The use (and misuse) of PISA in guiding policy reform: the case of Spain

Álvaro Choi [1]
John Jerrim [2]

1 Barcelona Institute of Economics & University of Barcelona, Barcelona, Spain
2 University College London, Institute of Education, London, England

Contact details: Álvaro Choi (alvarochoi@ub.edu). Departament of Public Economics, Political Economy and Spanish Economy, Faculty of Economics and Business, University of Barcelona. 690 Av. Diagonal, 08034, Barcelona, Spain.
Tel. +34 934021816

Álvaro Choi

Associate professor in the Public Economy, Political Economy and Spanish Economy of the Faculty of Economics and Business of the University of Barcelona. Ph. D. in Economics (2009). Member of the Interdisciplinary Research Group on Educational Policies (GIPE-Grup Interdisciplinari sobre Polítiques Educatives) and member of the Barcelona Institute of Economics (IEB-Institut d'Economia de Barcelona). His main research interests are the Economics of Education, the evaluation of educational policies and economic development.

John Jerrim

John Jerrim is Reader in Educational and Social Statistics at the Department of Social Science of UCL-Institute of Education. John's research interests include the Economics of Education, access to higher education, intergenerational mobility, cross-national comparisons and educational inequalities. He has worked extensively with the OECD Programme for International Student Assessment (PISA) data.

1

The use (and misuse) of PISA in guiding policy reform: the case of Spain

Abstract

In 2013 Spain introduced a series of educational reforms explicitly inspired by the Programme for International Student Assessment (PISA) 2012 results. These reforms were mainly implemented in secondary education – based upon the assumption that this is where Spain's educational problems lie. This paper questions this assumption by attempting to identify the point where Spanish children fall behind young people in other developed countries. Specifically, by drawing data from multiple international assessments, we are able to explore how cross-national differences in reading skills change as children age. Consideration is given to both the average level of achievement and the evolution of educational inequalities. Our conclusion is that policymakers have focused their efforts on the wrong part of the education system; educational achievement is low in Spain (and educational inequalities large) long before children enter secondary school. This study therefore serves as a note of caution against simplistic interpretation of the PISA rankings.

Keywords: Educational policy; academic performance; PISA; PIRLS.

1. Introduction

Since its return to democracy, a change in the colour of Spain's governing party has generally meant a new set of educational reforms. The latest is the Organic Act for the Improvement of Quality in Education (LOMCE); approved by the conservative government shortly after the release of PISA 2012 results. These reforms have been designed to tackle what the Spanish Ministry of Education (2013a) believe are the key weaknesses of Spain's education system: high rates of school failure, early school dropout[1], the low status of vocational education, lack of external evaluations, low levels of school autonomy, and generally low academic performance of students. It is the last of these which is perhaps the ruling government's greatest concern. This is driven, at least in part, by Spain's continual poor performance in three major international assessments - the Trends in Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Survey (PIRLS) and, most notably, the Programme for International Student Assessment (PISA). Indeed, in 2012 Spanish students performed below the OECD average in each of the three core PISA domains (reading, maths and science).

The Spanish government has explicitly used poor performance in PISA to justify its educational reforms[2]. For instance, the preamble of the LOMCE legislation states how:

> *"The PISA 2009 report presents for Spain results which underline an insufficient level in reading comprehension, mathematics competency and scientific competency"*
> (Spain 2013: 97859).

While the Spanish Minister of Education (J.I. Wert[3]) argued while approving the act that:

> *"it is not our intention using PISA as a weapon or a self-flagellation instrument. An international comparative study as this one must be used for finding a path to*

---

[1] According to the Ministry of Education, Culture and Sports (2013b), during 2010/11, 33% of 16 year-old students had not completed compulsory education. Moreover, early school dropout stood around 25%. This was well above the 15% target, and higher than any other European Union (EU) country.

[2] The previous 2006 Education Act (LOE) included the following generic statement: "*Some recent international assessments have clearly revealed it is possible to combine quality and equity and should not be considered opposing objectives*".

[3] The LOMCE is popularly referred to in Spain as the "Wert Act".

improvement, and this is shown by countries like Poland that, <u>having the same</u>
<u>orientation as the LOMCE</u>, have significantly improved their results" (Europa Press
2013).

This last concern has surged -as it has happened in other countries (Bulle, 2011; Pons, 2011)- with the relatively modest achievement of Spanish students in international assessment programmes such as Trends in Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Survey (PIRLS) and, especially, the Programme for International Student Assessment (PISA). It is thus clear that the low performance of Spanish children in PISA has had a significant impact upon important policymakers in this country. Indeed, it is to their credit that they have taken the results of such assessments so seriously, and are passionate in their desire to introduce educational reforms. However, although international assessments can be a useful tool for comparative education purposes, a naïve use by educational policymakers can be misleading, as will be shown in this paper.

Specifically, it is our belief that the Spanish government have focused their reforms upon the wrong part of the education system, due to their simplistic interpretation of the PISA data. Specifically, the main components of the LOMCE reforms are:

1. Raising the level of autonomy of schools, increasing the importance of school principals.
2. Introducing external evaluations to students at the end of the primary (year 6) and lower secondary levels (year 10)[4]. These evaluations will be performed for providing information to families and schools.
3. Simplifying the curriculum, putting more weight on instrumental competencies, ICT and foreign languages.
4. Making tracks more flexible, avoiding dead-ends in the educational system. For meeting these ends, tracking between the academic and vocational paths is advanced by one year (from age 16 to age 15).

The vast majority of the above are focused upon changes to lower-secondary education. But is this really when Spain's educational problems emerge? Or are low levels of

---

[4] Compulsory education in Spain begins at age 6 and comprises six years of primary education and four years of lower secondary education. Nevertheless, school enrolment rates at age 3 are over 95%.

academic achievement, and large educational inequalities, already apparent much earlier in young people's lives?

Unfortunately, despite the changes already underway in Spain, there is actually very little robust evidence on this important issue. This study therefore aims to fill this gap in the literature by investigating how Spanish children's reading skills develop over time (between the ages of 10 and 16) relative to children in a selection of other countries. Specifically, we address the following three research questions:

1. At what point in the schooling system does Spain fall behind other countries in terms of average reading achievement? Do other countries improve relative to Spain in secondary school, or is the achievement gap already stark by the end of primary school and then simply maintained? How does the gender gap evolve during this period?

2. How does the *distribution* of academic achievement change in Spain between the end of primary school (age 9/10) and the end of secondary school (age 15/16)? Do educational inequalities grow, shrink or remain the same?

3. Is the socio-economic (SES) gradient in children's reading skills large or small in Spain relative to other developed countries? Do these inequalities grow or narrow during secondary school, and does Spain differ significantly to other countries in this respect?

Although this exercise should have been conducted *before* the approval of the 2013 education act, this ex-post analysis will nevertheless reveal how well founded the 'Wert Act' educational reforms are. It therefore provides an illustration of how *not* to use international assessments (in this case PISA) in designing changes to national education systems[5] and, at the same time, how comparative education approaches can be useful for implementing country-level reforms.

