/job contents. Section 3 and 4 present empirical estimates of task determinants and the relation between tasks and wages. Section 5 concludes.

4.2 Data Sources, Task Measures and Descriptives across Countries

4.2.1 Data

Our empirical approach uses data from the Programme for the International Assessment of Adult Competencies (PIAAC), conducted by the Organization for Economic Co-operation and Development (OECD) in two waves first, in 2011-2012 and second, in 2014-2015 -in 31 participating countries: Austria, Belgium, Canada, Chile, Cyprus, Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France, Great Britain, Greece, Ireland, Israel, Italy, Japan, South Korea, Lithuania, Netherlands, Norway, New Zealand, Poland, Russian Federation, Singapore, Slovak Republic, Slovenia, Sweden, Turkey and the United States. The data sample contains 201,033 observations, which represent a total population of 815 million adults aged 16 to 65. The survey includes a personal interview comprising a questionnaire followed by a skills assessment of literacy, numeracy and problem-solving skills in technology environments.

The questionnaire contains information about personal background, education and training, current work status, work history and skills used at current job (or last job) and everyday life. As said previously, the variables of job contents (tasks) used at work are particularly appropriate for the analysis within occupations. In addition, the PIAAC skills assessment provides an accurate measurement of cognitive skills, an excellent proxy to control for unobserved heterogeneity. Beyond the assessment of specific reading, mathematical or technology contents, the skill assessment framework of PIAAC emphasizes the ability of workers to apply background knowledge, a unique feature used by OECD in their cognitive skill assessments.

This paper aims at understanding the task content of jobs in a framework where occupational sorting drives worker choices. Unfortunately, not all countries in the sample provide detailed occupation-level information at more than one digit. Therefore, we lose 11 countries from the sample and remain with 20 countries for the analysis: Belgium, Chile, Cyprus, Czech Republic, Denmark, Spain, France, Great Britain, Greece, Italy, Japan, South Korea, Lithuania, Netherlands, Poland, New Zealand, Russian Federation, Slovak Republic, Slovenia and Turkey.

4.2.2 Constructing Worker-level Task Measures

Using data from the worker responses of activities conducted at work, we construct measurements of task intensities. We follow the original Autor et al. (2003) approach that defines five task measurements: non-routine cognitive, non-routine personal, routine cognitive, routine manual and non-routine manual.

We construct a set of indices based on the PIAAC background questionnaire. The background questionnaire displays answers on work habits and tasks performed at the workplace, normally organized through different types of frequency answers. To keep consistency in the aggregation of answers into indexes, we restrict our set of items used to those with the frequency category on responses to the question starting by ”How often does does your job usually involve”. The responses are organized as five qualitative time inputs: (i) Every day; (ii) At least once a week but not every day; (iii) Less than once a week; (iv) Less than once a month; (v) Never.

TABLE 4.1: Task Framework with PIAAC Data

Task	PIAAC Questionnaire Item	Item No.
Non routine Cognitive	Write Reports	G Q02c
Non routine Cognitive	Write Reports	G Q02c
Non routine Cognitive	Read Diagrams, Maps or Schematics	G Q01h
Non routine Cognitive	Faced Complex Problems (¿30 mins)	F Q05b
Non - routine Personal	Planning the activities of others	F Q03b
Non-routine Personal	Persuading/Influencing People	F Q04a
Routine Cognitive	Learning-by-doing from task performed	D Q13b
Routine Cognitive	Organizing your own time	F Q03c
Routine Cognitive	Instructing, training or teaching people	F Q03c
Routine Cognitive	Making speeches or giving presentations	F Q02c
Routine Manual	Hand and finger dexterity	F Q06b

Notes: All questions provide the same time categorical answers: (i) Every Day; (ii) At least once a wee but not every day; (iii) Less than once a week; (iv) Less than once a month; (v) Never.

We follow Autor and Handel (2013) to construct the indexes for each of the dimensions using the first component of a principal component analysis[2]. Due to data limitations on manual tasks, we only construct four of the five job tasks: non-routine cognitive, non-routine personal, routine cognitive and routine manual. For the non-routine cognitive task index, we use three different items: (i) write reports; (ii) read diagrams, maps or schematics; and (iii) face complex problems that take at least 30 minutes. We construct the non-routine personal task index with the following items: (i) planning the activities of others; (ii) persuading or influencing people. The routine cognitive task index is constructed with four different items: (i) learning-by-doing at work; (ii) organizing own time; (iii) instructing, training or teaching people, individually or in groups; (iv) making speeches or giving presentations in front of five or more people. The last two items follow the approach used by Autor and Handel (2013), which construct the routine cognitive index emphasizing the importance of the lack of face-to-face interactions.

The manual routine index is constructed from the item using skill or accuracy with hands or fingers item, which has been widely used in the task literature (see Autor et al. (2003) and Autor and Dorn (2013)). Table 4.1 depicts the job task items from the PIAAC background questionnaire that are used to construct each of the task indexes. For non-routine manual tasks, an alternative would be to use the question on Physical Work.

[2] Autor and Handel (2013) follow a principal component analysis to derive continuous job task variables taking advantage of multiple responses of items. The data from Spitz Oener (2006) only contain binary information on whether the worker either performs a certain task or not, and aggregate measures are constructed as percentage of activities performed for each category of tasks. As a robustness check with our approach, data comparing both methods lead to similar results, with correlation of 0.90 for non-routine cognitive tasks, 0.78 for non-routine personal tasks, and 0.92 for routine cognitive tasks.

Nevertheless, the non-routine manual task constructed in, for example, the Autor et al. (2009) approach is a non-codifiable type of work[3]. These should include, among others, task such as dexterity, coordination, object handing or spatial orientation tasks. Unfortunately, the PIACC dataset does not include items to learn about these non-routine manual job contents and, hence, we decide not to use any non-routine manual task measure in our approach.

Our sample consists of 51,056 observations (from 20 countries for which occupation at more than 1-digit is available) that are well defined for each item of the task framework, as well as the worker covariates later used in the next section. We then compute the indexes for each task. Consistent with Autor and Handel (2013), the first component of the principal component analysis for each task explains a relatively large variance (see Table 4.10 in Section 4.6), being 0.55 for non-routine cognitive tasks, 0.71 for non-routine personal tasks and 0.45 for routine cognitive tasks[4]. We also compute the indexes into their standardized form and compare their results at the worker level. The indexes of non-routine cognitive and personal tasks are positively correlated (0.47), while both are negatively correlated (-0.49 and -0.60) with routine cognitive tasks. Routine manual task index does not correlate with the rest of indices in any relevant way.

4.2.3 Job Task Descriptives around the World

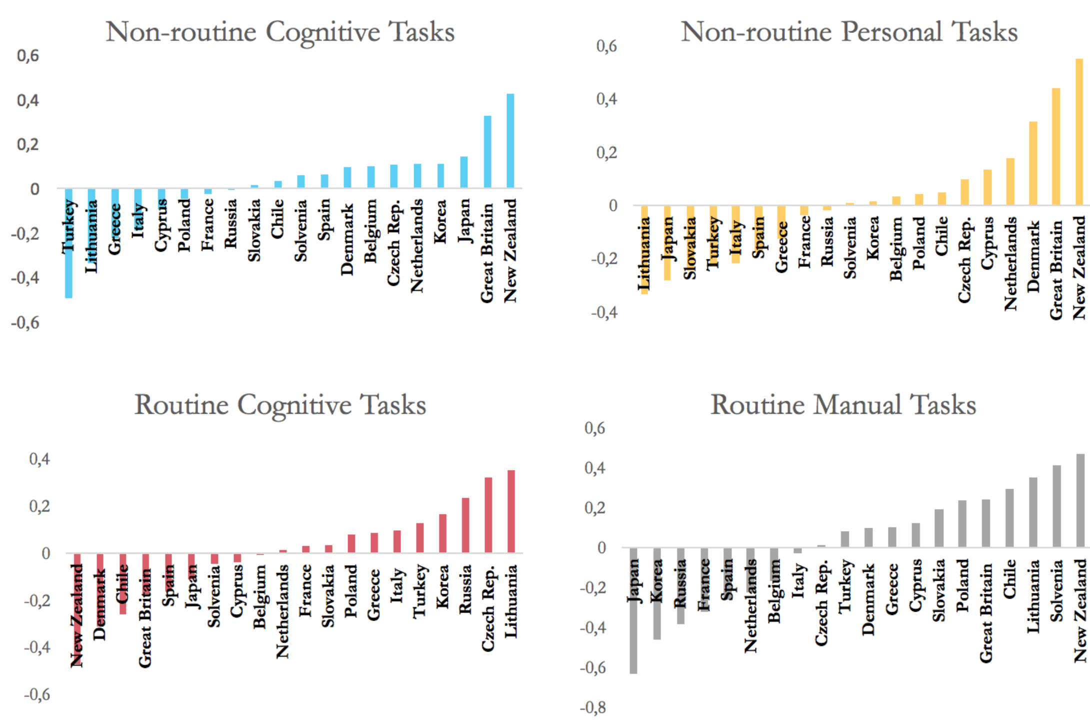
In Figure 4.1 we depict the average values of the standardized form of the task variables explained above. In particular, we find that New Zealand, Great Britain, Denmark, Japan, South Korea and Netherlands display the largest values of non-routine cognitive tasks, while these are lowest in Turkey, Lithuania and Greece. This is consistent with what is found for the case of routine cognitive tasks (where the order is

[3]For the manual non-routine category, both Spitz Oener (2006) and Autor and Handel (2013) use activities that are clearly identifiable as non-routine. Spitz Oener (2006) uses as response of activity: Repairing or renovating houses/apartments/machines/vehicles, restoring art/monuments, and serving or accommodating, while Autor and Handel (2013) use four activities: (i) operating vehicles, mechanized devices, or equipment; (ii) time spent using hands to handle, control, or feel objects, tools, or controls; (iii) manual dexterity; (iv) spatial orientation. From the Dictionary of Occupational Titles (DOT), Autor and Dorn (2013) use eye-hand-foot coordination variable for the manual (non-routine) task and finger dexterity to be included as the manual part of the routine construct. Finally, Acemoglu and Autor (2011) use from the Occupational Information Network (O*NET) pace determined by speed of equipment, controlling machines and processes and spend time making repetitive motions for routine manual tasks and operating vehicles, mechanized devices, or equipment, spend time using hands to handle, control or feel objects, tools or controls, manual dexterity or spatial orientation.

[4]Follwoing Autor and Handel (2013), Abstract and Routine scales are derived from a first component of a principal component analysis which explain 41% and 56% of the total variation respectively.

mostly inversed), with the exception of Netherlands. Moreover, we find large values of non-routine personal tasks in New Zealand, Great Britain, Denmark and Netherlands and low values in Japan, Lithuania and Slovakia. Finally, routine manual tasks are low in Japan, South Korea, France and Russia, whereas high in Lithuania, Slovenia and, surprisingly, New Zealand. The cross-country comparison between the three task indexes shows a very high and positive correlation between non-routine cognitive and personal (0.72), whereas those two display a negative correlation with routine cognitive tasks (-0.60 and -0.59 respectively). The routine manual task correlates positively with non-routine personal tasks (0.32) and negatively with routine cognitive tasks (-0.24).

