Estudios sobre la Economía Española - 2018/17

The PISA "shock" in the Basque Country: Contingent factors or structural change?

Lucas Gortazar
(University of the Basque Country and The World Bank Group)

fedea

Resumen Ejecutivo

En este artículo exploro las razones por las cuales el País Vasco experimentó un gran revés en los resultados de la prueba PISA de 2015. El sistema educativo vasco había tenido un desempeño alto en el pasado, tanto en evaluaciones nacionales como internacionales, así como en términos de acceso en la educación superior (comparable a la de los sistemas educativos del norte de Europa). Por lo tanto, entender esta caída es una cuestión clave desde un punto de vista de las políticas educativas, más aún cuando ese mismo año los resultados de PISA estuvieron acompañados por disminuciones en las evaluaciones diagnóstico implementadas por las autoridades vascas.

En particular, exploro las razones del declive analizando los factores socioeconómicos (dado el impacto de la Gran Recesión en la región), junto con los factores clave del sistema educativo vasco. Presto especial atención a factores que pueden estar relacionados con la naturaleza misma del sistema, como su modelo lingüístico, la coexistencia de servicios educativos públicos y privados o las altas tasas de repetición experimentadas por la región en comparación con la OCDE y la UE. Utilizo técnicas de descomposición derivadas de la literatura de la economía laboral para comprender los cambios promedio en el rendimiento entre 2015 con respecto a ediciones anteriores (2009 y 2012), así como los cambios en la distribución de resultados.

Los resultados solo permiten identificar parte de este declive por factores observables, donde el porcentaje explicado varía según el año de referencia y el dominio de la prueba. Hay tres factores principales que son responsables de esta caída explicada. En primer lugar, el aumento en las tasas de repetición en 2015 (tanto entre repetidores de un curso como repetidores de dos cursos) está asociado a una disminución en el rendimiento, y no es sorprendente dado la evidencia disponible sobre la repetición y su impacto negativo sobre el aprendizaje (Hattie, 2008). En segundo lugar, el lenguaje de prueba es importante para el rendimiento: el País Vasco es una comunidad bilingüe, donde el castellano es el idioma de uso en el hogar para la mayoría de los estudiantes, pero el euskera es el idioma principal de instrucción en la mayoría de los centros educativos. En relación con 2012, en 2015 una mayor proporción de estudiantes realizó la prueba PISA en un idioma diferente al idioma en el hogar. En particular, el cambio en 2015 tuvo que ver con que más estudiantes que hablan castellano en el hogar que están matriculados en el modelo de inmersión de euskera (Modelo D) y que en 2015 tomaron la prueba en euskera: esto lleva a que los estudiantes sean ligeramente penalizados en términos del aprendizaje medido por la prueba. Finalmente, un tercer factor está relacionado con la percepción del director del centro sobre el comportamiento del alumno en la escuela: entre 2009 y 2015 esta percepción se ha deteriorado, lo cuál que está relacionado con parte del declive observado. Finalmente, la crisis económica no está asociada a la caída de los resultados de PISA, como cabría esperar dado el deterioro de las condiciones económicas y sociales en muchos hogares.

Este artículo aporta evidencia clave sobre una pregunta importante de política educativa para el sistema educativo vasco. Lo hace siguiendo la literatura sobre métodos de descomposición de los resultados educativos para comparar las diferencias de aprendizaje entre y dentro de los países. La multiplicidad de factores detrás de los resultados y el hecho de que gran parte de la disminución permanece sin explicación requiere que los resultados sean tomados con cautela. Más aún cuando PISA es una prueba de tipo *low-stakes*, en la que el esfuerzo del alumno varía considerablemente y no está necesariamente relacionado con factores del sistema educativo, sino más bien con factores culturales (Zamarro et al. (2016)).