The paper is now structured as follows. Section 2 describes the PIRLS and PISA databases and our empirical methodology. Section 3 presents results, focusing upon how Spain's relative performance on important international reading tests changes

---

[5] That is, using Bieber and Martens' (2011) terminology, we will assess a real case of the role of PISA as a 'Soft Power' in education.

between the end of primary school and the end of secondary school. Conclusions and policy discussion follows in section 4.

2. Methodology and databases

The aim of this study is to investigate Spain's relative performance in international reading tests at ages 10 and 16. Ideally, longitudinal data would be available to track children's progress over time. Unfortunately, such data is not collected in Spain, nor in several other important comparator countries. Consequently, we follow an alternative strategy pursued by Goodman et al (2009) and Jerrim and Choi (2014). Specifically, we treat PIRLS 2006 and PISA 2012 as repeated cross-sectional data, with children aged 9/10 (4th year of primary school) in the former and 15/16 (3rd or 4th year of compulsory secondary education) in the latter. To maximise comparability, we retain only those countries that participated in both the PIRLS 2006 and PISA 2012 studies. Moreover, we only retain children born in either 1996 or 1997[6]. This leaves a total of 25 education systems[7] for whom we investigate change in relative reading test scores as children age[8].

   Although PIRLS and PISA both collect nationally representative samples, with similar survey designs and response rates[9] (see Mullis et al. 2007 and OECD 2011 for further information), raw test scores *cannot* be directly compared across the two surveys. First, the two surveys use different item-response theory models to scale the test score data (see Brown et al 2007). Second, there are some subtle conceptual differences in the skills the two tests measure, with PIRLS focused upon "curriculum-based" measures of literacy, while PISA measures children's ability to use their skills in "real-life" situations. Finally, the two studies contain different sets of countries (e.g. 41

---

[6] Any country where more than half the sample was born outside these years has also been excluded from our analysis. Sensitivity analyses using a lower threshold (25 percent) has also performed, with the main conclusions unaltered (results available upon request).

[7] We will refer to these education systems, throughout the article, as countries. The reader should however bear in mind that, among the units compared, there are also smaller administrative units such as, for example, five Canadian provinces.

[8] In PIRLS 2006, Iceland and Norway assessed their year 5 students too. However, in order to keep comparability with the rest of countries, we work with their year 4 pupils. Given the decentralized nature of the Spanish educational system, an analysis by Autonomous Communities would have been relevant. However, the information provided by PIRLS does not allow to perform analyses for Spain at the regional level.

[9] PIRLS 2006 and PISA 2012 response rates after replacement are available in Martin et al. (2007, 126) and OECD (2014, 271) respectively.

countries participated in PIRLS 2006 compared to 65 in PISA 2012) with test scores then scaled to a mean of 500 and standard deviation of 100 *within* each of the respective surveys. Consequently, a score of 500 in PIRLS is *not* equivalent to a score of 500 in PISA.

We deal with this issue by converting all test score data into international z-scores, following the lead of Brown et al (2007). In other words, we normalize test scores for each survey at the student level, resulting in a mean of 0 and a standard deviation of 1 across all 25 countries included in our sample. This has important implications regarding interpretation of results. Specifically, we are unable to comment upon how children's reading test scores change in Spain as children age in *absolute* terms. Rather we can only consider <u>*relative*</u> differences between Spain and other countries, and how this <u>*relative*</u> difference changes between the end of primary (PIRLS 2006) and end of secondary (PISA 2012) school. It is important for readers to bare this in mind when interpreting our results.

Our analysis begins by considering how *average* reading test scores (converted into the z-score metric) compares across countries at ages 9/10 and 15/16. This is followed by a consideration of how the distribution of children's reading scores changes as children age. We then turn to the issue of socio-economic inequalities, estimated using the following Ordinary Least Squares (OLS) regression model[10]:

$$A_{ijk} = \alpha + \beta_1 Sex_i + \beta_2 SES_i + \beta_3 I_i + \beta_4 SES_i * I_i + \varepsilon_{ij} \quad \forall k \qquad (1)$$

Where:

A_{ijk}= Performance on the PIRLS or PISA reading test (in terms of z-scores);

Sex_i = Pupil gender (0= boys, 1= girls);

I_i = Immigrant status (0= native, 1= immigrant);

SES_i = A set of dummy variables reflecting parental occupation

 i = Pupil i

j = School j

$\forall k$ = The model is estimated seperately for each K country[11]

---

[10] This specification follows Schütz et al. (2008); Wößmann (2008) or Jerrim and Choi (2014).
[11] We use the first plausible value only, both in PIRLS and PISA, throughout the analysis. As OECD (2010, 129) notes, 'analyzing one plausible value instead of five plausible values provides unbiased population estimates'.

The parameter of interest from (1) is β_2; the association between children's socio-economic background and performance on the reading test.

We estimate model (1) twice; once using father's occupation to measure SES (divided into four groups: elementary, semi-skilled blue collar, semi-skilled white collar and skilled white collar workers) and once using the number of books at home (Wößmann 2008, Evans et al. 2010, Hanushek and Wößmann 2011, Jerrim and Choi 2014)[12]. Both of the above have strengths and limitations. Although father's occupation is a widely accepted measure of SES in sociological research, and is reliably reported in international surveys (Jerrim and Micklewright 2014), such information is missing for up to half the sample in PIRLS 2006 for some countries. In contrast, missing data for books in the home is low (less than five percent in most countries), and is a frequently used proxy for SES in international comparative research (see Schütz, Ursprung and Wöessmann 2008). Concerns have been raised, however, regarding accuracy of measurement and whether the number of books is really a robust measure of social stratification (Jerrim 2012; Jerrim and Micklewright 2014).

This difficulty will be handled as follows. First, we estimate model 1 using father's occupation, with multiple imputation by chained equations used to account for missing data (in terms of observable characteristics)[13]. Then model 1 is re-estimated, but using books in the home to measure SES rather than father's occupation. Our interest is whether the same broad pattern of results holds whichever family background measure is used. For instance, do we consistently find that socio-economic inequality in reading achievement is greater in Spain than other countries? And is there consistent evidence that the SES gradient grows, shrinks or stays the same in Spain as children move from the end of primary school to the end of secondary school?

The clustering of pupils within school is accounted for throughout the analysis by either Huber-White adjustments, bootstrapping by cluster (using 50 replications) or application of the Jackknife (PIRLS) or Balanced-Repeated Replication (PISA) weights. Final student senate weights are also applied to correct estimates for non-response and to scale national samples up to population estimates. Standard errors for differences between countries and between surveys are calculated using a two sample t-test assuming independence between samples.

---

[12] Jerrim and Choi (2014) provide an extensive review of analyses which have used this variable with international assessments.

[13] Precise details on the imputation model used is available from the authors upon request. We have also conducted a 'complete case' analysis, with found little substantive difference to the results presented.