FIGURE 4.1: Job Task Measures by Countries.



Notes: Results display values of standardized indexes for each task, with mean 0 and standard deviation 1. Observations are weighted so that countries are weighted equally. Results are also presented in Table 4.12 Section 4.6

Secondly, we look at the task content of jobs across a set of individual covariates, which can be found in Table 4.2. The results show a larger value of non-routine cognitive tasks for males, whereas no difference in the other tasks. Moreover, the task-age curve displays a concave shape for non-routine tasks, whereas a convex shape is found for routine tasks. Workers' education and skills are positively related to non-routine (both cognitive and personal) tasks, whereas negatively related to routine tasks. Finally, workers receiving training on the job show larger shares of non-routine tasks and a lower share of routine cognitive tasks, compared to those which do not receive any training on the job. Regarding sectors, services display larger shares of non-routine personal tasks, whereas lower shares of routine tasks, both cognitive and manual.

TABLE 4.2: Distribution of Task Measures by Worker Covariates.

	NRC		NRP		RC		RM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender								
Female	-0.17	0.95	-0.02	0.99	0.03	1.00	0.00	1.01
Male	0.14	1.02	0.02	1.01	-0.02	1.00	0.00	0.99
Age								
25 or less	-0.27	0.94	-0.15	0.94	0.17	0.9	0.08	0.95
25 - 34	0.13	1.00	0.04	0.98	-0.08	0.98	0.03	0.98
35 - 44	0.07	1.00	0.07	1.01	-0.06	1.00	0.00	1.00
45 - 54	-0.03	1.00	-0.02	1.02	0.03	1.02	-0.01	1.01
55 - 64	-0.17	0.97	-0.09	1.01	0.12	1.02	-0.08	1.04
Education Level								
Lower Secondary or Less	-0.56	0.87	-0.44	0.89	0.46	0.86	0.13	0.94
Upper Secondary	-0.15	0.97	-0.13	0.99	0.20	0.92	0.09	0.96
Post-secondary or Tertiary Professional	0.12	0.94	0.13	0.98	-0.20	0.98	0.01	0.98
Tertiary (Bachelor/Master)	0.53	0.88	0.41	0.93	-0.56	0.96	-0.24	1.06
Literacy Skills (Quartile)								
Quartile 1	-0.38	0.95	-0.27	0.97	0.26	0.99	0.16	0.93
Quartile 2	-0.07	0.99	-0.04	1.00	0.06	1.00	0.07	0.97
Quartile 3	0.15	0.97	0.1	0.99	-0.09	0.98	-0.04	1.01
Quartile 4	0.42	0.91	0.29	0.94	-0.32	0.93	-0.25	1.06
Numeracy Skills (Quartile)								
Quartile 1	-0.44	0.93	-0.32	0.95	0.29	0.98	0.16	0.92
Quartile 2	-0.13	0.97	-0.07	1.00	0.09	1.00	0.07	0.97
Quartile 3	0.11	0.97	0.08	1.00	-0.05	0.98	-0.06	1.02
Quartile 4	0.46	0.91	0.3	0.95	-0.33	0.93	-0.18	1.05
On the Job Training								
No	-0.22	0.96	-0.17	0.97	0.2	0.96	-0.01	1.00
Yes	0.36	0.95	0.27	0.98	-0.33	0.97	0.01	0.99
Sector								
Manufacturing	0.06	1.04	-0.16	1.00	0.16	0.95	0.07	0.98
Construction	0.00	1.01	-0.12	1.02	0.15	0.94	0.18	0.90
Services	-0.03	0.98	0.1	0.99	-0.10	1.02	-0.06	1.02

Notes: the sample includes all employed respondents of all countries. Results display values of standardized indexes for each task, with mean 0 and standard deviation 1. Observations are weighted so that countries are weighted equally. NRC refers to non-routine cognitive, NRP refers to non-routine personal, RC refers to routine cognitive and RM refers to routine manual tasks.

4.2.4 Validity of Job Task Measures: Worker level vs Occupational Level (PIAAC and O*NET)

To better understand the reliability of our measures, we compute the worker-level measurement of tasks at the occupational four-digit level and compare them with the worker-level values. Table 4.3 shows a positive large correlation of worker and occupational-level task measurements.

Moreover, we can validate our individual-based task measures by comparing them with those which would be obtained when assigning job descriptors from O*NET to each of the 4-digit occupational level. For comparability reasons, we use O*NET dataset to construct non-routine cognitive, non-routine personal, routine cognitive and routine manual measures using exactly the same task measures used in Acemoglu and Autor (2011) at the 4-digit occupation level. As Table 4.3 shows, we find a positive correlation of around 0.54 for non-routine cognitive tasks, 0.46 for non-routine personal tasks, 0.26 for routine cognitive tasks (which is lower than expected) and 0.30 for routine manual tasks. The fact that the PIAAC occupation-level correlation O*NET is slightly larger compared to the PIAAC worker-level correlation (except for routine cognitive tasks) is consistent with Autor and Handel (2013)[5].

TABLE 4.3: Correlation between Worker and Occupation-Level Task Measures (PIAAC and O*NET)

	Worker-level PIAAC		Occupation-level PIAAC
	Occupation-level PIAAC	Occupation-level O*NET	Occupation-level O*NET
Non-routine cognitive	0.672	0.417	0.536
Non-routine personal	0.644	0.368	0.456
Routine cognitive	0.679	0.235	0.264
Routine manual	0.585	0.197	0.304

Notes: The sample includes employed respondents aged 20-64 and currently working for which variables in Section 4.3 are well defined and have non missing values. Observations are weighted so that countries are weighted equally.

In order to better understand the consistency of our estimates beyond the international sample, we further investigate for which countries the relation between occupation-level task measures (PIAAC and O*NET) showed in the last column of Table 4.3 is stronger. In Table 4.4, we find a large occupation-level correlation of both task measures in Belgium, Denmark France, Netherlands, Great Britain, or Slovenia. Moreover, we find low levels of correlation for Russia, South Korea (which is even negative for routine manual tasks), Cyprus, or Lithuania. A potential explanation for these differences may be found in differences in the degree of similarity of these economies (institutions, degree of technological changes, etc) compared to the US. As mentioned before, O*NET dataset provides job descriptors at occupational level for the US economy, and

[5] Autor and Handel (2013) find positive and high occupation-level correlation between their measures and O*NET measures for Abstract (0.65), Routine (0.48) and Manual tasks (0.63).

this may be applicable to some countries, such as some Western European countries but not to others, such as Eastern or South European Countries.

TABLE 4.4: Correlation of task measures at occupation level between PIAAC and O*NET measures, by country

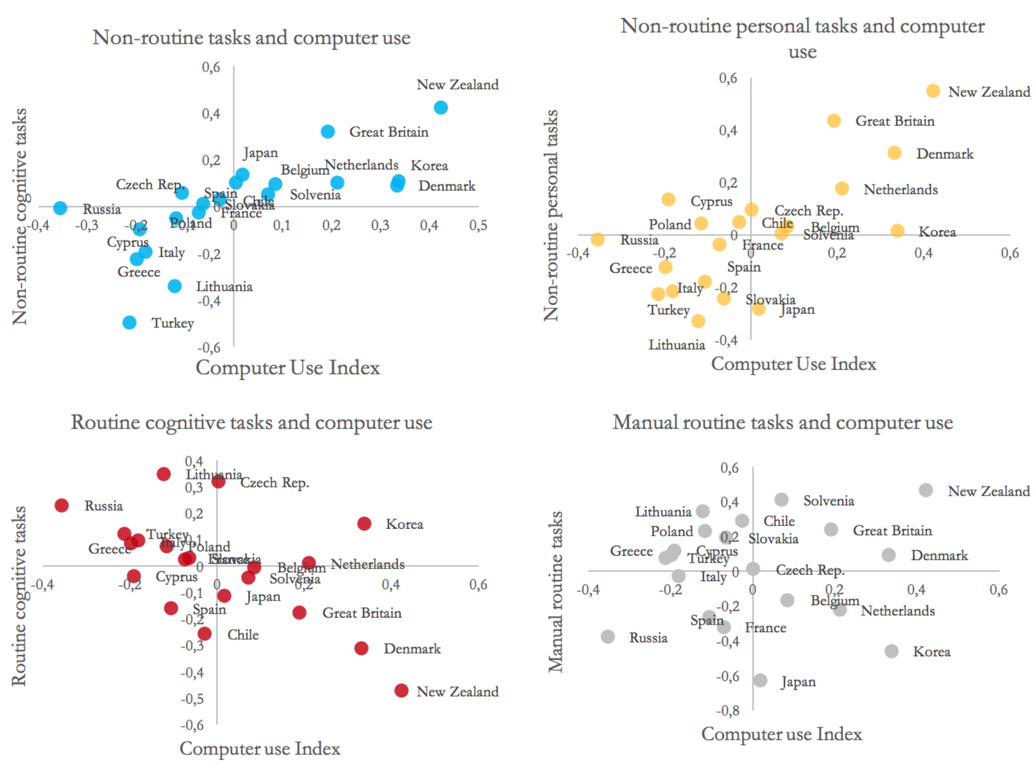
	Non routine Cognitive	Non routine Personal	Routine Cognitive	Routine Manual
Belgium	0.572	0.47	0.318	0.556
Chile	0.45	0.396	0.284	0.408
Cyprus	0.479	0.378	0.241	0.205
Czech Republic	0.59	0.49	0.356	0.277
Denmark	0.547	0.442	0.269	0.35
Spain	0.557	0.418	0.241	0.563
France	0.63	0.591	0.383	0.602
Great Britain	0.548	0.528	0.255	0.434
Greece	0.541	0.411	0.211	0.385
Italy	0.632	0.474	0.211	0.338
Japan	0.595	0.431	0.245	0.343
Korea	0.487	0.367	0.245	-0.173
Lithuania	0.493	0.47	0.285	0.107
Netherlands	0.621	0.489	0.303	0.43
Poland	0.528	0.464	0.258	0.414
New Zealand	0.522	0.531	0.318	0.276
Russia	0.408	0.467	0.219	0.228
Slovakia	0.602	0.545	0.317	0.171
Slovenia	0.622	0.617	0.385	0.219
Turkey	0.513	0.425	0.121	0.396
Overall Mean	0.536	0.456	0.264	0.304

Notes: The sample includes employed respondents aged 20-64 and currently working for which variables in Section 4.3 are well defined and have non missing values. We use one observation per occupation and country.

4.2.5 Job Tasks and Computer Adoption at Work around the world

As stated before, the implicit assumption underlying the relation between task changes and technological change is that the decline in the price of computer capital (the exogenous driver of digitalization) is expected to increase computer adoption at work. Measurement of a precise definition of computer adoption at work in a cross-country setting and its relation to job tasks is another important contribution of this paper. Previous authors have used binary indicators of computer use at work in relating technological change and the task content of jobs (Spitz Oener (2006)) or a measure of adjusted computers-per-worker at the firm level (Autor and Dorn (2013)). To exploit further variation at the worker level, we construct a measure for the individual degree of computer use at work through four different items with frequency inputs, the same qualitative responses as of those in Table 4.1[6]. In particular, we use the following items: (i) use the Internet to better understand issues related to work; (ii) conduct transactions on the Internet, for example buying or selling products or services; (iii) use spreadsheet software; (iv) participate in real-time discussions on the Internet, for example on line conferences or chat groups[7]. We then compute the first component of the principal component analysis of these four items (which explains 57.5% of the total variance) and present it in a standardized form[8].

