

# Spain: The Evidence Provided by International Large-Scale Assessments About the Spanish Education System: Why Nobody Listens Despite All the Noise



Montse Gomendio

**Abstract** ILSAs show that student performance in Spain is lower than the OECD average and has shown no progress from 2000 until 2011/2012. One of the main features is the low proportion of top performers. During this long period of stagnation, the education system was characterized by having no national (or standardized regional) evaluations and no flexibility to adapt to the different needs of the student population. The fact that the system was blind and rigid, plus the lack of common standards at the national level, gave rise to three major deficiencies: a high rate of grade repetition, which led to high rates of early school leaving, and large differences between regions. These features of the Spanish education system represent major inequities. However, PISA findings were used to reinforce the misguided view that the Spanish education system prioritized equity over excellence. After the implementation of an education reform, some improvements in student performance took place in 2015 and 2016. Unfortunately, the results for PISA 2018 in reading were withdrawn for Spain, apparently due to changes in methodology which led to unreliable results. To this date, no explanation has been provided raising concerns about the reliability and accountability of PISA.

## 1 The Value of International Comparisons: Uses and Misuses

The main goal of education systems is to equip students with the knowledge and skills that are required to succeed in current and future labour markets and societies. These are changing fast due to the impact of megatrends, such as technological change, globalization, demographic trends and migration. In particular, digitalization is leading to major changes in the workplace due to the automation of jobs and tasks, and it has modified dramatically the way people communicate, use services and obtain information (OECD 2019a). In order to be able to adapt to and benefit

---

M. Gomendio (✉)  
Spanish Research Council, Madrid, Spain  
e-mail: [montseg@csic.es](mailto:montseg@csic.es)

from these changes, people need increasingly higher levels of knowledge and skills, as well as new sets of skills (OECD 2019b). In such demanding and uncertain environments, education systems are under huge pressures to become more efficient, more responsive to changing needs, and to be able to identify which bundles of skills do people need.

Education and training systems remain the responsibility of countries and most of them have become decentralized to different extents and in different ways. Thus, in many countries national governments have transferred to subnational entities (regions, states and/or local authorities) the management of schools (OECD 2019b). Depending on the model of decentralization, regions may be responsible for raising the funding or may receive transfers from national governments. In most cases national governments retain the responsibility of defining the goals for each educational stage and, therefore, for defining the standards to evaluate student outcomes. Countries differ to a large extent in how ambitious these educational standards are.

The belief that education should remain a national policy is so ingrained, that even when countries organize themselves under the umbrella of supranational entities, such as the European Union, these have no direct responsibilities over education systems and they can only support their member states by defining overall targets and offering support (funding, tools and advice). Thus, the curricular contents, teacher training and professional development programmes, the degree of ambition in terms of the student outcomes required to obtain degrees and the way to measure them, are defined by each national government. For this reason, education has been regarded for a long time as one of the policy sectors which shows greater heterogeneity between countries. For a long time, this led to the widespread conclusion that international comparisons were difficult or worthless, because education systems were so unique and adapted to the national context that no common metric would be able to capture meaningful differences.

In this context, the international large-scale assessments (ILSAs) which started in 1995 (IEA: PIRLS and TIMSS) and 2000 (OECD: PISA) initially faced scepticism over their true value. The main critics argued that the methodology was flawed, that differences between countries were meaningless or that they focused too much on a narrow set of subjects and failed to capture important outcomes of the education systems. Overtime this has changed and the main ILSAs are increasingly regarded as useful tools to compare student performance between different countries. In fact, the international surveys have revealed large differences between countries in student performance which are equivalent to several years of schooling, showing that differences in the quality of education systems are much larger than expected. This has shifted the focus of the educational policy debate from an almost exclusive emphasis on input variables (the amount of resources invested) to output variables (student outcomes).

For countries and governments, the value of ILSAs lies in providing international benchmarks, which allows them to compare their performance directly with that of other countries, as well as evidence on trends over time. International surveys can also be useful to measure the impact of educational policies on student performance, although drawing causal inferences remains controversial mainly due to the

cross-sectional nature of the samples (Cordero et al. 2013, 2018; Gustafsson and Rosen 2014; Hanushek and Woessmann 2011, 2014; Klieme 2013; Lookheed and Wagemaker 2013).

As an increasing number of countries has joined these international surveys and trust on them has strengthened, the media impact has grown and with it the political consequences. This has raised the profile of international surveys, PISA in particular, but it has also turned them into a double-edged sword. On the positive side, as the impact grows, more people become aware of the level of performance of their country in relation to others and to the past. They have also promoted much needed analyses on which are the good practices that lead to improvements in certain countries, what policies have top-performing countries implemented, and to what extent are good practices context-specific or useful in other contexts (Cordero et al. 2018; Hanushek and Woessmann 2014; Hopfenbeck et al. 2018; Johansson 2016; Lockheed and Wagemaker 2013; Striethold et al. 2014). On the dark side, this leads to a very narrow focus on the ranking between countries and to oversimplistic hypotheses concerning the impact of policies implemented by different governments. Thus, international surveys, PISA in particular, have become powerful tools in the political debate. This is a reality that must be acknowledged and raises the bar for ILSAs to be reliable and accountable.