3. Results

3.1. *Average reading scores*

Cross-country differences in average reading test scores (converted into the z-score metric) are presented in Table 1. The first point of note is that, at both age 9/10 and 15/16, Spain falls below the international median. Specifically, in both surveys, it is ranked 19th out of the 25 countries included. Moreover, there is little change in the average z-score for Spain between the two studies; it stands at -0.071 standard deviation at age 9/10 and -0.079 at age 15/16[14]. This highlights two important points. First, even by age 9/10, Spanish children's reading proficiency is behind that of most other countries included in our analysis. For instance, average reading achievement in Spain is already 0.34 standard deviations lower than in the United States, 0.41 standard deviation lower than in Italy and more than half a standard deviation behind the top performer (Hong Kong). Second, there is little evidence that the gap in *relative* performance between Spain and other countries either shrinks or grows during secondary school. On the one hand, this suggests that Spanish secondary schools are unable to compensate for the comparatively poor reading skills children have developed during their first ten years of life. On the other, it is clearly *not* during secondary school where Spain's educational problems start to emerge. This finding has important policy (and political) implications – the "blame" for Spain's poor performance in PISA should not be directed at secondary schools. Rather Spain's educational problems seem to emerge much earlier in children's lives, which the secondary education system then struggles to reverse.

[INSERT TABLE 1]

This point is further emphasized in the last column of Table 1, which illustrates the change in average z-scores between ages 9/10 and 15/16 across the selected countries. In total, seven jurisdictions saw significantly more improvement than in Spain, including Norway, Poland and Taiwan. This was balanced out by six countries significantly declining relative to Spain, including several major OECD countries such as Austria, Italy, the Netherlands and the United States. Hence one can actually make a

---

[14] Note that two very low performing countries (Indonesia and Qatar) are included in the analysis. This explains why the average score for Spain is close to zero, despite being well behind most other OECD countries.

case for Spain's secondary schools being *superior* to those in several other European and North American countries (in that children make, on average, more progress). This serves as a valuable lesson to policymakers (particularly those in Spain) – disappointing performance in PISA does not necessary mean that secondary schools are 'failing' or that this part of the education system is the root-cause of a country's educational problems.

Table 1 also highlights some other interesting findings. Notably, countries performing well above the international average at the end of primary school generally managed to maintain their strong performance to the end of secondary school (Italy and Slovakia are notable exceptions). The same is also true at the other extreme, with countries performing poorly at primary school also tending to perform poorly at secondary school. Norway and Poland are two examples of low performing countries at age 9/10 which have improved significantly by ages 15/16. Their experiences may be particularly relevant for understanding features of secondary school systems that enable children to make strong progress (though some caution is required here, due to the possibility of statistical artifacts such as 'regression to the mean' – see Jerrim and Vignoles 2013). Nevertheless, these results seem to stress the importance of the early stages of education and the difficulty of overcoming large initial achievement gaps. In other words, once a country falls behind in the educational achievement race, it is difficult to then catch up. This should be particularly worrying for policymakers in Spain, given both this country's poor performance in PIRLS, and the fact that the 2013 LOMCE educational act introduced very few changes at the primary and pre-primary school levels. We believe this to be a grave mistake, driven by policymakers' naive use of the international educational achievement rankings.

We conclude this subsection by analysing differences in progress by gender. Previous research has consistently shown that, in almost every OECD country, girls outperform boys in international reading assessments (OECD 2010:16). The unique contribution of Table 2 is in considering whether the gender gap in relative reading test scores shrinks or grows during secondary education, and how this varies across countries. Interestingly, in almost every economy the 'change' coefficient is positive – not only do girls outperform boys in the international reading assessments – they also make significantly more progress during secondary school. Moreover, in most countries this cannot simply be attributed to sampling variation – the change is statistically significant in 21 out of the 25 countries considered (the exceptions are England,

Scotland, Indonesia and Nova Scotia). This includes Spain, where the gender gap increases from 0.03 standard deviations at age 9/10 to 0.29 standard deviations at ages 15/16. This is an important finding; it suggests that it is indeed during secondary school where the gender gap in reading skills in Spain seems to emerge. Hence, to the extent that Spanish policymakers should be looking at policy reforms to the secondary education system, it would seem one of the most fruitful targets may be to reduce the gender gap in reading achievement – by making sure the reading skills of boys keeps pace with their female peers.

[INSERT TABLE 2]

3.2. Inequality in educational outcomes

We now turn to inequality in children's educational outcomes, along with change in reading performance of the highest and lowest achievers. To begin, the standard deviation of children's test scores is presented as the preferred measure of educational inequality[15]. Results can be found in Figure 1. The length of the bars illustrate the standard deviation at age 15/16, with triangles providing analogous figures at age 9/10. The most unequal countries at the end of primary school are Israel, Qatar, England and Scotland, with greatest equality found in the Netherlands, Flemish-Belgium and Hong Kong. The standard deviation for Spain at age 10 (0.813) is around the international average, with educational inequalities neither standing out as particularly large or small. There is a modest increase of 0.086 standard deviations in educational test scores in Spain between ages 9/10 and 15/16. Yet similar increases are observed in other countries. Consequently, educational inequality in Spain remains around the international average even at the end of secondary school. Thus neither the magnitude nor the change in educational inequality stands out as particularly pronounced in Spain relative to other countries.

[INSERT FIGURE 1]

---

[15] See Ferreira and Gignoux (2014) for a discussion on educational inequality measures and the validity of the standard deviation.

To gain further insight into this issue, Tables 3 and 4 consider change in the 10th (P10) and the 90th (P90) percentile of the reading test distribution across the two studies. The former can be interpreted as the performance of the *lowest* achievers in a country, while the latter refers to the *highest* achievers. Unsurprisingly, countries that saw an increase in mean performance also tended to see an increase in P10 and P90. As Figure 2 shows, there was a modest but statistically significant increase in the 90th percentile in Spain between ages 9/10 and 15/16 (from 0.92 to 1.03), while the opposite holds true for the 10th percentile (from -1.12 to -1.26). This is an important finding – it suggests that already high achieving Spanish children saw a relative improvement in their reading scores (compared to children in other countries) while low achieving children in primary school fall further behind[16]. Consequently, if action is to be taken in Spanish secondary education, it should be targeted at the country's lowest performing schools and pupils.

[INSERT TABLE 3]

[INSERT TABLE 4]

[INSERT FIGURE 2]

3.3. Inequality of educational opportunity

To conclude, we turn to socio-economic differences in educational achievement. Table 5 measures the socio-economic gradient as differences in test scores between children whose father works in a skilled white collar occupation versus those whose father works in an elementary occupation. The robustness of these results are considered in Table A1 (Appendix), where the socio-economic gradient is alternatively measured as the differences in test scores between children living in homes with more than 200 books versus those with 25 books or less (as noted in section 2, books in the home is a frequently used proxy for socio-economic status in cross-national research).

[INSERT TABLE 5]

---

[16] As shown in Table 1, these two effects largely cancel one another out, meaning there was little change in mean scores for Spain between 9/10 and 15/16.