FIGURE 4.2: Computer Use and Job Tasks across Countries.



Notes: The sample includes employed respondents aged 20-64 and currently working for which variables in section 4.3 are well defined and have non missing values. For regression purposes and due to few observations, we exclude workers in non-profit firms and workers in Armed Forces and Skilled Agricultural and Fishery occupations.

Figure 4.2 depicts an index for each of the countries under analysis. As can be seen,

[6]The five item responses are again: (i) Every Day; (ii) At least once a wee but not every day; (iii) Less than once a week; (iv) Less than once a month; (v) Never.

[7]We include a subset of all related items provided in the questionnaire when constructing the index of ICT use. This arbitrary decision is based on two reasons: we exclude items with little variation in responses and pick only one item from those that are highly correlated.

[8]When comparing the final index with the dummy used by Spitz Oener (2006), we find an individual correlation of 0.67. When comparing the average index at the country level with the proportion of workers responding yes to computer use dummy, the cross-country correlation is 0.83.

New Zealand, Denmark, Korea and Great Britain display the largest values in terms of computer use, whereas the lowest values correspond to Russia, Greece, Italy and Cyprus. Moreover, the cross-country comparison shows a strong positive correlation of computer use and non-routine cognitive (0.72) and personal (0.67) tasks, whereas a negative correlation with routine cognitive (-0.55) tasks. There is no observed relevant relation between computer use at work and manual routine task at work.

4.3 Job Tasks, Occupation Sorting and Computer Use at Work

The main advantage of job task individual-level measures is that research can exploit the variance within occupations, an issue widely discussed in the literature (Spitz Oener (2006); Autor and Handel (2013)). This section explores the potential use of job tasks measures at the individual level in order to better understand the relation between occupational sorting, education and skills and task measures. Additionally, we consider the extent to which computer use at work plays a role mostly between or within occupations in determining the relation of task measures with computerization.

Conceptually, we follow the Autor and Handel (2013) approach and start by assessing the determinants of job tasks across countries, by particularly looking at the relationship between tasks at the worker-level with key covariates within occupations. To explore this relationship, we consider for a given worker a pooled linear model with country fixed effects for countries, where:

$$T_{kij} = \alpha + \sum_{m} \beta_{1m} X_{ijm}^{Ind} + \sum_{o} \beta_{2o} X_{ijo}^{Job} + \theta ICT_{ij} + \delta_{kj} + \tau_{pj} + \epsilon_{kij} \tag{4.1}$$

where T_{kij} represents the intensity of k^{th}-job task for individual i in country j. In addition, X_{ijm}^{Ind} is a vector of m individual worker characteristics (such as gender, age, level of education or literacy and numeracy skills), X_{ijo}^{Job} is a vector of o job characteristics (public or private firm, firm size and on-the-job training), ICT_{ij} captures the use of computer for worker i at work. Finally, δ_{kj} and τ_{pj} represent country and occupational dummies. Country fixed effects capture the cross-country differences in the respective task measure index that cannot be explained by the model, whereas occupational dummies provide estimates net of occupational assignment. It is to be noted, once again, that this model accounts for differences in individual cognitive skills through the individual literacy and numeracy test scores. Job contents are undoubtedly driven at least partially by individual ability, which is in most instances unobservable to researchers. The possibility to account for them in an estimation of the determinants of job tasks allows time invariant unobserved heterogeneity to be controlled and hence, the relationship between job contents and other covariates to be cleaned from such correlation. Indeed, to our knowledge, it is the first time that cognitive skills are controlled for in an estimation of the determinants of Job Contents in a cross-country analysis.

Table 4.5 presents regression results for the task measures of non-routine cognitive, non-routine personal, routine cognitive and routine manual tasks. The first model of each task includes country dummies, individual characteristics, job characteristics as well as the worker-level index value of computer use. To control for occupational sorting, we include occupation fixed effects. Finally, in order to control for potential differences in elasticity of tasks with respect to computer use across countries, we include an interaction term of country fixed effects and computer use. First, adding occupation fixed effects increases the explanatory power of the model substantially, as we might expect. Second, it is particularly interesting to see how the importance of gender, age (not for all tasks) and education levels of workers in determining job tasks decrease significantly when controlling for occupational sorting. However, they still account for important socio-demographic marginal effects within occupation estimates. The importance of job characteristics and the economic sector also decreases significantly. Third, computer use is a key determinant of job task and, more importantly, this relation does remain constant even when accounting for occupational sorting.

The interaction terms display some variation which is significant, but overall the positive relation between computer use and non-routine tasks and the negative relation of computer use and routine cognitive tasks keeps the same sign for all countries. Nevertheless, for the case of manual routine tasks, the marginal term is positive, small and significant (Column (11)), when we include the interaction with country fixed effects.

TABLE 4.5: Task Content, Occupation Sorting and Computer Use at Work

	Non routine cognitive			Non routine personal			Routine cognitive			Routine manual		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Computer Use (Index)	0.404***	0.377***	0.422***	0.299***	0.278***	0.357***	-0.268***	-0.262***	-0.285***	-0.0578***	0.0517***	-0.208***
	(0.00535)	(0.00628)	(0.0172)	(0.00589)	(0.00673)	(0.0178)	(0.00579)	(0.00672)	(0.0159)	(0.00632)	(0.00742)	(0.0214)
Male	0.284***	0.195***	0.192***	0.102***	0.138***	0.139***	-0.123***	-0.141***	-0.142***	-0.0472***	-0.0360***	-0.0431***
	(0.00982)	(0.0109)	(0.0109)	(0.0108)	(0.0118)	(0.0118)	(0.0105)	(0.0113)	(0.0112)	(0.0119)	(0.0135)	(0.0134)
Age 25-34	0.121***	0.0824***	0.0852***	0.0404**	0.0586***	0.0635***	-0.0242	-0.0246	-0.0262	0.0272	0.0348*	0.0283
	(0.0183)	(0.0175)	(0.0174)	(0.0201)	(0.0191)	(0.0191)	(0.0192)	(0.0178)	(0.0178)	(0.0203)	(0.0193)	(0.0191)
Age 35-44	0.118***	0.0683***	0.0696***	0.124***	0.126***	0.132***	-0.0487***	-0.0341*	-0.0343**	-0.025	0.00926	0.0133
	(0.0179)	(0.0172)	(0.0172)	-0.0198	(0.0188)	(0.0188)	(0.0186)	(0.0175)	(0.0175)	(0.0201)	(0.0192)	(0.019)
Age 45-54	0.129***	0.0691***	0.0698***	0.115***	0.107***	0.113***	-0.0478**	-0.013	-0.0142	-0.0763***	-0.0277	-0.0106
	(0.0182)	(0.0176)	(0.0176)	(0.0199)	(0.019)	(0.019)	(0.0187)	(0.0176)	(0.0176)	(0.0205)	(0.0196)	(0.0195)
Age 55-64	0.0479**	-0.0237	-0.0237	0.0965***	0.0853***	0.0876***	0.00865	0.0498**	0.0494**	-0.146***	-0.0689***	-0.0490**
	(0.0198)	-0.0193	-0.0193	-0.0223	-0.0215	-0.0215	-0.0212	-0.0201	-0.0201	-0.0232	-0.0222	-0.0221
High School or Sec. VET	0.138***	0.0846***	0.0843***	0.145***	0.0805***	0.0853***	-0.154***	-0.0914***	-0.0966***	-0.0127	-0.00196	0.0143
	(0.0141	-0.0138	-0.014	-0.0154	-0.0146	-0.0147	-0.015	-0.0141	-0.0143	-0.0166	-0.0161	-0.0161
Tertiary professional	0.238***	0.115***	0.114***	0.240***	0.101***	0.107***	-0.301***	-0.123***	-0.129***	-0.0235	-0.0277	-0.00661
	(0.0179)	(0.0176)	(0.0177)	(0.0196)	(0.0187)	(0.0188)	(0.0195)	(0.0188)	(0.019)	(0.0206)	(0.0204)	(0.0204)
Tertiary college	0.362***	0.208***	0.207***	0.331***	0.133***	0.136***	-0.554***	-0.209***	-0.211***	-0.222***	-0.161***	-0.157***
	(0.0176)	(0.0182)	(0.0183)	(0.0194)	(0.0195)	(0.0196)	(0.0194)	(0.019)	(0.0192)	(0.0212)	(0.0217)	(0.0217)
Literacy Skills	-0.000382	-0.000181	-0.000228	-0.000760**	-0.000466*	-0.000485*	0.000624**	0.000576**	0.000633**	-0.00127***	-0.00102***	-0.000977***
	(0.000259)	(0.000245)	(0.000245)	(0.000299)	(0.000271)	(0.000272)	(0.000301)	(0.000271)	(0.000271)	(0.000306)	(0.000294)	(0.000293)
Numeracy Skills	0.00207***	0.00133***	0.00136***	0.00186***	0.00121***	0.00123***	-0.00114***	-0.000559**	-0.000611**	-0.000347	-0.00017	0.0000135
	(0.000244)	(0.000232)	(0.000232)	(0.000277)	(0.000253)	(0.000253)	(0.000284)	(0.000259)	(0.000259)	(0.00028)	(0.000273)	(0.000271)
Private firm	-0.107***	-0.00616	-0.00718	0.00211	0.0622***	0.0600***	0.207***	-0.0119	-0.0105	0.00744	-0.00583	0.00147
	(0.0114)	(0.0126)	(0.0125)	(0.0129)	(0.0138)	(0.0137)	(0.0132)	(0.0135)	(0.0135)	(0.0141)	(0.0157)	(0.0155)
Size: 11 to 50	0.121***	0.102***	0.102***	0.00654	0.0117	0.0125	-0.132***	-0.0978***	-0.101***	-0.00381	0.00013	-0.00194
	(0.0119)	(0.0116)	(0.0116)	(0.0134)	(0.0127)	(0.0127)	(0.0129)	(0.012)	(0.012)	(0.0142)	(0.0137)	(0.0135)
Size: 51 to 250	0.152***	0.130***	0.130***	-0.0324**	-0.00364	-0.000547	-0.117***	-0.111***	-0.116***	-0.0350**	-0.0164	-0.0198
	(0.0135)	(0.0132)	(0.0132)	(0.0152)	(0.0144)	(0.0143)	(0.0146)	(0.0137)	(0.0137)	(0.0157)	(0.0153)	(0.0151)
Size: 251 to 1000	0.193***	0.156***	0.156***	-0.0665***	-0.00832	-0.00308	-0.0465***	-0.105***	-0.113***	-0.0189	-0.00697	-0.0105
	(0.0164)	(0.0165)	(0.0165)	(0.0182)	(0.0174)	(0.0174)	(0.0174)	(0.0167)	(0.0167)	(0.0194)	(0.0192)	(0.019)
Size: more than 1000	0.240***	0.181***	0.183***	-0.0393*	0.00771	0.015	-0.110***	-0.171***	-0.178***	-0.0345	-0.0489**	-0.0544**
	(0.0203)	(0.0198)	(0.0198)	(0.0212)	(0.0204)	(0.0204)	(0.0217)	(0.0207)	(0.0206)	(0.0235)	(0.0227)	(0.0225)
On the job Training	0.246***	0.197***	0.196***	0.203***	0.149***	0.147***	-0.267***	-0.215***	-0.213***	0.0737***	0.0634***	0.0590***
	(0.0101)	(0.00978)	(0.00977)	(0.0113)	(0.0105)	(0.0106)	(0.0111)	(0.0101)	(0.0101)	(0.0115)	(0.0112)	(0.0111)
Construction	0.139***	0.0303	0.0312	0.150***	0.116***	0.118***	-0.156***	-0.107***	-0.109***	0.0565***	-0.0179	-0.0174
	(0.0195)	(0.0238)	(0.0238)	(0.0221)	(0.0269)	(0.0268)	(0.0205)	(0.0247)	(0.0246)	(0.0212)	(0.0267)	(0.0265)
Services	-0.101***	-0.0686***	-0.0652***	0.194***	0.022	0.0223	-0.123***	-0.009	-0.0105	-0.103***	-0.00961	-0.00845
	(0.0117)	(0.0134)	(0.0133)	(0.0129)	(0.0146)	(0.0146)	(0.012)	(0.0135)	(0.0135)	(0.0136)	(0.0168)	(0.0167)