The PISA "shock" in the Basque Country: Contingent factors or structural change?*

Lucas Gortazar †

August 10, 2018

Abstract

I study the sharp decline in performance experienced in 2015 by the Basque Country region (Spain) in secondary education student outcomes, as measured by the OECD's PISA assessment. I construct a harmonized and comparable dataset for the Basque Country since 2003 and examine how differences in observable student and school characteristics have affected changes in student outcomes. Despite the economic crisis experienced by the Basque economy since 2008, the increase in socioeconomic characteristics of students slows down but does not decrease in 2015, hence having little effect on the decline in performance. Conversely, I find three factors that may help explain part of the decline in scores since 2009 and 2012 and which affect students across all the performance distribution: an increase of repeaters, an increase of students which take the test in a language different from their regular language at home, and the perceived increases by the school principals in student behavior problems at school. The results call for a cautious response and interpretation by public authorities given the low-stakes nature of the PISA test and the diversity of factors affecting the decline.

JEL Codes: I21, I24, I28.
Keywords: Student learning, Quality of Education, PISA, Achivement gaps.

*I am deeply thankful to Xabier Aizpurua, Francisco Luna, and Eduardo Ubieta from the Basque Institute for Research and Evaluation in Education (ISEI-IVEI), for responding to complementary data requests. I thank Sara De La Rica, David Martinez de Lafuente, Fernando Fernandez-Monge, Juan Manuel Moreno, Miguel Ruiz and Alfonso Unceta for their helpful comments. I am grateful to the seminar participants at the BRIDGE Bilbao workshop (University of the Basque Country). All errors and omissions are of my own.

†E-mail: lucas.gortazar@gmail.com. University of the Basque Country and The World Bank Group.

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 (Sellar and Lingard [41],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 (Breakspear [6], 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.

[1] See ISEI-IVEI [21] (2016a) and ISEI-IVEI [20] (2016b).

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 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 discusses the literature related to changes in education quality in education systems using cross-section data. Section 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 4. The results and its analysis are shown in Section 5. Section 6 provides a final discussion and concludes.

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. Amermueller [1] (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 [30], 2015; Ojima and Von Below [36], 2010). Other studies have analyzed the nature of certain observable gaps within countries: Nieto et al [37](2013) 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 [5](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, Gortázar et al [15] (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-Blinder [31](1973). 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 (Di Nardo et al [10](1996), Machado and Mata [26](2001) and Firpo, Fortin, Lemieux [13](2009)). 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, Fortin, Lemieux [13](2009) decomposition has already been used in the literature with education production functions. For example, Lounkaew [25](2013) uses a Firpo-Fortin-Lemieux approach to decompose differences between rural and urban student performance in Thailand using PISA 2009 data. Moreover, Goussé et Ledonné [16] (2015) llook 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 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 [15] (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. Among high-performing students, the improvement of socioeconomic conditions and the changes in peer composition at schools accounted for most of the learning improvements.

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].

lthough 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 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 et al [39](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 Guell [7][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].

[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.

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

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

[8]Source: Basque Government [9] (2016)

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, 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 [22], 2013). 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 [18], 2004; ISEI-IVEI [19], 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

[9] Source:EUSTAT

[10] See Costa and Sebastián-Gallés [8] (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 [2] (1997) find positive returns to bilingual education in the labor market in Morocco.

[12] Source: EUSTAT

[13] Source: Basque Government [9]. 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

increased the pressure on the system to provide quality education to a more diverse student population, and at the 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 [40], 2015). On the other hand, Jackson et al [24](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 [34], 2017[16]).

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[11], 2017), 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

5,270 were born in EU countries (around 15%), whereas 14,436 (more than 40%) where born in African countries.

[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

opportunities to students lagging behind (Hattie [17], 2008; Manacorda, [27] 2012;Jacob [23], 2009). García-Perez et al [14] (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 [33], 2014).

4 Data and Variables

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].

[17] I do so by using the REPEST command in STATA (Avvisati et al [3], 2014)

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.

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.

Regarding student characteristics, I include student's gender, age, language at home, and the country of birth. I include the Index of Economic, Social, and Cultural Status (ESCS), built 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 (1 or 2 grades) previously. I do not include the variable of years in pre-primary school as there was a change in the question structure in 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 be-

[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.

cause 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 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).

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. Once the items are identified, I follow a similar approach as OECD (OECD [32], 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

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

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 [38] (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 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 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 [42], 2017).

[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 A.1 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 [32], 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.

Table 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.

Beyond ESCS, I re-compute the rest of the indexes, which can be seen in Table 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 A.1 of the Appendix.

Table 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 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 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 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.