Results in Table 5 illustrate there exists a sizeable socio-economic gradient in Spanish children's reading skills at ages 9/10 (0.59 standard deviation points). There is a slight reduction of this gap by age 15/16 (to 0.48 standard deviations) but this change does not reach statistical significance at conventional thresholds. A similar finding holds across most of the selected countries, with a significant increased observed in only three (Netherlands, Flemish Belgium and Taiwan) and a decrease in just one (Scotland). These results therefore strongly suggest that inequality of educational opportunity in Spain is largely generated before the age of 9/10.

However, some caution is needed here, as our analysis using books in the home produces a somewhat different result (see Appendix 1). In particular, in most countries a significant increase in the impact of this SES measure is observed, including in Spain. In particular, the difference in test scores between the lowest (less than 25 books) and highest (more than 200 books) socio-economic groups increases from 0.63 (age 9/10) to 0.94 (age 9/10) standard deviations. This is of broadly similar magnitude to the increase observed in most other countries.

What do we therefore conclude from these results? First, there seems robust evidence that SES inequality in Spain does not appreciably decline between the end of primary and secondary school. Rather, inequalities in educational opportunities are either maintained or increased – with somewhat different results depending upon which SES measure one chooses to use. Secondly, both Table 5 and Appendix 1 suggest SES inequalities in Spain do not seem to change by any more or less than is observed in most other countries. Finally in Spain, as in many other countries, socio-economic differences in educational attainment are large – and require urgent policy action to be reduced.

Despite the LOMCE reforms noting the importance of this last point, few details are provided on how such a reduction in SES achievement gradients might be achieve. We believe that our evidence suggests Spanish policymakers should target their interventions early in young people's lives (i.e. before secondary school). In particular, both Table 5 and Appendix 1 illustrate how, once SES inequalities in educational attainment emerge, they are very difficult to reverse.

4. Discussion and conclusions

Reducing school failure and increasing the 'quality' of education were among the main objectives of Spain's latest educational reforms. The Ministry of Education has acknowledged these reforms were inspired by Spain's poor performance in international assessments, and the subsequent recommendations for improvement made by international organizations. The aim of this article was to scrutinize Spain's performance in these educational assessments in more detail, in order to provide a more nuanced view of this country's educational problems. Our focus has been whether Spain's disappointing performance in important international reading assessments really emerges during secondary education, or if it already lags behind other countries towards the end of primary school. We not only considered performance on average, but also changes in the distribution of reading achievement and the evolution of educational inequalities between ages 9/10 and 15/16. Our four key findings can be summarized as follows.

First, the gap in *average* reading test scores between Spain and other countries is just as stark at age 9/10 as it is at age 15/16. In other words, Spain's poor performance on international reading assessments seems to be generated in primary (and pre-primary) education, and does not appreciable decline (or improve) during secondary school. This is consistent with the work of Mena et al. (2010), who describe how low primary school performance can harm children's educational expectations, self-concept and engagement in school – with slow progress and early school dropout the result. Thus improving the poor reading skills of *primary school* children seems to be critical if Spain is to significantly improve its position in the PISA achievement rankings.

Second, although there is little change in mean reading test scores between ages 9/10 and 15/16, this masks some interesting changes to the *distribution* of reading achievement. In particular, whereas the reading skills of Spain's lowest achieving children declines (relative to other countries) during secondary school, the reading skills of its top performers actually improves. In other words, there is a small increase in educational inequality, with the least able children falling further behind the average and the more able moving further ahead. This has important implications for Spanish policymakers; improving basic skills amongst the country's lowest performing pupils – in both primary and secondary school – may be an effective way to simultaneously reduce educational inequality while improving average levels of achievement.

Third, our results have highlighted the socio-economic differences in educational achievement that exist in the Spanish educational system. Such inequality is established early in young people's lives, and then either maintained or exacerbated during secondary education. Consequently, our evidence suggests that once social inequalities in educational attainment have emerged, they become very difficult to reverse. This again points towards early action, long before children reach secondary school.

Finally, we provide empirical evidence on the usefulness for policymaking of some of the existing international assessments. Indeed, we show it is precisely the comparative nature of PIRLS and PISA that enables us to provide guidelines at the national-level. However, we reach this conclusion having taken the Spanish case as our starting point, Spain being a case of misuse of international assessments for policymaking. This strategy has allowed us to draw out the limitations and risks of simplistic approaches to cross-national studies such as PISA.

One must of course recognize the limitations of this paper and to stress the need for further work. Ideally, this study would have been conducted using longitudinal data, following exactly the same group of pupils over time. Unfortunately, cross-nationally comparable data of this nature does not yet exist, leading us to take the alternative 'repeated cross-section' approach instead. Nevertheless, this study has illustrated one of many interesting questions such data could address, and highlighted the need for international assessment like PISA to begin to track the *progress* of children over time. Second, our results are based upon observing young people at two time points - age 9/10 and 15/16. This limits our ability to identify the exact point when Spanish children fall behind their peers in other countries (in terms of their reading skills). For instance, we do not know how Spain compares to other countries at the approximate point of school entry (e.g. age 5/6), and thus whether educational problems actually emerge in this country even before compulsory schooling has begun. Finally, the focus of this study has been children's reading skills. We are unable to comment upon whether similar patterns are likely to hold for other cognitive (or indeed non-cognitive) domains, including science and mathematics. For example, Spain only started to participate in the Trends in Mathematics and Science Study (TIMSS) in 2011[17], meaning an investigation of children's performance in these domains over time is not currently possible.

---

[17] A Spanish region, the Basque Country, participated in previous TIMSS waves as a benchmarking participant.

Nevertheless, this may be a fruitful direction for future research once further data become available (e.g. results from PISA 2015).

Despite these limitations, we believe this paper has the potential to make an important contribution to contemporary education policy in Spain. Despite not being clear from international achievement rankings such as PISA, Spain's major educational problems emerge long before children enter secondary school. Yet, due to their naive interpretation of such rankings, Spain's politicians have nevertheless decided to concentrate the recent LOMCE reforms at the secondary education level. Although analyzing the impact of these reforms is beyond the scope of this paper, we believe that they have been designed and developed on a rocky foundation. Indeed, despite containing a number of well-meaning and potentially sensible measures, we believe the LOMCE reforms are unlikely to get to the heart of Spain's under-achievement – which occurs much earlier in the schooling system. Much more emphasis should have been given to primary and pre-school education when these reforms were being designed. As such, our study uncovers the paradox of LOMCE; international assessments such as PISA have been used to justify their existence –a clear case of the so-called 'tyranny of numbers' (Ball, 2015)-, yet the measures being introduced would have benefitted immensely from a more nuanced approach to their use. This study therefore acts as an important warning to policymakers from other countries. International assessments like PISA may have some role to play in directing education reforms and encouraging policy change. Yet their naïve use (and misuse) by policymakers may lead to a waste of resources, with sub-optimal changes to the education system being made.

5. References

Ball, Stephen. J. 2015. "Education, governance and the tyranny of numbers." *Journal of Education Policy* 30 (3):299-301. doi:10.1080/02680939.2015.1013271.

Bieber, Tonia, and Kerstin Martens. 2011. "The OECD PISA Study as a Soft Power in Education? Lessons from Switzerland and the US." *European Journal of Education* 46 (1): 101–116.