Notes: Robust standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. We use Belgium as country of reference for country dummies. Observations are weighted so that countries are weighted equally.

TABLE 4.6: Task Content and Differential Country Effects of Computer Use

	Non routine cognitive		Non routine personal		Routine cognitive		Routine manual	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Computer Use (Index)	0.377***	0.422***	0.278***	0.357***	-0.262***	-0.285***	0.0517***	-0.208***
	(0.00628)	(0.0172)	(0.00673)	(0.0178)	(0.00672)	(0.0159)	(0.00742)	-0.0214
Belgium as Reference								
Chile		-0.0893***		-0.128***		0.120***		0.138***
		(0.0267)		(0.028)		(0.0283)		(0.0322)
Cyprus		-0.0707**		-0.116***		0.0694***		0.308***
		(0.0277)		(0.03)		(0.0267)		(0.0327)
Czech		0.0379		0.00724		0.0686***		0.226***
		(0.0295)		(0.0303)		(0.0234)		(0.0359)
Denmark		-0.0688***		-0.135***		0.0201		0.179***
		(0.0204)		(0.0215)		(0.0193)		(0.0258)
Spain		0.0272		-0.0482*		0.0391*		0.0171
		(0.0247)		(0.0267)		(0.0236)		(0.031)
France		0.00784		-0.0474**		-0.0121		-0.105***
		(0.0214)		(0.0224)		(0.0209)		(0.0268)
Great Britain		-0.123***		-0.118***		0.0321		0.267***
		(0.026)		(0.0268)		(0.0253)		(0.0309)
Greece		-0.00669		0.0133		0.0204		0.338***
		(0.0283)		(0.0308)		(0.0304)		(0.0408)
Italy		0.0234		-0.0499*		0.0398		0.149***
		(0.0277)		(0.0291)		(0.0256)		(0.0353)
Japan		-0.0137		-0.132***		0.0192		0.298***
		(0.0223)		(0.0234)		(0.0232)		(0.0292)
Korea		-0.0415**		-0.0920***		0.018		0.672***
		(0.021)		(0.0224)		(0.0205)		(0.0268)
Lithuania		-0.0878***		0.0806***		-0.00886		0.418***
		(0.0248)		(0.0258)		(0.0234)		(0.0282)
Netherlands		-0.0359		-0.0764***		-0.0534**		0.185***
		(0.0222)		(0.0236)		(0.0211)		(0.0301)
New Zealand		-0.135***		-0.186***		0.117***		0.354***
		(0.0225)		(0.0241)		(0.0222)		(0.0245)
Poland		-0.0654***		-0.0700***		0.0261		0.148***
		(0.0239)		(0.0245)		(0.0224)		(0.029)
Russia		-0.0395		0.0349		-0.0417		0.239***
		(0.0339)		(0.0356)		(0.0362)		(0.0455)
Slovenia		-0.0129		-0.0695***		-0.0600**		0.296***
		(0.0242)		(0.0266)		(0.0248)		(0.0293)

Notes: Robust standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Results display country differential effects of regression results in Table 4.5. We use Belgium as country of reference for country dummies. Observations are weighted so that countries weight equally.

These can be seen in more detail in Table 4.6, which shows the differential effect of computer use at work for each country, keeping Belgium (arbitrarily) as the reference country. The largest effects of computer use on non-routine cognitive tasks (Column 2) are found for the Czech Republic, Spain, France, Greece, Italy, Japan, Netherlands, Russia, Slovenia and Slovakia. Such relation is clearly smaller for Chile, Great Britain, New Zealand and Turkey. Regarding non-routine personal tasks (Column 4), the effect is largest for the Czech Republic, Greece, Russia and Slovakia, while much lower for Japan, New Zealand and Chile. With respect to routine cognitive tasks (Column 6), the negative effect is stronger in the Netherlands, Slovenia and Slovakia, and much lower in Chile and New Zealand. Finally, the relation of computer use and routine manual tasks is negative and large in France and Belgium, while positive in a large set of countries, in particular South Korea, with a very large positive relation.

Overall results show that computer use at work is a key factor in explaining differences in job contents, both between and within occupations, and that this holds even after controlling for detailed individual and job characteristics, including individual cognitive skills, which are usually unobservable. Although we cannot attribute this to be a causal relation, this brings suggestive evidence of the importance of technology adoption in understanding the task content of jobs across and within countries, whereas at the same time validates our measures of job content from a more policy perspective.

4.4 The rewards to Job Tasks in the Labor Market around the World

The previous sections defined and validated different job task measures for a variety of OECD countries and addressed the association between job tasks and important demographic and job characteristics, as well as with a measure of digitalization, approximated by the intensity of computer use at work. In this section, we turn to estimating the rewards to the different job tasks on the labor market, through the empirical relation between wages and job tasks around the world. To our knowledge, there is not a single study that addresses this relationship in a cross-country framework. In particular, we are interested in computing empirically the differences in the rewards to the different job tasks, as well as differences in these rewards across different countries. We start by presenting basic empirical relationships between (log) wages and different job contents, to learn about the predictive power of job tasks for wages, both across and within occupations. As before, we validate our task measures by comparing our job tasks definition with those obtained from O*NET dataset. However, we must assess the endogeneity of job tasks for wages to obtain true job task rewards in order to be consistent with the task approach, by which individuals decide upon which job tasks to perform on the basis of their comparative advantage. We do that by following the methodology proposed by Autor and Handel (2013) to control for the self-selection of workers into different jobs based on their comparative advantage.

4.4.1 Wage Data

TABLE 4.7: Hourly Wages (USD) PPP Corrected, by Countries

	Obs.	Mean	S.D.
Denmark	3,781	25.2	11.14
Belgium	2,507	22.5	10.48
New Zealand	2,520	19.38	12.67
Netherlands	2,739	19.29	5.73
Great Britain	2,852	18.99	12.76
Korea	2,874	18.46	17.05
Cyprus	1,932	17.68	12.66
Italy	1,648	16.8	10.74
Japan	2,971	16.79	13.75
France	3,346	16.05	8.5
Spain	2,248	15.87	11.17
Greece	1,143	11.21	8.1
Chile	1,967	11.1	11.13
Poland	3,616	9.75	7.89
Slovenia	2,113	9.47	4.82
Czech Rep.	2,390	9.37	6.06
Slovakia	2,253	8.81	6.93
Lithuania	2,404	7.76	5.83
Russia	1,240	5.16	5.14

Notes: Data reflects hourly earnings, including bonuses for wage and salary earners, in PPP corrected USD1. We exclude earnings below USD1 and above USD150.

The wage data reported by PIAAC that we use corresponds to hourly earnings with bonuses for wage and salary earners. There is no public data on earnings at the individual level for the case of Turkey among countries for which we have constructed task measures. For this reason, we exclude the Turkey data when preparing of variables for wage analysis, in particular when re-constructing task measures with the new sample of nineteen countries. Moreover, we use the conversion data to USD, corrected in Purchasing Power Parity (PPP), constructed by OECD for consistent comparisons. As can be seen in Table 4.7, Nordic European as well as Anglo-Saxon countries form the group of countries with highest hourly wages, followed by Central European, Asian and Southern European countries. Eastern European countries display the lowest mean wages.

4.4.2 The Predictive Power of Job Tasks for Wages

Our basic empirical approach consists of estimating a pooled linear model for the nineteen countries ($j = 1...19$) described above:

$$LogW_{ij} = \alpha + \sum_m \beta_{1m} X^{Ind}_{ijm} + \sum_o \beta_{2o} X^{Job}_{ijo} + \sum_k \beta_{3k} Task_{ijk} + \sum_k \beta_{4k} \overline{Task_{jk}} + \delta_j + \epsilon_{kij} \quad (4.2)$$

with $LogW_{ij}$ being the hourly log-wage, X_{ijm} individual m worker characteristics: these include gender, age or level of education, but more importantly, a unique standardized measure of individual ability approximated by the individual test scores on reading and mathematical competencies. Using this last measure, a greater part of the supply component of unobserved worker ability are rewarded to the task. X^{Job}_{ijo} includes a vector of job characteristics (public or private firm, firm size and on-the-job training), and $Task_{ijk}$ is the intensity of non-routine cognitive, non-routine personal, routine cognitive and routine manual tasks which each worker reports to exert in her work. We also include the average mean of each task at occupation-country level $\overline{Task_{jk}}$ in order to net out the pure individual job task reward from the association between job tasks and occupations. As before, this mean is a leave-out mean, representing the average intensity of k-th task for all workers from a particular country in occupation p except for the i-th worker.

Table 4.8 displays several empirical specifications of the relationship between wages and each of the previously defined tasks. In particular, Column 1 addresses the raw (or unconditional) relationship between wages and individual and job characteristics in a Mincer regression where we include country dummies. Columns 2 and 3 consider task measures, with and without controlling for worker ability measures. The first key result is that we observe a positive relation between non-routine tasks and hourly earnings, while a negative and small relation of routine cognitive and manual tasks with earnings. Returns to individual and job characteristics slightly decrease when considering job task measures. Moreover, the returns to the index of computer use at work also decrease when considering measures of worker cognitive skills.