5 Results

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 [31] for the mean and the method one proposed by Firpo, Fortin and Lemieux [12](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 1:

$$Y_{ij} = \alpha + X_i^{'}\beta + P_j^{'}\theta + S_j^{'}\gamma + \epsilon_i \tag{1}$$

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:

$$Y_{ij,t_1}-Y_{ij,t_0} = (\alpha_{t_1}-\alpha_{t_0})+(X_{i,t_1}^{'}\beta_{t_1}-X_{i,t_0}^{'}\beta_{t_0})+(P_{j,t_1}^{'}\theta_{t_1}-P_{j,t_0}^{'}\theta_{t_0})+(S_{j,t_1}^{'}\gamma_{t_1}-S_{j,t_0}^{'}\gamma_{t_0})+(\epsilon_{i,t_1}-\epsilon_{i,t_0}) \tag{2}$$

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 and rearranging terms gives the expression:

$$
\begin{aligned}
Y_{i,t_1} - Y_{i,t_0} = \{(X'_{i,t_1} - X'_{i,t_0})\beta_{t_1} + (P'_{j,t_1} - P'_{j,t_0})\theta_{t_1} + (S'_{j,t_1} - S'_{j,t_0})\gamma_{t_1})\}+ \\
\{X'_{i,t_0}(\beta_{t_1} - \beta_{t_0}) + P'_{j,t_0}(\theta_{t_1} - \theta_{t_0}) + S'_{j,t_0}(\gamma_{t_1} - \gamma_{t_0}) + (\alpha_{t_1} - \alpha_{t_0}) + (\epsilon_{i,t_1} - \epsilon_{i,t_0})\}
\end{aligned} \tag{3}
$$

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) [25],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 3 is multiplied by the following term b^*:

$$
b^* = W\beta_{t_1} + (I - W)\beta_{t_0} \tag{4}
$$

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 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:

$$
R = (X_{t_1} - X_{t_0})'\{W\beta_{t_1} + (I - W)\beta_{t_0}\} + \{X'_{t_1}(I - W) + X'_{t_0}W\}(\beta_{t_1} - \beta_{t_0}) \tag{5}
$$

In the context of the OLS regression, I follow the Neumark [29] (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, Fortin and Lemeiux (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:

$$RIF(I; q_\tau) = q_\tau + \frac{\tau - D(I \leq q_\tau)}{f_I(q_\tau))} \tag{6}$$

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:

$$RIF(I; \hat{q_\tau}) = \hat{q_\tau} + \frac{\tau - D(I \leq \hat{q_\tau})}{f_I\hat{(}q_\tau)} \tag{7}$$

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.