Brown, Giorgina, Micklewright, John, Schnepf, Sylke V., and Robert Waldmann. 2007. "International Surveys of Educational Achievement: How Robust are the Findings?" *Journal of the Royal Statistical Society Series A* 170 (3): 623–646. doi: 10.1111/j.1467-985X.2006.00439.x.

Bulle, Nathalie. 2011. "Comparing OECD Educational Models through the Prism of PISA." *Comparative Education* 47 (4): 503–521.

Bushnik, Tracey, T., Telford, Lynn-Barr and Patrick Bussière. 2004. *In and out of high school: First results from the second cycle of the Youth in Transition Survey, 2002*. Ottawa: Statistics Canada, Culture, Tourism and the Centre for Education Statistics.

Europa Press. 2013. "Wert is convinced the LOMCE will improve PISA results". Europa Press, December 11. [In Spanish]. http://www.europapress.es/sociedad/educacion/noticia-wert-convencido-lomcecontribuira-mejorar-resultados-espana-pisa-20131211110703.html

Evans, Mariah D. R., Kelley, Jonathan., Sikora, Joanna, and Donald J. Treiman. 2010. "Family Scholarly Culture and Educational Success: Books and Schooling in 27 Nations." *Research in Social Stratification and Mobility* 28 (2): 171–197.

Ferreira, Francisco H. G., and Jérémie Gignoux. 2014. "The Measurement of Educational Inequality: Achievement and Opportunity." *The World Bank Economic Review* 28 (2): 210-246. doi: 10.1596/1813-9450-5873.

Goodman, Alissa, Sibieta, Luke, and Elizabeth Washbrook. 2009. "Inequalities in educational outcomes among children aged 3 to 16". *Final report for the National Equality Panel*. London: Institute for Fiscal Studies.

Hanushek, Eric A., and Ludger Wößmann. 2011. "The Economics of International Differences in Educational Achievement". In *Handbook of Economics of Education* vol.3: 89-200, ed., Eric. A Hanushek, Stephen Machin and Ludger Wößmann, L. St.Louis (MO): Elsevier.

Jerrim, John. 2012. "The socio-economic gradient in teenagers' literacy skills: how does England compare to other countries?" *Fiscal Studies* 33(2):159–84.

Jerrim, John. 2013. "The reliability of trends over time in international education test scores: is the performance of England's secondary school pupils really in relative decline?" *Journal of Social Policy* 42(2): 259–279.

Jerrim, John and Álvaro Choi. 2014. "The mathematics skills of school children: how does England compare to the high-performing East Asian Jurisdictions?" *Journal of Education Policy* 29 (3), 349-376.

Jerrim, John, and John Micklewright. 2014. "Socioeconomic Gradients in Children's Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background?" *European Sociological Review* 30 (6): 766-781. doi: 10.1093/esr/jcu072 .

Jerrim, John and Anna Vignoles. 2013. "Social mobility, regression to the mean and the cognitive development of high ability children from disadvantaged homes". *Journal of the Royal Statistical Society: Series A* 176 (4): 887-906. DOI: 10.1111/j.1467-985X.2012.01072.x.

Martin, Michael O., Mullis, Ina V. S., and Ann M. Kennedy. 2007. *PIRLS 2006 Technical Report*. Boston MA: Lynch School of Education, Boston College.

Mena, Luis, Fernández Enguita, Mariano, and Jaime Riviére Gómez. 2010. "Disengaged from education: Processes, experiences, motivations and strategies of early school dropout and school failure." *Revista de Educación* número extraordinario 2010: 119-145.

Ministry of Education, Culture and Sports. 2013a. *LOMCE: Proposals for improving the educational quality*. Madrid: Ministry of Education, Culture and Sports. [In Spanish].

Ministry of Education, Culture and Sports. 2013b. *Education in figures. Academic year 2010-2011 (2013 Edition)*. Madrid: Ministry of Education, Culture and Sports. [In Spanish].

Mullis, Ina V. S., Martin, Michael O., Kennedy, Ann M., and P. Foy. 2007. *PIRLS 2006 International Report*. Boston MA: Lynch School of Education, Boston College.

OECD. 2010. *PISA 2009 at a Glance.* Paris: OECD.

OECD. 2011. *PISA 2009 Technical Report*. Paris: OECD.

OECD. 2013. *PISA 2012 Results: Excellence through Equity (vol.II).* Paris: OECD.

OECD. 2014. *PISA 2012 Results: What Students Know and Can Do – Student's Skills in Teaching Real-Life Problems (Volume V)*. Paris: OECD.

Pons, Xavier. 2011. "What Do We Really Learn from PISA? The Sociology of its Reception in Three European Countries (2001–2008)." *European Journal of Education* 46 (4): 540–548.

Schütz, Gabriela, Heinrich W. Ursprung, and Ludger Wößmann. 2008. "Education Policy and Equality of Opportunity." *Kyklos* 61 (2): 279–308.

Spain. Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educativa. *Boletín Oficial del Estado*, 10th December 2013, 295 (I): 97858-97921.

Wößmann, Ludger. 2008. "How Equal are Educational Opportunities? Family Background and Student Achievement in Europe and the United States." *Zeitschrift für Betriebswirtschaft* 78 (1): 45–70.

Appendix 1

Table A1. Socio-economic differences in the reading competency between ages 9/10 and 15/16 (international Z-scores): books at home

	Age 9/10		Age 15/16		Change	
	SES	SE	SES	SE	Change	SE
France	0.64*	0.04	1.18*	0.07	0.54*	0.08
Netherlands	0.36*	0.04	0.89*	0.06	0.53*	0.07
Belgium (French)	0.53*	0.14	1.05*	0.08	0.52*	0.16
Slovak Republic	0.76*	0.06	1.25*	0.09	0.49*	0.11
Canada (Alberta)	0.51*	0.06	0.90*	0.08	0.39*	0.09
Belgium (Flemish)	0.40*	0.04	0.76*	0.06	0.36*	0.07
United States	0.47*	0.07	0.82*	0.07	0.35*	0.10
Austria	0.64*	0.05	0.98*	0.06	0.33*	0.07
Italy	0.53*	0.06	0.86*	0.03	0.33*	0.07
Spain	0.63*	0.13	0.94*	0.03	0.32*	0.14
Norway	0.53*	0.07	0.84*	0.05	0.32*	0.08
Canada (Brit. Columbia)	0.62*	0.07	0.93*	0.08	0.31*	0.10
Canada (Ontario)	0.44*	0.09	0.74*	0.07	0.30*	0.11
England	0.82*	0.07	1.11*	0.06	0.28*	0.09
Canada (Nova Scotia)	0.57*	0.05	0.85*	0.06	0.28*	0.08
Iceland	0.60*	0.05	0.84*	0.06	0.24*	0.07
Canada (Quebec)	0.60*	0.06	0.82*	0.07	0.22*	0.09
Hong Kong	0.41*	0.05	0.63*	0.07	0.21*	0.08
Taiwan	0.65*	0.05	0.85*	0.05	0.20*	0.07
Slovenia	0.69*	0.04	0.86*	0.07	0.17*	0.08
Poland	0.79*	0.05	0.87*	0.05	0.08	0.07
Scotland	0.77*	0.06	0.86*	0.06	0.08	0.09
Israel	0.73*	0.10	0.72*	0.09	-0.01	0.13
Qatar	0.36*	0.05	0.27*	0.05	-0.09	0.07
Indonesia	0.41*	0.18	0.29*	0.10	-0.12	0.20

Source: Authors' calculations using the PIRLS 2006 and PISA 2012 microdata.