When ability measures are included in Column 3, the wage returns to tasks decrease mostly for non-routine tasks (from 3.1% of hourly wages of one standard deviation of the cognitive task measure to 2.6% and from 2.0% to 1.8% in the personal task measure), but also for manual routine tasks. In Columns 4 and 5 we include variables at the occupational level: we first consider task measures of PIAAC averaged at the occupational level (Column 4) and then include occupational dummies (Column 5). The positive returns to tasks remain, except for non-routine cognitive tasks when considering tasks measures at the occupational level, which reveals the importance of within occupation task measures to understand wage returns in the sample of countries. The returns to task measures at the occupational level are also substantial. This indicates that heterogeneity of job tasks within and between occupations is important, and hence measuring it at the occupational level only would be greatly misleading.

Finally, it is interesting to compare the empirical relationship between tasks and wages found in this paper, where 20 OECD and non-OECD countries are used, to that found by Autor and Handel (2013), who uses a similar framework but with only US data. Autor and Handel (2013) find that within occupations, a one standard deviation increase in abstract tasks is related to a 7 log point wage premium, whereas we find that the same increase in abstract tasks (non-routine cognitive and non-routine personal) predicts a 1 to 3 log point wage premium. Hence, on average, in the countries considered in this study, the rewards to abstract tasks are much lower than those found by Autor and Handel (2013) for the US economy.

To check the validity of our results, we also consider separate regressions for PIAAC and O*NET occupation-level tasks. In Table 4.9, we find positive and similar returns to non-cognitive routine tasks and modest to insignificant returns to non-routine personal tasks. First, the PIAAC estimate in Column 1 has (slightly) more predictive power to explain wage differences than the O*NET estimates in Column 2. With respect to routine cognitive tasks, only PIAAC occupation-level data remains significant in explaining the negative wage returns to routine cognitive tasks, whereas O*NET data does not display significant results. When including tasks at the individual level (PIAAC) together with occupation-level data (using PIAAC or O*NET) in Columns 3 and 4, the magnitude of returns at the occupation-level tasks remains constant, except for non-routine personal tasks, which changes to a negative (although small) effect in the case of PIAAC data at the occupation level. The individual measure estimates are complemented by statistically significant estimates for both measures of occupation-level tasks. Finally, including both measures at the same time (Column 5) together with individual task measures shows complementarity for non-routine cognitive tasks using O*NET and PIAAC at the occupation level, but a negligible effect in the rest of O*NET task measures.

TABLE 4.8: Log-Wage Returns to Tasks, Demographics and Occupation Variables

	(1)	(2)	(3)	(4)	(5)
Male	0.186***	0.168***	0.156***	0.147***	0.121***
	(0.00616)	(0.00627)	(0.00651)	(0.00846)	(0.00824)
Upper Secondary	-0.0838***	-0.0650***	-0.0301***	-0.0204**	-0.0231***
	(0.00877)	(0.00875)	(0.00889)	(0.00921)	(0.00884)
Post-secondary or Tertiary Professional	0.0973***	0.0839***	0.0708***	0.0552***	0.0408***
	(0.00955)	(0.00951)	(0.00946)	(0.00997)	(0.00919)
Tertiary (Bachelor/Master)	0.286***	0.249***	0.217***	0.181***	0.139***
	(0.0083)	(0.00841)	(0.00861)	(0.0119)	(0.0107)
25-34	0.117***	0.112***	0.114***	0.107***	0.0992***
	(0.012)	(0.0119)	(0.0118)	(0.0134)	(0.0124)
35-44	0.240***	0.232***	0.235***	0.225***	0.210***
	(0.0119)	(0.0118)	(0.0117)	(0.0154)	(0.0145)
45-54	0.279***	0.270***	0.277***	0.264***	0.245***
	(0.0119)	(0.0118)	(0.0117)	(0.0183)	(0.0174)
55 plus	0.298***	0.293***	0.310***	0.292***	0.274***
	(0.0139)	(0.0137)	(0.0136)	(0.0226)	(0.0206)
On the Job Training	0.0725***	0.0500***	0.0463***	0.0427***	0.0401***
	(0.00644)	(0.00645)	(0.00641)	(0.00666)	(0.00649)
Private sector	-0.0244***	-0.00896	-0.0104	0.0101	0.00612
	(0.00727)	(0.00728)	(0.0072)	(0.0106)	(0.00995)
Firm size	-0.129***	-0.119***	-0.115***	-0.110***	-0.107***
	(0.00865)	(0.00862)	(0.00858)	(0.0104)	(0.00983)
Firm size	-0.0332***	-0.0339***	-0.0334***	-0.0328***	-0.0314***
	(0.00797)	(0.00788)	(0.00784)	(0.0073)	(0.00747)
Firm Size	0.0598***	0.0650***	0.0628***	0.0631***	0.0584***
	(0.00995)	(0.00984)	(0.0097)	(0.00866)	(0.00886)
Firm size	0.131***	0.131***	0.129***	0.129***	0.115***
	(0.0134)	(0.0132)	(0.013)	(0.0132)	(0.0127)
Computer Use Index	0.115***	0.0821***	0.0716***	0.0593***	0.0461***
	(0.00355)	(0.00378)	(0.00382)	(0.00564)	(0.00528)
Literacy Skills			-0.000169	-0.000152	-0.000174
			(0.00019)	(0.000169)	(0.000167)
Numeracy Skills			0.00154***	0.00137***	0.00124***
			(0.000179)	(0.000155)	(0.000153)
Non-routine cognitive (PIAAC Worker)		0.0311***	0.0258***	0.00216	0.0126***
		(0.00393)	(0.00392)	(0.00407)	(0.00386)
Non-routine personal (PIAAC Worker)		0.0202***	0.0177***	0.0202***	0.0220***
		(0.00403)	(0.004)	(0.00421)	(0.00491)
Routine cognitive (PIAAC Worker)		-0.0443***	-0.0450***	-0.0174***	-0.0228***
		(0.00429)	(0.00424)	(0.00488)	(0.00467)
Routine manual (PIAAC Worker)		-0.0225***	-0.0194***	-0.0152***	-0.0159***
		(0.00326)	(0.00324)	(0.00379)	(0.00373)
Non-routine cognitive (PIAAC Occ.)				0.0812***	
				(0.0118)	
Non-routine personal (PIAAC Occ.)				-0.0285*	
				(0.015)	
Routine cognitive (PIAAC Occ.)				-0.0889***	
				(0.0156)	
Routine manual (PIAAC Occ.)				-0.00305	
				(0.0142)	
Worker Ability measures	No	No	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes
Occupational Dummies	No	No	No	No	Yes
Constant	2.397***	2.411***	2.042***	2.099***	2.264***
	(0.019)	(0.0189)	(0.0308)	(0.0423)	(0.0423)
Observations	46,544	46,530	46,530	46,530	46,530
R-squared	0.523	0.531	0.536	0.543	0.566

Notes: Robust standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Data reflects log hourly earnings, including bonuses for wage and salary earners, in PPP corrected USD. We exclude earnings below USD1 and above USD150. All regressions with occupation-level variables cluster standard errors at the occupation level.

TABLE 4.9: Log-wage Returns to Tasks within and between Occupations comparing O*NET and PIAAC

	(1)	(2)	(3)	(4)	(5)	(6)
NRC (PIAAC Worker)			0.00216	0.0166***	0.00302	0.00279
			(0.00407)	(0.00517)	(0.00404)	(0.004)
NRP (PIAAC Worker)			0.0202***	0.0162***	0.0203***	0.0199***
			(0.00421)	(0.00504)	(0.00422)	(0.00419)
RC (PIAAC Worker)			-0.0174***	-0.0366***	-0.0177***	-0.0159***
			(0.00488)	(0.00594)	(0.00488)	(0.00494)
RM (PIAAC Worker)			-0.0152***	-0.0176***	-0.0158***	-0.0114***
			(0.00379)	(0.00413)	(0.00381)	(0.00392)
NRC (PIAAC Occ.)	0.0766***		0.0812***		0.0552***	0.0845***
	(0.0115)		(0.0118)		(0.0156)	(0.0121)
NRP (PIAAC Occ.)	-0.00867		-0.0285*		-0.0235	-0.0247*
	(0.0143)		(0.015)		(0.0162)	(0.0149)
RC (PIAAC Occ.)	-0.104***		-0.0889***		-0.0741***	-0.0799***
	(0.0154)		(0.0156)		(0.0168)	(0.016)
RM (PIAAC Occ.)	-0.0163		-0.00305		-0.00505	0.00394
	(0.0136)		(0.0142)		(0.0129)	(0.0141)
NRC (O*NET Occ.)		0.0775***		0.0757***	0.0551***	
		(0.00896)		(0.00873)	(0.00982)	
NRP (O*NET Occ.)		0.00347		-0.00765	-0.0098	
		(0.00978)		(0.00985)	(0.00983)	
RC (O*NET Occ.)		-0.0162		-0.012	-0.00928	
		(0.0107)		(0.0101)	(0.00889)	
RM (O*NET Occ.)		0.0105		0.0147	0.0206**	
		(0.00996)		(0.00947)	(0.00849)	
PIAAC NRC (worker)*PIAAC NRC (occ.)						-0.0126*
						(0.00674)
PIAAC NRP (worker)*PIAAC NRP (occ.)						0.00198
						(0.00804)
PIAAC RC (worker)*PIAAC RC (occ.)						0.0132**
						(0.00605)
PIAAC RM (worker)*PIAAC RM (occ.)						0.0191***
						(0.00576)
Individual, Job and Ability measures	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Constant	2.091***	2.070***	2.099***	2.098***	2.125***	2.093***
	(0.0424)	(0.0446)	(0.0423)	(0.0429)	(0.0422)	(0.0431)
Observations	46,544	46,544	46,530	46,530	46,530	46,530
R-squared	0.541	0.538	0.543	0.542	0.545	0.543

Notes: Robust standard errors in parentheses. *, ** and *** indicate significance at 10-, 5- and 1-percent level, respectively. Data reflects log hourly earnings, including bonuses for wage and salary earners, in PPP corrected USD. We exclude earnings below USD1 and above USD150. All regressions with occupation-level variables cluster standard errors at the occupation level. NRC refers to non-routine cognitive, NRP refers to non-routine personal, RC refers to routine cognitive and RM refers to routine manual tasks.

Overall, it can be said that returns to tasks are partially captured by worker occupations, but there is still substantial heterogeneity of tasks within occupations. Hence, we need to consider not only definition of job contents at occupational level, but also at individual level to capture average returns to tasks appropriately. Additionally, compared to the PIAAC occupation-level measures, the O*NET measures at the occupation level are, if anything, less relevant and less precise, as the standard errors of O*NET measures are substantially higher than their PIAAC counterpart. These results are highly consistent with those found by Autor and Handel (2013).

4.4.3 Accounting for Sorting of Workers into Occupations

Within the task approach, workers decide their job contents based on their comparative advantage. Consequently, an OLS approach of Job Tasks on Wages would not drive consistent estimates of the job task rewards. Instead, we consider, as in a Roy (1951) model, that workers are utility (wage) maximizers and hence select into those jobs/occupations where they attain the highest wage given their skills. Each job requires a different combination of different tasks to be performed simultaneously.