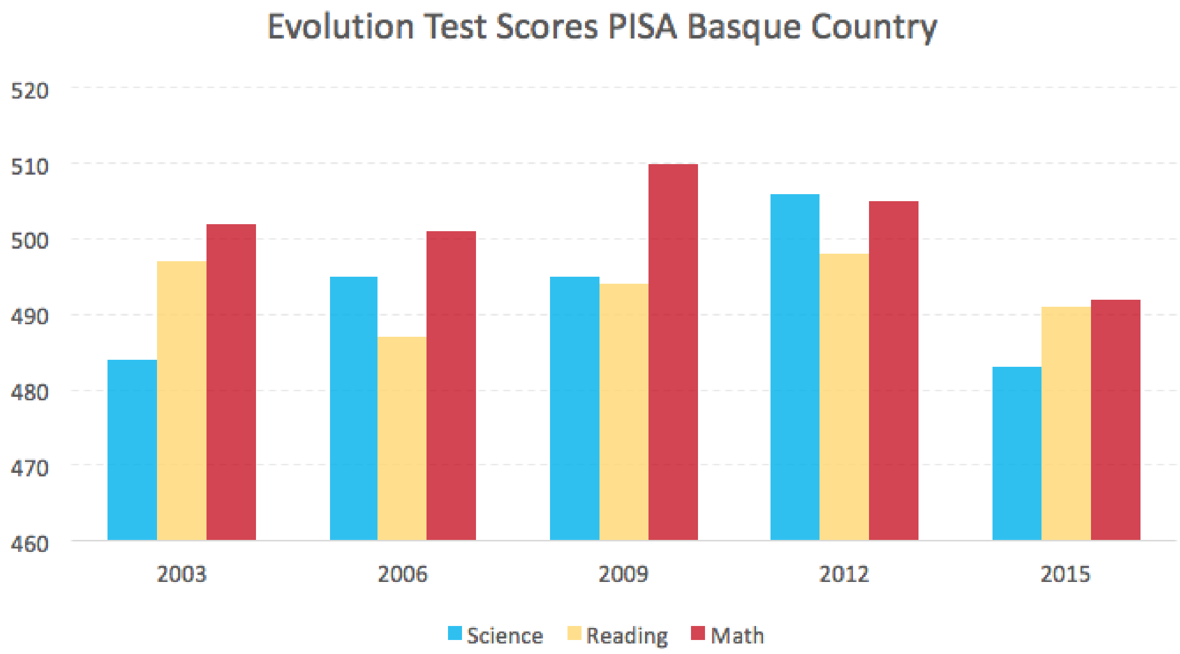
5.2 Descriptive Statistics

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 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 (see OECD[35](2016a)). 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.

[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.

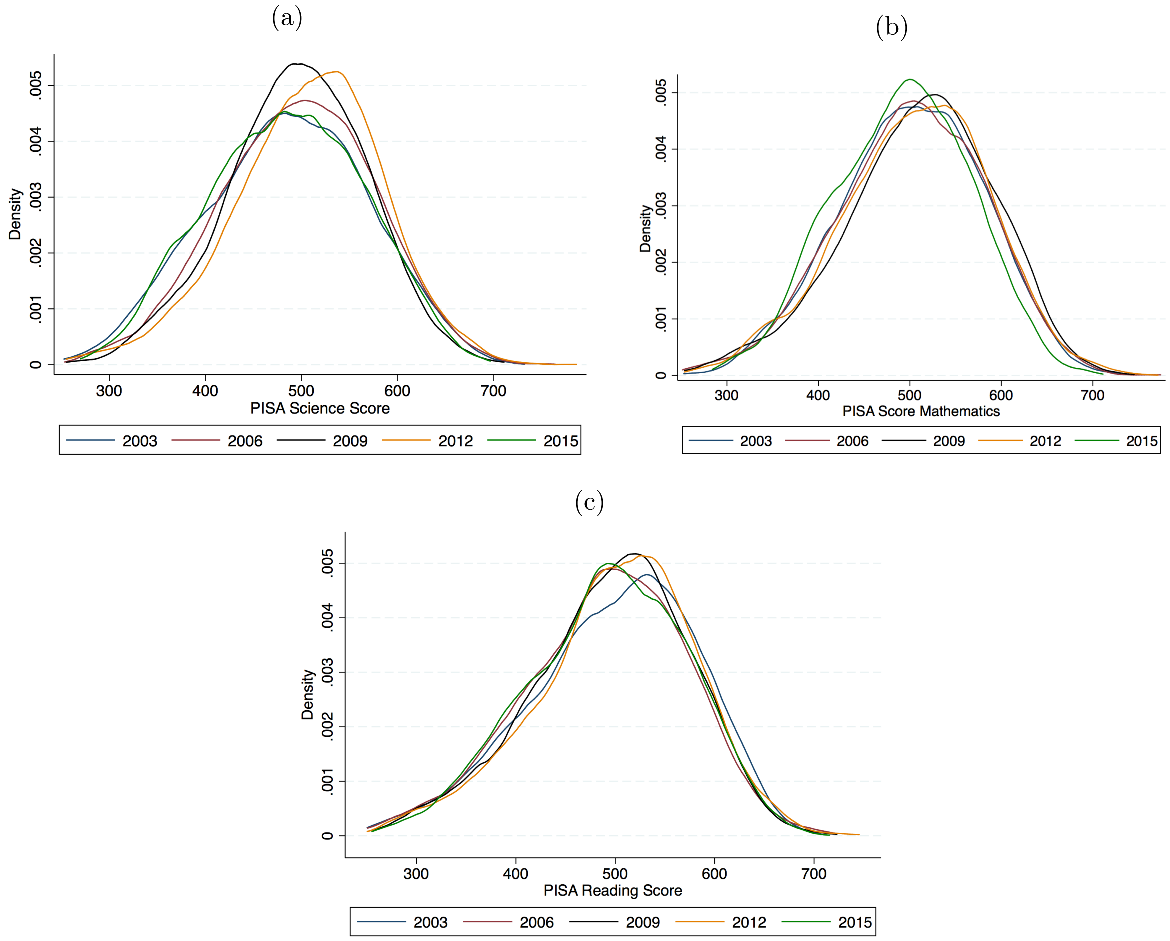
Figure 1: Changes in PISA Student Performance in the Basque Country