Note: * = Significantly different from zero at the five percent level. Senate weights applied.

Table 1. Average test scores in reading competency between ages 9/10 and 15/16 (international Z-scores).

	Age 9/10		Age 15/16		From 9/10 to 15/16	
	Mean	**SE**	**Mean**	**SE**	**Change**	**SE**
Qatar	-2.00*	0.05	-1.06*	0.08	0.93*	0.09
Norway	-0.30*	0.03	0.07*	0.03	0.38*	0.04
Indonesia	-1.25*	0.05	-0.98*	0.05	0.28*	0.07
Poland	-0.04	0.03	0.22*	0.03	0.25*	0.04
Taiwan	0.08*	0.03	0.26*	0.05	0.18*	0.05
Belgium (French)	-0.16*	0.03	0.01	0.07	0.17*	0.08
Canada (Quebec)	0.12*	0.03	0.24*	0.04	0.12*	0.05
Scotland	0.05*	0.04	0.10*	0.04	0.05	0.05
Israel	-0.15	0.08	-0.11	0.06	0.04	0.10
Iceland	-0.15*	0.02	-0.13	0.04	0.02	0.05
France	0.10*	0.03	0.09*	0.06	-0.01	0.06
Hong Kong	0.49*	0.03	0.48*	0.05	-0.01	0.05
Spain	-0.07	0.03	-0.08	0.02	-0.01	0.04
Canada (Brit. Columbia)	0.42*	0.03	0.38*	0.05	-0.04	0.05
Canada (Ontario)	0.39*	0.03	0.31*	0.04	-0.07	0.05
England	0.12*	0.05	0.04*	0.05	-0.08	0.07
Canada (Nova Scotia)	0.21*	0.03	0.12*	0.04	-0.10	0.05
Slovenia	-0.02	0.03	-0.16	0.05	-0.13	0.06
Belgium (Flemish)	0.36*	0.02	0.21*	0.06	-0.14	0.06
Canada (Alberta)	0.47*	0.03	0.29*	0.04	-0.17*	0.05
Netherlands	0.36*	0.02	0.14*	0.06	-0.22*	0.06
United States	0.27*	0.04	0.01	0.04	-0.26*	0.06
Austria	0.24*	0.03	-0.07	0.05	-0.31*	0.06
Italy	0.34*	0.04	-0.07	0.03	-0.40*	0.05
Slovak Republic	0.15*	0.03	-0.32*	0.06	-0.47*	0.07

Source: Authors' calculations using the PIRLS 2006 and PISA 2012 microdata.

Note: * = significantly different from Spain at the five percent level. Senate weights applied.

Table 2. Average test scores gender gap in reading competency between ages 9/10 and 15/16 (international Z-scores).

	Age 9/10		Age 15/16		Change 15/16 – 9/10	
	Gap	SE	Gap	SE	Difference	SE
France	0.098*	0.028	0.440*	0.042	0.342*	0.049
Slovenia	0.230*	0.027	0.563*	0.046	0.333*	0.055
Belgium (French)	0.016	0.031	0.346*	0.053	0.330*	0.063
Italy	0.096*	0.028	0.387*	0.028	0.291*	0.040
Iceland	0.227*	0.030	0.502*	0.038	0.275*	0.051
Qatar	0.425*	0.087	0.691*	0.091	0.266*	0.121
Spain	0.034	0.029	0.289*	0.019	0.255*	0.031
Austria	0.116*	0.031	0.358*	0.057	0.242*	0.063
Norway	0.207*	0.030	0.439*	0.031	0.232*	0.045
Belgium (Flemish)	0.065*	0.026	0.284*	0.057	0.219*	0.057
United States	0.084*	0.035	0.298*	0.030	0.214*	0.047
Poland	0.204*	0.033	0.412*	0.031	0.208*	0.042
Israel	0.228*	0.048	0.431*	0.079	0.203*	0.091
Slovak Republic	0.181*	0.035	0.379*	0.043	0.198*	0.055
Netherlands	0.061	0.032	0.258*	0.032	0.197*	0.046
Canada (Ontario)	0.177*	0.042	0.372*	0.038	0.195*	0.056
Canada (Alberta)	0.102*	0.026	0.294*	0.038	0.192*	0.045
Canada (Quebec)	0.178*	0.038	0.347*	0.038	0.169*	0.054
Hong Kong	0.097*	0.024	0.247*	0.044	0.150*	0.050
Taiwan	0.170*	0.023	0.319*	0.053	0.149*	0.055
Canada (Nova Scotia)	0.259*	0.035	0.400*	0.068	0.141	0.076
Canada (Brit. Columbia)	0.123*	0.031	0.259*	0.046	0.136*	0.056
Indonesia	0.190*	0.042	0.268*	0.030	0.078	0.050
England	0.253*	0.040	0.241*	0.052	-0.012	0.067
Scotland	0.275*	0.040	0.261*	0.036	-0.014	0.054

Source: Authors' calculations using the PIRLS and PISA dataset.

Note: Senate weights applied. * = statistically significant from 0 at the five percent level. As Mullis et al. (2007:48, Exhibit 1.4) report, Spain was one of the only two countries participating in PIRLS-2006 where gender differences were statistically insignificant.

Table 3. Average test scores for the 10th percentile in the reading competency between ages 9/10 and 15/16 (international Z-scores).

	Age 9/10		Age 15/16		From 9/10 to 15/16	
	P10	SE	P10	SE	Difference	SE
Qatar	-3.48*	0.07	-2.48*	0.09	1.00*	0.10
Indonesia	-2.52*	0.07	-1.93*	0.06	0.60*	0.10
Poland	-1.24	0.06	-0.89*	0.05	0.36*	0.06
Scotland	-1.25	0.08	-1.02*	0.05	0.24*	0.10
Israel	-1.87*	0.12	-1.64*	0.09	0.23	0.21
Norway	-1.36*	0.05	-1.18	0.06	0.18*	0.06
England	-1.28	0.08	-1.24	0.07	0.04	0.10
Taiwan	-0.92*	0.04	-0.96*	0.06	-0.04	0.07
Canada (Brit. Columbia)	-0.67*	0.05	-0.76*	0.07	-0.09	0.08
Canada (Nova Scotia)	-0.95*	0.05	-1.06*	0.08	-0.11	0.10
Canada (Quebec)	-0.86*	0.04	-0.99*	0.06	-0.13	0.09
Spain	-1.12	0.04	-1.26	0.03	-0.14	0.06
Iceland	-1.28*	0.04	-1.44*	0.05	-0.17	0.07
Canada (Ontario)	-0.69*	0.05	-0.86*	0.05	-0.17	0.08
Slovenia	-1.16	0.03	-1.34	0.05	-0.18	0.06
Belgium (French)	-1.19	0.04	-1.40	0.12	-0.21	0.12
Hong Kong	-0.42*	0.05	-0.66*	0.09	-0.25	0.08
Canada (Alberta)	-0.58*	0.04	-0.88*	0.08	-0.30	0.09
United States	-0.83*	0.05	-1.17	0.05	-0.33	0.07
France	-0.86*	0.04	-1.37	0.07	-0.51*	0.08
Austria	-0.73*	0.04	-1.27	0.07	-0.54*	0.08
Italy	-0.74*	0.04	-1.35	0.04	-0.61*	0.06
Belgium (Flemish)	-0.44*	0.03	-1.08	0.10	-0.64*	0.09
Netherlands	-0.42*	0.04	-1.06*	0.08	-0.64*	0.06
Slovak Republic	-0.93*	0.06	-1.71*	0.08	-0.79*	0.11