Accounting empirically for the self-selection of workers into different jobs to estimate adequately job tasks rewards is not straightforward. One possibility is to find adequate instruments to circumvent the endogeneity of tasks due to the self-selection of workers into jobs, but unfortunately we do not have such instruments. Instead, we hypothesize, as the task approach does, that workers self-select into those occupations with higher returns to those tasks which they perform relatively better i.e., have comparative advantage. From the empirical point of view, this implies that occupational level task returns must show a positive covariance with individual level task returns. Hence, and following Autor and Handel (2013), we consider an augmented version of a classical mincerian regression, where in addition to the individual and occupation endowment of tasks, presented above, we include an interaction term between individual and occupation task returns. Moreover, Autor and Handel (2013) shows that if all interaction terms are positive, this means that workers that are highly efficient in conducting specific tasks self-select into occupations which reward those tasks, resulting in a comparative, rather than absolute advantage. Instead, the absolute advantage hypothesis is less restrictive and requires that only one of the interaction coefficients is positive[9].

Results are displayed in Column 6 of Table 4.9 and reveal that the interaction between routine (cognitive and manual) individual and occupation based measures are

[9] The test presented by Autor and Handel (2013) does not provide dispositive evidence, but represents rather an exploratory approximation to the issue.

significantly positive, whereas such correlation is not found (at 5% level) for the non-routine (cognitive and personal) interaction between individual and occupation measures. As Autor and Handel (2013) discuss, these results are consistent with absolute advantage. Additionally, when we include such interactions, individual measures of job tasks, with the exception of non-routine cognitive, still show significant predictive power to explain wage differences, even within occupations. In particular, a one standard deviation increase in non-routine personal tasks predicts a 2 log point wage premium, whereas the same magnitude of increase in routine cognitive or in routine manual tasks predicts a 2 log point wage penalty.

This can be mostly extrapolated to the same argument within countries. We look at the sign and magnitude of results by countries in Table 4.12 in Section 4.6. The results are not exactly homogeneous within countries, with variation in terms of the sign of the interaction terms and the significance. It is true, at the same time, that at least two interaction terms are positive for all countries, and three out of four task interaction terms are positive for five of them (Denmark, Spain, France, Great Britain, and Greece), with joint significance for France and Great Britain.

4.5 Conclusion

In this paper, we investigate for the first time cross-country differences in the intensity of job tasks presented in the Autor et al. (2003) model. We therefore use for the first time the background questionnaire items from the the Programme for International Assessment of Adult Competences (PIACC) database, implemented in more than 30 countries between 2011 and 2014. The dataset provides very precise information on job contents at the worker level, which allows for job task heterogeneity within occupations when accounting for differences on job contents, a unique feature only available in very few national surveys in the past. Additionally, the data includes novel data of an accurate measurement of cognitive skills in literacy, numeracy and problem solving skills, so that unobserved worker characteristics can be accounted for.

This allows us to have a clear picture of job tasks at the individual level and compare it across individual and job characteristics, including key variables of the ALM model such as computer use at work. The Nordic, Anglo-Saxon and East Asian advanced countries (Great Britain, Denmark, New Zealand, Japan and Korea) form the group of countries where jobs exhibit a higher intensity of non-routine task content of jobs (both cognitive and personal dimensions). At the other end, there is the group of Eastern European countries, such as Lithuania, Turkey Russia and the Czech Republic, which display low levels of non-routine cognitive tasks, whereas high levels of routine cognitive tasks.

Our measures derived from PIAAC data for 20 countries are consistent and display large correlation with O*NET, the other main measure used at the occupational level in the literature. In particular, relative to O*NET, the PIAAC measures shows larger consistency for the measures of non-routine tasks and for Western European countries, such as Belgium, Denmark France, Netherlands, and Great Britain, compared to Eastern European countries. This validity approach confirms that the task measurement followed is highly valuable for this current and future research, whenever new PIAAC waves are implemented again. It may allow for dynamic analyses of the consequences of technology adoption on job polarization or job de-routinization and the redistributive impact this may be having on the workforce population.

At the same time, when looking at the determinants of job tasks at work, we find that the task measures are relevant to explain differences both between but also within occupation task measures, but do not matter if we control for individual and job characteristics. The importance and statistical significance of computer use at work is large and relevant overall, hence revealing a high predictive power of computer use at work to explain tasks determinants. We also find that task heterogeneity within occupation is relevant across and within countries, even after controlling for individual and job characteristics. Moreover, we find a positive and consistent relation of computer use at work and non-routine tasks, but a negative relation with routine tasks, consistent with the Autor et al. (2003)) framework. Although this assessment has a static point of view, we observe that the relation between computer use at work and task content of jobs is relevant both across and within countries. Finally, similar to what Autor and Handel (2013)) describe, we observe clear signs of self-selection of workers based on their comparative advantage to be rewarded in tasks which they more effectively perform.

Finally, results on how tasks affect earnings show clear signs of worker self-selection into jobs according to their ability to perform tasks at job. Moreover, it yet again shows the consistency (and even complementarity) of PIAAC task measures when comparing additional task measures such as O*NET regarding the wage returns to task measures.

The returns to task measures at the occupational level are also substantial, but worker-level information is still key. This indicates that heterogeneity of job tasks within and between occupations is important and hence measuring it at the occupational level only would be greatly misleading. These two findings (the comparison with O*NET and the importance of within-occupation worker level data) are highly consistent to what is found by Autor and Handel (2013) for the United States. It is to be noted that our paper goes beyond a single country approach, as it allows us to extend such results to a large set of diverse countries, hence providing external validity to the Autor and Handel (2013) approach. Finally, and as a robustness check, we show that our approach is consistent with a Roy model of self-selection.

From a policy perspective, this study is important to level the playing field of a fundamental policy discussion such as the impact of technological change and automation of the job content of workers. This will be at the core of social sciences in the coming decades. Our paper provides a consistent and promising avenue of future research on the consequences of technological change on the labor markets from an international perspective, whenever second or third waves of PIAAC (or national longitudinal studies) are implemented worldwide. Moreover, it confirms the importance of collecting worker level information of job activities and habits (already discussed in previous studies for the US and Germany), and hence the need to move away from occupation-level studies.

4.6 Appendix

TABLE 4.10: Results of principal component analysis and cross-tasks correlations.

	Computation		Correlations			
	Number of components	Variation of first component	Non routine Cognitive	Non routine Personal	Routine Cognitive	Routine Manual
Non-routine cognitive	3	0.546	1	-	-	-
Non-routine personal	2	0.712	0.466	1	-	-
Routine cognitive	4	0.449	-0.487	-0.597	1	-
Routine manual	1	-	-0.029	0.0005	-0.014	1

Notes: The sample includes employed respondents aged 20-64 currently working for which variables in section 4.3 are well defined and have non missing values. Observations are weighted so that countries are weighted equally.

TABLE 4.11: Distribution of Task Measures by Countries.

	Obs.	NRC		NRP		RC		RM	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Belgium	1,883	-0.006	1.002	0.094	1.005	0.032	0.926	0.022	1.012
Chile	1,055	0.166	1.013	0.24	0.995	-0.372	1.011	0.442	0.865
Cyprus	1,267	-0.201	1.046	0.141	1.035	0.01	1.026	0.445	0.886
Czech Rep.	1,639	0.097	1.052	0.216	1.045	0.386	0.852	0.255	0.952
Denmark	3,175	-0.095	0.933	0.28	0.888	-0.204	0.897	0.354	0.896
Spain	1,370	0.129	1.097	-0.074	1.062	-0.214	1.024	-0.127	1.019
France	2,390	-0.09	1.024	0.036	1.011	0.037	1.014	-0.196	0.996
Great Britain	2,193	0.2	1.013	0.478	0.978	-0.122	0.935	0.533	0.81
Greece	676	-0.252	1.049	0.02	0.945	0.036	1.024	0.444	0.888
Italy	982	-0.079	1.049	-0.048	1.028	0.061	0.92	0.133	1.033
Great Britain	2,243	0.017	0.963	-0.337	0.878	-0.012	0.98	-0.28	0.973
Ireland	1,939	0.112	1.026	0.074	0.94	0.159	1.008	0.119	0.986
Italy	1,337	-0.328	0.917	-0.152	1.014	0.197	0.927	0.664	0.712
Japan	2,306	-0.073	1.017	0.158	0.952	0.114	0.942	0.052	1.022
Korea	1,953	0.288	0.97	0.532	0.927	-0.366	0.921	0.748	0.595
Netherlands	2,043	-0.041	0.914	0.216	1.027	-0.047	0.996	0.347	0.913
Norway	670	0.021	0.91	0.208	1.023	0.095	1.029	-0.205	0.963
Poland	1,263	0.021	1.045	-0.139	1.088	-0.099	1.097	0.445	0.881
Russia	1,467	-0.044	1.02	0.082	1.051	-0.038	1.079	0.641	0.74
Slovakia	691	-0.526	0.936	-0.081	0.963	-0.039	1.136	0.319	0.897

Notes: Results display values of standardized indexes for each task, with mean 0 and standard deviation 1. Observations are weighted so that countries are weighted equally. NRC refers to non-routine cognitive, NRP refers to non-routine personal, RC refers to routine cognitive and RM refers to routine manual tasks.

TABLE 4.12: Interaction terms and joint significance test results.

	NRC	NRP	RC	RM	F (joint)	p-value
Belgium	-0.024	-0.002	0.011	0.036	2.873	0.023
Chile	0.053	0.007	-0.065	0.003	1.568	0.183
Cyprus	-0.049	-0.003	0.064	0.03	5.226	0
Czech Rep.	-0.013	0.027	-0.015	0.025	0.816	0.516
Denmark	0.015	0.01	0.005	0.023	1.301	0.27
Spain	-0.001	0.027	0.004	0.024	1.381	0.241
France	-0.001	0.022	0.02	0.072	12.903	0
Great Britain	0.001	0.051	0.011	-0.028	2.143	0.078
Greece	-0.029	0.012	0.036	0.042	0.95	0.436
Italy	-0.036	0.075	0.023	-0.005	4.691	0.001
Japan	-0.034	-0.005	0.065	0.008	1.51	0.199
Korea	-0.023	0.006	-0.018	0.042	0.834	0.504
Lithuania	-0.035	-0.054	0.045	0.005	1.775	0.134
Netherlands	-0.002	0.013	-0.002	0.025	2.283	0.06
New Zealand	-0.015	0.029	-0.004	0.084	2.889	0.023
Poland	-0.011	0.06	-0.054	0.039	5.473	0
Russia	-0.22	-0.026	0.067	0.087	4.217	0.003
Slovakia	-0.01	0.037	-0.031	0.047	1.916	0.108
Solvenia	-0.006	0.045	-0.008	0.012	3.075	0.017

Notes: Data reflects interaction terms of tasks at the worker and occupational level for country regressions of log hourly earnings, including bonuses for wage and salary earners, in PPP corrected USD. We exclude earnings below USD1 and above USD150. All regressions with occupation-level variables cluster standard errors at the occupation level. We include individual and job controls. NRC refers to non-routine cognitive, NRP refers to non-routine personal, RC refers to routine cognitive and RM refers to routine manual tasks.

Chapter 5

Conclusion

This chapter presents the main findings of Chapter 2, Chapter 3 and Chapter 4, and discusses the contributions in the context of the previous research and briefly comments on the policy implications.