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

Beyond computing the evolution of mean scores, I report the changes in the distribution of scores across years for the three domains. Figure 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 4.

Figure 2: Distribution of PISA performance in the Basque Country across domains and years.



I compute the average scores of the variables of interest described in the previous section by year. Results in Table 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 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].

[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.

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 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.

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

[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 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
Private (Model A)	0.25	0.43	0.21	0.41	0.15	0.36	0.11	0.31	0.08	0.27
Private (Model B)	0.17	0.38	0.16	0.37	0.20	0.40	0.22	0.41	0.23	0.42
Private (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.

5.3 Results

I first display results of the linear regression model for 2015, with the covariates included in Table 4. Table 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 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 [34], 2017). 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 Garcia-Perez et al [14](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 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
Private (Model A)	-1.61	-0.77	2.45
Private (Model B)	-12.18	-9.10	-12.09
Private (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 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 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.

[29] In Table A.2, 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.

Table 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	-23.43***	-15.60***	-7.56	-9.02**	-15.85***	-0.06
Explained	-8.66**	-6.93*	-6.79**	-4.71	-5.92*	-2.73
Unexplained	-14.78***	-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
Private (Model A)	-0.14	-0.12	-0.16	1.50	0.27	1.51
Private (Model B)	-0.08	0.06	0.07	0.07	0.21	0.05
Private (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 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 5 and Table 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, $p20$, $p50$, and $p80$. Regarding the comparison for 2012, I find that the explained differences are larger in the low percentile of the distribution ($p20$), 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 ($p80$), 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 7[31]. For the case of the comparison with respect to 2009, the explained differences are larger in the low-achieving students ($p20$). 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 $p20$.

[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 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.

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.

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.

[32]See MECD(2017) [28].

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 Välijärvi et al [43](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 [44] (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 [4] (2014)-, 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 of the test (as shown by Zamarro et al [44], 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.

References

[1] A. Ammermueller. PISA: What makes the difference?: Explaining the gap in test scores between Finland and Germany. *Empirical Economics*, 33(2):263–287, 2007.

[2] J. D. Angrist and V. Lavy. The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco. *Journal of Labor Economics*, 15(1, Part 2):S48–S76, 1997.

[3] F. Avvisati and F. Keslair. *REPEST: Stata module to run estimations with weighted replicate samples and plausible values.* 2016.

[4] P. Balart and A. Cabrales. *La maratón de PISA: la perseverancia como factor del éxito en una prueba de competencias.* 2017.

[5] F. Barrera-Osorio, V. García-Moreno, H. A. Patrinos, and E. Porta. 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(3):65–84, 2011.

[6] S. Breakspear. The policy impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance. *OECD Journals*, (71):1–32, 2012.

[7] C. Calsamiglia and M. Güell. Priorities in school choice: The case of the boston mechanism in barcelona.

[8] A. Costa and N. Sebastián-Gallés. How does the bilingual experience sculpt the brain?, 2014.

[9] A. Departamento de Educación Política Lingüística y Cultura: Basque Government. *Datos de la Enseñanza No Universitaria en Euskadi: Curso 2014-2015.* 2016.

[10] J. DiNardo, N. M. Fortin, and T. Lemieux. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, 64(5):1001, 1996.

[11] Eurydice. *Grade Retention during Compulsory Education in Europe: Regulations and Statistics.* 2011.

[12] S. Firpo, N. Fortin, and T. Lemieux. Unconditional Quantile Regressions. *Econometrica*, pages 1–29, 2009.

[13] S. Firpo, N. M. Fortin, and T. Lemieux. Unconditional quantile regression. *NBER Technical Working Paper Series*, (339), 2007.