Source: Authors' calculations using the PIRLS 2006 and PISA 2012 microdata.

Note: * = significantly different from Spain at the five percent level. Senate weights applied.

Table 4. Average test scores for the 90th percentile in the reading competency between ages 9/10 and 15/16 (international Z-scores).

	Age 9/10		Age 15/16		Change 15/16 – 9/10	
	P90	**SE**	**P90**	**SE**	**Difference**	**SE**
Qatar	-0.54*	0.06	0.38*	0.08	0.91*	0.09
Norway	0.65*	0.03	1.29*	0.04	0.64*	0.04
Belgium (French)	0.81*	0.03	1.25*	0.07	0.43*	0.07
France	1.02*	0.04	1.42*	0.06	0.40*	0.06
Taiwan	1.02*	0.02	1.34*	0.04	0.32*	0.05
Canada (Quebec)	1.07*	0.04	1.38*	0.04	0.31*	0.06
Belgium (Flemish)	1.13*	0.02	1.37*	0.05	0.24*	0.04
Poland	1.06*	0.03	1.28*	0.05	0.22	0.05
Iceland	0.84*	0.02	1.04	0.03	0.20	0.04
Hong Kong	1.34*	0.03	1.47*	0.05	0.13	0.04
Netherlands	1.13*	0.02	1.26*	0.04	0.13	0.05
Spain	0.92	0.02	1.03	0.02	0.12	0.04
Indonesia	-0.09*	0.08	-0.04*	0.06	0.05	0.08
Canada (Ontario)	1.42*	0.04	1.44*	0.03	0.02	0.06
Israel	1.25*	0.04	1.26*	0.04	0.01	0.06
Canada (Brit. Columbia)	1.45*	0.04	1.44*	0.04	-0.01*	0.05
Slovenia	1.01*	0.03	0.98	0.06	-0.02	0.06
Scotland	1.19*	0.05	1.16*	0.04	-0.03*	0.06
Canada (Alberta)	1.47*	0.04	1.40*	0.03	-0.07*	0.05
Austria	1.15*	0.02	1.05	0.05	-0.10*	0.05
Canada (Nova Scotia)	1.30*	0.04	1.18*	0.07	-0.12*	0.06
United States	1.29*	0.04	1.16*	0.04	-0.14*	0.05
England	1.38*	0.05	1.22*	0.05	-0.15*	0.06
Slovak Republic	1.11*	0.04	0.94	0.06	-0.17*	0.07
Italy	1.37*	0.04	1.09	0.03	-0.28*	0.05

Source: Authors' calculations using the PIRLS 2006 and PISA 2012 microdata.

Note: * = significantly different from Spain at the five percent level. Senate weights applied.

Table 5. Socio-economic differences in the reading competency between ages 9/10 and 15/16 (international Z-scores): father's occupation

Country	Age 9/10		Age 15/16		Change	
	Beta	**SE**	**Beta**	**SE**	**Beta**	**SE**
Netherlands	0.31*	0.09	0.58*	0.05	0.27*	0.10
Belgium (Flemish)	0.38*	0.05	0.61*	0.05	0.23*	0.08
Taiwan	0.48*	0.06	0.68*	0.05	0.19*	0.08
Hong Kong	0.16*	0.07	0.32*	0.04	0.16	0.09
Belgium (French)	0.58*	0.09	0.75*	0.06	0.16	0.11
Italy	0.44*	0.18	0.57*	0.03	0.12	0.19
Slovak Republic	0.86*	0.16	0.95*	0.08	0.09	0.18
France	0.62*	0.09	0.70*	0.05	0.07	0.11
Iceland	0.40*	0.08	0.47*	0.06	0.06	0.10
Canada (Ontario)	0.36*	0.09	0.42*	0.05	0.06	0.11
Canada (British Columbia)	0.38*	0.08	0.43*	0.06	0.04	0.10
Norway	0.41*	0.07	0.42*	0.04	0.01	0.09
Qatar	0.57*	0.23	0.57*	0.04	-0.00	0.23
Canada (Quebec)	0.42*	0.09	0.41*	0.04	-0.01	0.10
Canada (Alberta)	0.46*	0.10	0.44*	0.05	-0.01	0.11
Poland	0.59*	0.07	0.52*	0.04	-0.06	0.09
Slovenia	0.66*	0.07	0.56*	0.05	-0.10	0.09
Indonesia	0.50*	0.09	0.39*	0.08	-0.10	0.12
Spain	0.58*	0.06	0.48*	0.03	-0.10	0.07
Israel	0.89*	0.15	0.75*	0.06	-0.14	0.16
Canada (Nova Scotia)	0.52*	0.07	0.35*	0.06	-0.17	0.10
England	0.84*	0.19	0.66*	0.05	-0.18	0.20
Scotland	0.84*	0.10	0.50*	0.05	-0.33*	0.11

Source: Authors' calculations using the PIRLS 2006 and PISA 2012 microdata.

Note: Figures refer to the estimated test score difference between children from working class and professional background. All figures reported in terms of international standard deviations. * indicates significant differences from zero at the five percent level. Parental occupation not available for the Austrian case in the PISA 2012 database downloaded from the OECD site. United States not included, as no information is available in PIRLS 2006 on the relevant SES variable. Senate weights applied.