5.1 Chapter 2

Motivated by the sharp decline in the Programme for International Student Assessment (PISA) test scores in 2015 in the Basque Country region in Spain, I study the reasons behind the abrupt changes across socio-demographic factors, school factors, and idiosyncratic factors of the Basque education system. I use decomposition methods between PISA 2015 results and previous rounds in 2012, 2009, 2006 and 2003, to understand the changes in performance of students in the Basque Country.

The analysis in Chapter2 allows only part of this decline to be clarified with observable factors, where the explained part share varies depending on the reference year and the subject I compare. There are three major factors which are responsible for this explained decline. First of all, the increase in repetition rates in 2015 (both among 1-year repeaters and 2-year repeaters) causes a decrease in performance, and it is no surprise given that repetition is known to be related to ineffective learning progress (Hattie, 2008). Second, the testing language matters for performance: the Basque Country is a bilingual region with Spanish being the language of use at home for the majority of students, but Basque is the primary language of instruction at the majority of schools. Relative to 2012, in 2015 a larger proportion of students took the test in a language different to the language at home. In particular, the change in 2015 had to do with more students speaking Spanish at home that are enrolled in the Basque immersion model (Model D) at school and who now take the test in Basque: this leads to the students being slightly penalized in terms of test performance. Finally, a third factor is related to the school principal perception of student behavior at school: between 2009 and 2015 this perception has deteriorated, an issue which is related to part of the observed decline. What is more important, the economic crisis has not played a key role in the decline, as one would expect given the deterioration of economic and social conditions in certain households.

The paper contributes with key evidence on a major policy question for the Basque education system, following the literature on decomposition methods that using results from international assessments to compare learning differences across and within countries. The multiplicity of factors behind the results and the fact that a large share of the decline remains unexplained requires the findings to be taken with caution. More importantly, PISA remains as a low-stakes test where student effort varies considerably and is not necessarily related to factors of the education system, but rather to cultural reasons (Zamarro et al. (2016)).

5.2 Chapter 3

Chapter 3 addresses a key question in the literature of economics of education: whether more choice generates differential family responses and hence leads to greater student sorting. This paper uses novel and rich administrative data to address the impact of a large school choice reform implemented in the Madrid region and its capital Madrid: the reform generated single school choice zones for all medium and large-size municipalities, a measure which significantly increased the number of schools available to households for students entering primary schools. By looking at new entrants to the system aged 3, we find that the reform increased the level of geographic mobility of students, with a large response by higher educated families and families of non-immigrant children, whereas families with immigrant children did not respond to the reform at all. We then look at whether the reform had an indirect effect on the school composition of students, and find that the reform mildly increased the average level of social diversity, but with heterogeneous results by school performance: while schools in the bottom of the average performance distribution experienced a decrease (increase) in diversity (segregation), schools in the 3^{rd}, 4^{th} and 5^{th} quintile of the distribution showed a decreased in the level of sorting. What is more important, results present solid evidence of a large increase of immigrant segregation with the implementation of this policy, which was exacerbated by residential dynamics of immigrant population. We conduct several robustness checks to strengthen our identification, including a control group of municipalities in the region which implemented the reform with no time to anticipate the policy changes.

In terms of the interpretation and mechanisms behind the reform effects, it is also noticeable that most of the changes are driven by the interest of households in attending semi-public schools (semi-publicly funded but privately managed) and schools with low and high scores. The puzzling results in terms of school segregation (small decrease of social segregation while a large increase in immigrant segregation) will require further research, although the data on parental education may be implying an under-estimation of the true effects of the reform on social segregation. The results are important because they contribute to closing the gap in terms of validity of previous research given the usual lack of identification of large-scale reforms described by Epple, Romano, and Urquiola (2017) in a recent literature review.

With respect to the policy implications, one cannot ignore the context of the reform. At the time of the reform, the region of Madrid reported very high levels of social segregation in secondary schools in comparison with other Spanish and European regions. Our paper looks at primary education students, but in any case, the small changes in social segregation may be explained by the fact that the original level of social segregation was potentially near to a maximum. In contrast, the effects with regard to immigrant segregation are better understood when put in its policy context: the region of Madrid shows relatively low levels of immigrant segregation (in secondary schools), and hence the increase observed in the data becomes a more natural response. In any case, a broader policy discussion is required to better understand the transition from primary to secondary schooling and the impact this has on the student composition of schools. For the future policy agenda, the results acknowledge the existence of a trade-off between equity and choice.

5.3 Chapter 4

Chapter 4 studies the cross-country differences in the intensity of job tasks at the worker level following the seminal work of Autor et al. (2003). This is an important aspect for the literature studying the recent changes in the tasks workers perform at work and its relation with strong forces such as technology, globalization or aging. We use the OECD's PIAAC database in more than 30 countries, which brings in two new features. First, the worker survey enables job tasks to be measured at the worker level (a unique feature only available in two national surveys in the past), which allows within-occupation differences of job tasks to be understood. Second, the survey is accompanied by a cognitive skill assessment, which introduces a very precise ability measure of workers in order to control for unobserved characteristics.

We are able to depict a very clear picture of job tasks at the individual level across countries. Consistent with previous evidence from occupational data, countries in Northern Europe, countries with Anglo-Saxon influence (such as Great Britain or New Zealand) and East Asian countries (such as Japan and Korea) display the largest intensity in non-routine job content. At the other end, there is a group of Eastern European countries, such as Lithuania, Turkey Russia and the Czech Republic, which display low levels of non-routine cognitive tasks, whereas high levels of routine cognitive tasks.

What is more important, our measures are consistent and display a high degree of correlation with O*NET at the occupational level. This is fundamental, given that O*NET data for occupations is the most commonly used measure in the literature of job tasks and the discussion on the future of work. In particular, with respect to O*NET, our PIAAC task constructs display larger consistency for the measurement of non-routine tasks for Western European countries, such as Belgium, Denmark France, Netherlands, and Great Britain, compared to Eastern European countries. On the other hand, we find that the task measures are indeed relevant in explaining variation within occupations (controlling for several individual and job characteristics), which improves the relevance of the findings as opposed to O*NET occupation data.

Another two findings are important and introduce key contributions. First, we find computer use to be highly important and statistically significant to understand task differences at work, hence revealing the highly predictive power of computer use at work to explain task content of jobs: we find a positive and consistent relation of computer use at work and non-routine tasks, whereas a negative relation with routine tasks, consistent with the Autor et al. (2003)) framework. With respect to task and earnings, our results display that workers self-select into jobs according to their ability to perform tasks at work. Once again, the results on the wage returns to tasks add consistency to our PIAAC task measures.

Our approach confirms that the task measurement followed is highly valuable for this current and future research, whenever new PIAAC waves are implemented again, as it will allow worker-level dynamics to be tracked in terms of job polarization or other hypothesis with a more robust approach. From a policy perspective, this study is important to level the playing field of a fundamental policy discussion such as the impact of technological change and automation of the job content of workers. This will be at the core of social sciences in the coming years.

References

ABDULKADIROĞLU, A., Y.-K. CHE, AND Y. YASUDA (2011): "Resolving conflicting preferences in school choice: The" boston mechanism" reconsidered," *American Economic Review*, 101, 399–410.

ABDULKADIROGLU, A., P. PATHAK, A. E. ROTH, AND T. SONMEZ (2006): "Changing the Boston school choice mechanism," Tech. rep., National Bureau of Economic Research.

ABDULKADIROGLU, A. AND T. SÖNMEZ (2003): "School choice: A mechanism design approach," *The American Economic Review*, 93, 729–747.

ACEMOGLU, D. AND D. AUTOR (2011): *Skills, tasks and technologies: Implications for employment and earnings*, vol. 4.

AEDO, C., J. HENTSCHEL, J. LUQUE, AND M. MORENO (2013): "From occupations to embedded skills : a cross-country comparison," *Background Paper for the World Development Report*, 1–26.

AMMERMUELLER, A. (2007): "PISA: What makes the difference?: Explaining the gap in test scores between Finland and Germany," *Empirical Economics*, 33, 263–287.

ANDERSON, S., A. PALMA, AND J. THISSE (1992): *Discrete Choice Theory of Product Differentiation*, The MIT Press.

ANGHEL, B., A. CABRALES, AND J. M. CARRO (2016): "Evaluating a bilingual education program in spain: The impact beyond foreign language learning," *Economic Inquiry*, 54, 1202–1223.

ANGRIST, J. D. AND V. LAVY (1997): "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco," *Journal of Labor Economics*, 15, S48–S76.

ARELLANO, M. AND G. ZAMARRO (2007): "The choice between public and private schools with or without subsidies in Spain," *Unpublished*.

AUTOR, D. (2013): "The" task approach" to labor markets: an overview," *National Bureau of Economic Research*.

AUTOR, D. H. AND D. DORN (2013): *American Economic Review*, 103, 1553–1597.

AUTOR, D. H. AND M. J. HANDEL (2013): "Putting Tasks to the Test: Human Capital, Job Tasks, and Wages," *Journal of Labor Economics*, 31, S59–S96.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2009): "The Polarization of the U . S . Labor Market," *The American Economic Review*, 96, 189–194.

AUTOR, D. H., F. LEVY, AND R. J. MURNANE (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly Journal of Economics*, 118, 1279–1333.

AVVISATI, F. AND F. KESLAIR (2016): *REPEST: Stata module to run estimations with weighted replicate samples and plausible values*.

BALART, P. AND A. CABRALES (2017): *La maratón de PISA: la perseverancia como factor del éxito en una prueba de competencias*.

BARRERA-OSORIO, F., V. GARCÍA-MORENO, H. A. PATRINOS, AND E. PORTA (2011): "Using the oaxaca-blinder decomposition technique to analyze learning outcomes changes over time: An application to indonesia's results in pisa mathematics," *Regional and Sectoral Economic Studies*, 11, 65–84.

BILLINGS, S. B., D. J. DEMING, AND J. ROCKOFF (2014): "School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg," *Quarterly Journal of Economics*, 129, 435–476.

BÖHLMARK, A., H. HOLMLUND, AND M. LINDAHL (2016): "Parental choice, neighbourhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform," *Journal of Population Economics*, 29, 1155–1190.

BÖHLMARK, A. AND M. LINDAHL (2007): "The impact of school choice on pupil achievement, segregation and costs: Swedish evidence," *IZA Discussion Paper Series*, 65.

BREAKSPEAR, S. (2012): "The policy impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance," *OECD Journals*, 1–32.

BURGESS, S. AND A. BRIGGS (2010): "School Assignment, School Choice and Social Mobility," *Economics of Education Review*, 29, 639–649.

BURGESS, S., E. GREAVES, A. VIGNOLES, AND D. WILSON (2015): "What Parents Want: School Preferences and School Choice," *The Economic Journal*, 125, 1262–1289.

BURGESS, S., C. PROPPER, AND D. WILSON (2007): "The Impact of School Choice in England," *Policy Studies*, 28, 129–143.

CALSAMIGLIA, C. (2014): "Matching Practices for Elementary and Secondary Schools Spain," Tech. rep., MiP Country Profile 17.