[14] J. I. García-Pérez, M. Hidalgo-Hidalgo, and J. A. Robles-Zurita. Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics*, 46(12):1373–1392, 2014.

[15] L. Gortazar, K. Herrera-Sosa, and D. Kutner. *How can Bulgaria improve its education system? : an analysis of PISA 2012 and past results.* 2014.

[16] M. Goussé and M. Ledonné. Why Do Inequalities in 15-Year-Old Cognitive Skills Increased So Much in France? 2014.

[17] J. Hattie. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* 2008.

[18] ISEI-IVEI. *Influencia de la lengua de la prueba en los resultados de las evaluaciones.* 2004.

[19] ISEI-IVEI. *Investigación sobre la influencia de la lengua de la prueba en las evaluaciones internacionales.* 2012.

[20] ISEI-IVEI. *Evaluación de Diagnóstico. 2015. Informe Ejecutivo. 2º ESO.* 2016.

[21] ISEI-IVEI. *Evaluación de Diagnóstico. 2015. Informe Ejecutivo 4º EP.* 2016.

[22] A. Ivlevs and R. M. King. 2004 Minority Education Reform and pupil performance in Latvia. *Economics of Education Review*, 38:151–166, 2014.

[23] B. A. Jacob and L. Lefgren. The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3):33–58, 2009.

[24] K. Jakcson, C. Wigger, and H. Xiong. Do School Spending Cuts Matter? Evidence from the Great Recession. *National Bureau of Economic Research (No. w24203)*, 2018.

[25] K. Lounkaew. Explaining urban-rural differences in educational achievement in Thailand: Evidence from PISA literacy data. *Economics of Education Review*, 37:213–225, 2013.

[26] J. A. Machado and J. Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20(4):445–465, 2005.

[27] M. Manacorda. The Cost of Grade Retention. *Review of Economics and Statistics*, 94(2):596–606, 2012.

[28] MECD. *Sistema estatal de indicadores de la educación: Edición 2017.* 2017.

[29] D. Neumark. 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(3):279–295, 1988.

[30] S. Nieto and R. Ramos. Educational outcomes and socioeconomic status: A decomposition analysis for middle-income countries. *Prospects*, 45(3):325–343, 2015.

[31] R. Oaxaca. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693, 1973.

[32] OECD. PISA 2009 Technical Report. Technical report, 2012.

[33] OECD. Are disadvantaged students more likely to repeat grades? *PISA in Focus*, (09):1–4, 2014.

[34] OECD. *Education at a Glance 2016: OECD Indicators.* 2016.

[35] OECD. PISA 2015 Results (Volume I): Excellence and Equity in Education. Technical report, 2016.

[36] F. Ojima and S. Von Below. Family background, school system and academic achievement in Germany and in Japan. In *Quality and Inequality of Education: Cross-National Perspectives*, pages 275–297. 2010.

[37] R. Ramos, J. Duque, and S. Nieto. Decomposing the rural-urban differential in student achievement in Colombia using PISA microdata. *IZA Discussion Paper*, 5731(6515):1–25, 2012.

[38] G. Rasch. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* 1960.

[39] J. Rogero-García and M. Andrés-Candelas. Elección de Centro educativo y preferencias familiares en España. *Paper Manuscript*, 2018.

[40] J. Ruiz-Valenzuela. Job loss at home: children's school performance during the Great Recession in Spain. *London School of Economics and Political Science, LSE Library*, 2015.

[41] S. Sellar and B. Lingard. The OECD and the expansion of PISA: New global modes of governance in education. *British Educational Research Journal*, 40(6):917–936, 2014.

[42] S. the Children Spain. *Desheredados: Desigualdad infantil, igualdad de oportunidades y políticas públicas en España.* 2017.

[43] J. Välijärvi and H. Alasuutari. The Current Challenges of Finnish Education: Changes of the Society Affecting Quality and Equity in Schools. *World Bank, Washington, DC.*, 2017.

[44] G. Zamarro, C. Hitt, and I. Mendez. When Students Don't Care: Reexamining International Differences in Achievement and Non-Cognitive Skills. *EDRE Working Paper No. 2016-18*, 2016.

7 Appendix

Table A.1: 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 A.2: 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 A.3: 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 A.1: 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			