Figure 1. The standard deviation of reading test scores at ages 9/10 and 15/16 (international Z-scores).
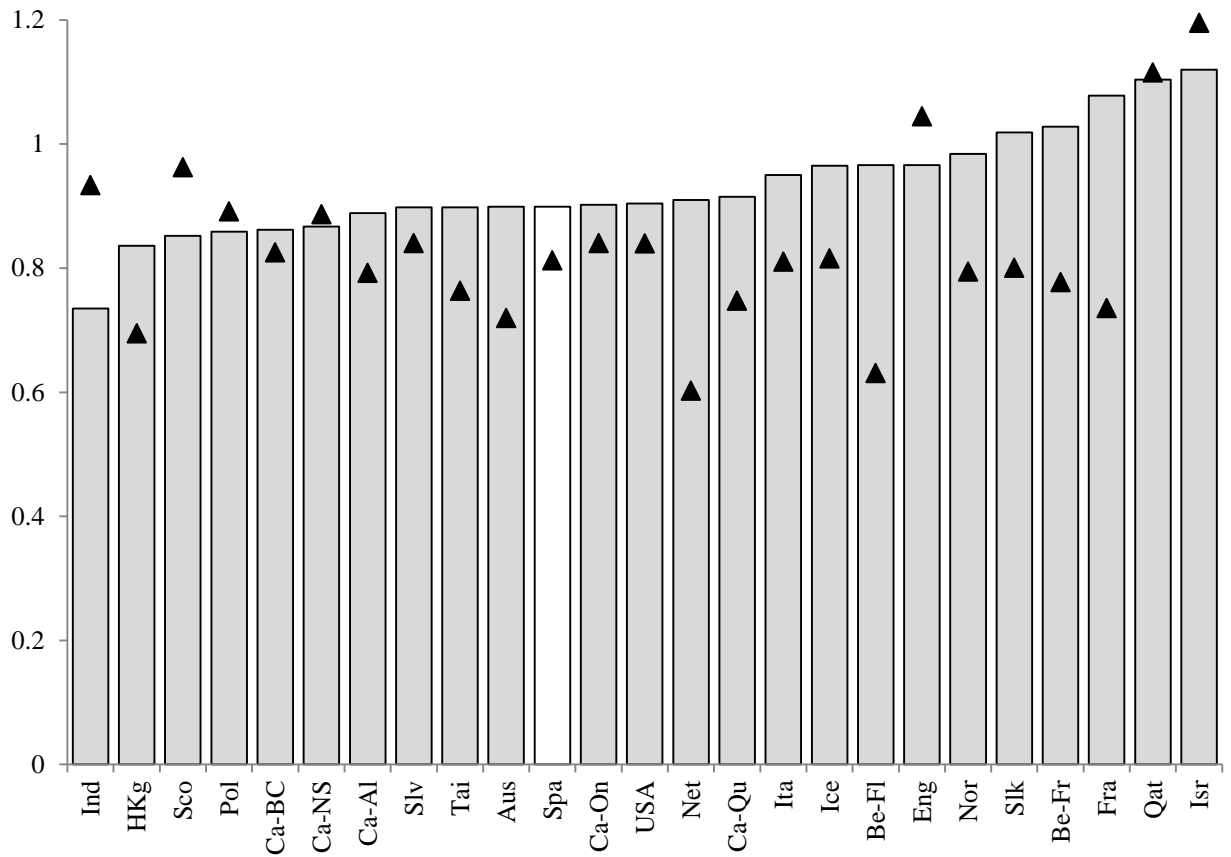
Source: Self-elaboration from PIRLS 2006 and PISA 2012 microdata.

Note: ▲ Standard deviation at age 9/10. Senate weights applied. Bar for Spain highlighted in white.

Figure 2. Changes in the 10th and 90th percentile of the reading competency test distribution between ages 9/10 and 15/16 (international Z-scores).
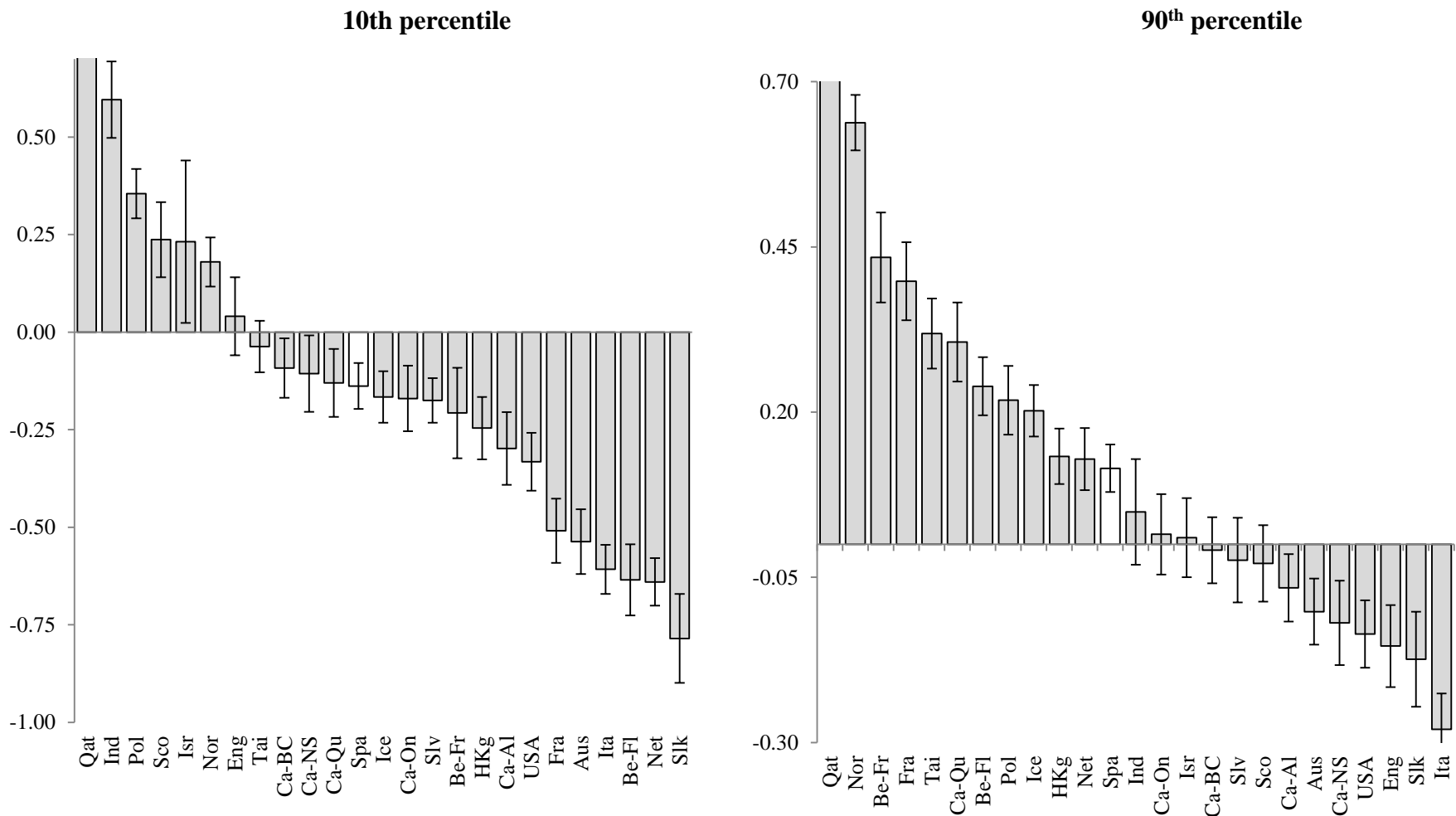
Source: Self-elaboration from PIRLS 2006 and PISA 2012 microdata.
Note: The line running through the bars is the estimated 95% confidence interval. Senate weights applied. Bar for Spain highlighted in white.