CALSAMIGLIA, C. AND M. GÜELL (2018): "Priorities in School Choice: The case of the Boston mechanism in Barcelona," *Journal of Public Economics, Forthcoming*.

CARD, D., M. D. DOOLEY, AND A. A. PAYNE (2010): "School competition and efficiency with publicly funded catholic schools," *American Economic Journal: Applied Economics*, 2, 150–176.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *The American Economic Review*, 106, 855–902.

COSTA, A. AND N. SEBASTIÁN-GALLÉS (2014): "How does the bilingual experience sculpt the brain?" .

CULLEN, J., B. JACOB, AND S. LEVITT (2006): "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica*, 74, 1191–1230.

DE HAAN, M., P. A. GAUTIER, H. OOSTERBEEK, AND B. VAN DER KLAAUW (2015): "The performance of school assignment mechanisms in practice," .

DEPARTAMENTO DE EDUCACIÓN POLÍTICA LINGÜÍSTICA Y CULTURA: BASQUE GOVERNMENT, A. (2016): *Datos de la Enseñanza No Universitaria en Euskadi: Curso 2014-2015*.

DINARDO, J., N. M. FORTIN, AND T. LEMIEUX (1996): "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001.

DUFLO, E., P. DUPAS, AND M. KREMER (2011): "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review*, 101, 1739–1774.

DUNCAN, O. D. AND B. DUNCAN (1955): "A Methodological Analysis of Segregation Indexes," *American Sociological Review*, 20, 210.

EPPLE, D., R. E. ROMANO, AND M. URQUIOLA (2017): "School Vouchers: A Survey of the Economics Literature," *Journal of Economic Literature*, 55, 441–492.

EURYDICE (2011): *Grade Retention during Compulsory Education in Europe: Regulations and Statistics*.

FACK, G. AND J. GRENET (2010): "When do better schools raise housing prices? Evidence from Paris public and private schools," *Journal of public Economics*, 94, 59–77.

FIRPO, S., N. FORTIN, AND T. LEMIEUX (2009): "Unconditional Quantile Regressions," *Econometrica*, 1–29.

FIRPO, S., N. M. FORTIN, AND T. LEMIEUX (2007): "Unconditional quantile regression," *NBER Technical Working Paper Series*.

FRIEDMAN, M. (1955): *The Role of Government in Education*, University of Rutgers Press, New Brunswick.

GARCÍA-PÉREZ, J. I., M. HIDALGO-HIDALGO, AND J. A. ROBLES-ZURITA (2014): "Does grade retention affect students' achievement? Some evidence from Spain," *Applied Economics*, 46, 1373–1392.

GIBBONS, S., S. MACHIN, AND O. SILVA (2006): "The Educational impact of parental choice and school competition," .

GIMPELSON, V. AND R. KAPELIUSHNIKOV (2016): "Polarization or upgrading? Evolution of employment in transitional Russia," *Russian Journal of Economics*, 2, 192–218.

GOOS, M., A. MANNING, AND A. SALOMONS (2014): "Explaining job polarization: Routine-biased technological change and offshoring," *American Economic Review*, 104, 2509–2526.

GORTAZAR, L., K. HERRERA-SOSA, AND D. KUTNER (2014): *How can Bulgaria improve its education system? : an analysis of PISA 2012 and past results*.

GOUSS, M. AND M. LEDONN (2014): "Why Do Inequalities in 15-Year-Old Cognitive Skills Increased So Much in France?" .

HARDY, W., R. KEISTER, AND P. LEWANDOWSKI (2016): "Technology or Upskilling? Trends in the Task Composition of Jobs in Central and Eastern Europe," *HKUST Institute for Emerging Market Studies*.

HASTINGS, J., T. KANE, AND D. STAIGER (2009): "Heterogeneous preferences and the efficacy of school choice." *NBER Working Papers No. 2145*.

HASTINGS, J., C. NEILSON, AND S. ZIMMERMAN (2012): "The Effect of School Choice on Intrinsic Motivation and Academic Outcomes," *NBER Working Paper Series 18324, National Bureau of Economic Research, Inc.*

HASTINGS, J. AND J. WEINSTEIN (2008): "Formation, School Choice, and Academic Achievement: Evidence from Two Experiments," *The Quarterly Journal of Economics*, 123, 1373–1414.

HATTIE, J. (2008): *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*.

HOXBY, C. (2000): "Does Competition among Public Schools Benefit Students and Taxpayers?" *The American Economic Review*, 90, 1209–1238.

HSIEH, C. AND M. URQUIOLA (2006): "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program," *Journal of Public Economics*, 1477–1503.

HUBER, S. AND C. RUST (2016): "Calculate Travel Time and Distance with OpenStreetMap Data Using The Open Source Routing Machine (OSRM)," *The Stata Journal*, 16, 1–8.

ISEI-IVEI (2004): *Influencia de la lengua de la prueba en los resultados de las evaluaciones*.

——— (2012): *Investigación sobre la influencia de la lengua de la prueba en las evaluaciones internacionales*.

——— (2016a): *Evaluación de Diagnóstico. 2015. Informe Ejecutivo. 2 ESO*.

——— (2016b): *Evaluación de Diagnóstico. 2015. Informe Ejecutivo 4 EP*.

IVLEVS, A. AND R. M. KING (2014): "2004 Minority Education Reform and pupil performance in Latvia," *Economics of Education Review*, 38, 151–166.

JACOB, B. A. AND L. LEFGREN (2009): "The Effect of Grade Retention on High School Completion," *American Economic Journal: Applied Economics*, 1, 33–58.

JAKCSON, K., C. WIGGER, AND H. XIONG (2018): "Do School Spending Cuts Matter? Evidence from the Great Recession," *National Bureau of Economic Research (No. w24203)*.

LADD, H. F. AND E. B. FISKE (2001): "The uneven playing field of school choice: Evidence from New Zealand," *Journal of Policy Analysis and Management*, 20, 43–64.

LAVY, V. (2010): "Effects of Free Choice Among Public Schools," *Review of Economic Studies*, 74, 11641191.

LEVIN, H. M. (1998): "Educational vouchers: Effectiveness, choice, and costs," *Journal of Policy Analysis and Management*, 17, 373–392.

——— (2002): "A Comprehensive Framework for Evaluating Educational Vouchers," *Educational Evaluation and Policy Analysis*, 24, 159–174.

LOUNKAEW, K. (2013): "Explaining urban-rural differences in educational achievement in Thailand: Evidence from PISA literacy data," *Economics of Education Review*, 37, 213–225.

MACHADO, J. A. AND J. MATA (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20, 445–465.

MANACORDA, M. (2012): "The Cost of Grade Retention," *Review of Economics and Statistics*, 94, 596–606.

MANCEBÓN, M. J., J. CALERO, Á. CHOI, AND D. P. XIMÉNEZ-DE-EMBÚN (2012): "The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006," *Journal of the Operational Research Society*, 63, 1516–1533.

MECD (2017): *Sistema estatal de indicadores de la educacin: Edicin 2017*.

MIZALA, A. AND M. URQUIOLA (2013): "School markets: The impact of information approximating schools' effectiveness," *Journal of Development Economics*, 103, 313–335.

MURILLO, F. J. AND C. GARRIDO (2018): "Magnitud de la Segregación escolar por nivel socioeconómico en España y sus Comunidades Autónomas y comparación con los países de la Unión Europea," *RASE: Revista de la Asociación de Sociología de la Educación*, 1, 37–58.

MURILLO, F. J., C. GARRIDO, AND G. BELAVI (2017): "Segregación escolar por origen nacional en España," *OBETS. Revista de Ciencias Sociales*, 2, 395–423.

MUSSET, P. (2012): "Shool choice and equity: current policies in OECD countries and literature review," *Education Working Paper*, 1–51.

NEUMARK, D. (1988): "The Board of Regents of the University of Wisconsin System Employers' Discriminatory Behavior and the Estimation of Wage Discrimination," *Source: The Journal of Human Resources*, 23, 279–295.

NIETO, S. AND R. RAMOS (2015): "Educational outcomes and socioeconomic status: A decomposition analysis for middle-income countries," *Prospects*, 45, 325–343.

OAXACA, R. (1973): "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14, 693.

OECD (2012): "PISA 2009 Technical Report," Tech. rep.

——— (2014): "Are disadvantaged students more likely to repeat grades?" *PISA in Focus*, 1–4.

——— (2016a): *Education at a Glance 2016: OECD Indicators*.

——— (2016b): "PISA 2015 Results (Volume I): Excellence and Equity in Education," Tech. rep.

OJIMA, F. AND S. VON BELOW (2010): "Family background, school system and academic achievement in Germany and in Japan," in *Quality and Inequality of Education: Cross-National Perspectives*, 275–297.

RAMOS, R., J. DUQUE, AND S. NIETO (2012): "Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata," *IZA Discussion Paper*, 5731, 1–25.

RASCH, G. (1960): *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.*

REARDON, S. AND G. FIREBAUGH (2002): "Measures of multigroup segregation," *Sociological methodology, 32(1), 33-67.*

ROGERO-GARCÍA, J. AND M. ANDRÉS-CANDELAS (2018): "Elección de Centro educativo y preferencias familiares en España," *Paper Manuscript*.

ROTHSTEIN, J. (2006): "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions," *The American Economic Review*, 96, 1333–1350.

ROY, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3, 135–146.

RUIZ-VALENZUELA, J. (2015): "Job loss at home: childrens school performance during the Great Recession in Spain," *London School of Economics and Political Science, LSE Library*.

SAVE THE CHILDREN SPAIN, S. (2017): *Desheredados: Desigualdad infantil, igualdad de oportunidades y políticas públicas en España*.

SELLAR, S. AND B. LINGARD (2014): "The OECD and the expansion of PISA: New global modes of governance in education," *British Educational Research Journal*, 40, 917–936.

SHAFIQ, M. N. AND J. P. MYERS (2014): "Educational Vouchers and Social Cohesion: A Statistical Analysis of Student Civic Attitudes in Sweden, 19992009," *American Journal of Education*, 121, 111–136.

SÖDERSTRÖM, M. AND R. UUSITALO (2010): "School choice and segregation: Evidence from an admission reform," *Scandinavian Journal of Economics*, 112, 55–76.

SPITZ OENER, A. (2006): "Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure," *Journal of Labor Economics*, 24, 235–270.

TESKE, P. AND M. SCHNEIDER (2001): "What research can tell policymakers about school choice," *Journal of Policy Analysis and Management*, 20, 609–631.

VÄLIJAÄRVI, J. AND H. ALASUUTARI (2017): "The Current Challenges of Finnish Education: Changes of the Society Affecting Quality and Equity in Schools." *World Bank, Washington, DC.*

YANG HANSEN, K. AND J. E. GUSTAFSSON (2016): "Causes of educational segregation in Sweden school choice or residential segregation," *Educational Research and Evaluation*, 22, 23–44.

ZAMARRO, G., C. HITT, AND I. MENDEZ (2016): "When Students Don't Care: Reexamining International Differences in Achievement and Non-Cognitive Skills." *EDRE Working Paper No. 2016-18*.