As mentioned before, education systems need to evolve to continue to improve and to ensure that students are equipped with higher and more complex skills that allow them to adapt to an ever-changing landscape. This puts ILSAs in a dilemma. On the one hand, the metrics need to change and adapt to these changes in order to remain a meaningful tool to compare countries. On the other hand, the metrics need to remain stable and consistent in order to measure change over time (e.g. Klieme 2013). The balance between these two opposing forces lies in leaving enough anchor items unchanged, so that the information provided about change at the systemic level is robust.

## 2 The Spanish Case: Shedding Light on the Darkness

The media impact of PISA is much greater in Spain than in other countries (Martens and Niemann 2010). One plausible explanation is that Spain does not have national evaluations, so PISA scores represent the only information available concerning how Spain performs in relation to other countries and over time. The reasons for the lack of national evaluations are complex. The Spanish education system has followed a rather radical version of the “comprehensive” model since 1990 when a major education reform was approved: the LOGSE (Delibes 2006; Wert 2019). The comprehensive model is based on the premise that all students should be treated equally and its most extreme forms regard evaluations as a discriminatory tool that unfairly segregates students who fail (for a discussion of comprehensive education models see also Adonis 2012; Ball 2013; Enkvist 2011). In addition, political parties on the left of the ideological spectrum often argue that evaluations are a tool designed

to prevent students from disadvantaged socioeconomic backgrounds from entering university. Finally, most regions fear that national evaluations represent an important step towards the re-centralization of education and do not recognize the responsibility that the national government has in defining the standards required to attain the degrees that are provided by the Ministry of Education for the whole country.

As a consequence, there are no national evaluations and many regions do not have evaluations at the regional level either. In other words, the Spanish education system is blind, since no information is available on how students perform according to homogeneous standards. This has important consequences. The lack of evaluations in the first years of schooling means that it is not possible to detect early enough students lagging behind in order to provide the additional support required. Thus, throughout primary students of different levels of performance advance from one grade to the next. When students enter secondary, many of them have not acquired the basic knowledge and skills, leading to a high rate of grade repetition. This defining feature of the Spanish education system is somewhat surprising since grade repetition takes place in the absence of uniform standards or strict rules, instead it's the result of the decisions made by teachers. The lack of national (and regional) evaluations at the end of each educational stage, implies that there is no signalling system in place to inform students, teachers and families, of what the expected outcomes are. Thus, each school and each teacher develops its own standards. Obviously, this leads to increasing heterogeneity which has generated huge differences between regions.

Therefore, ILSAs are the only instrument available to measure student performance with the same standards in the whole country and they have been increasingly used to compare the performance of different regions. Since Spain joined PISA much earlier than other ILSAs and has participated in every cycle, the strongest body of evidence comes from PISA. In this rather unique context, the impact of PISA results in Spain is not only (or not so much) about how Spain performs in relation to other countries. Instead it is the result of intense political debates about the impact of different policies and the causes of large differences between regions.

Given that PISA is held in high regard in Spain it seems particularly unfortunate that the results of the main domain (reading) in PISA 2018 have not been released and many unanswered questions remain about the reliability of results for science and mathematics. For this reason, most of the analyses in this Chapter use data from PISA 2015. I will discuss more general implications of what has happened in Spain with the findings from the PISA 2018 cycle in the last section of the Chapter.

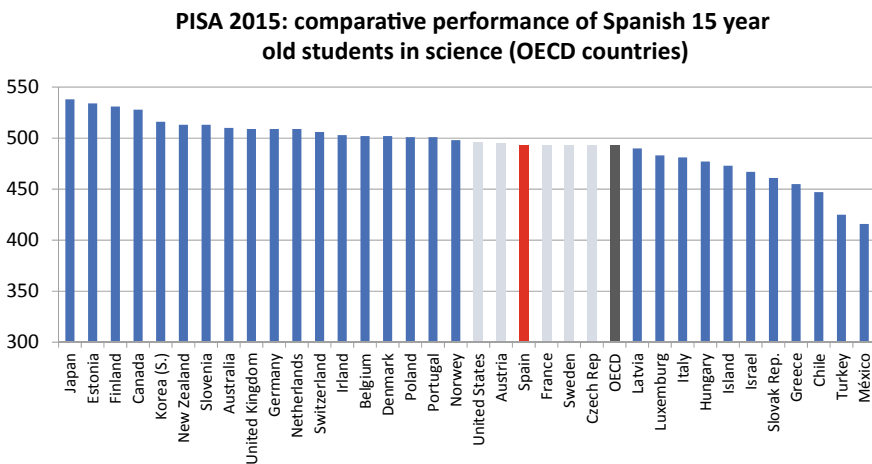
## ***2.1 The Performance of Spain in Comparison to Other Countries: Ample Room for Improvement***

The three major international large-scale assessments (PIRLS, TIMSS and PISA) measure the same domains: reading, mathematics and science, but the methodology, lengths of the cycles and the target population (as defined by student age or grade)

are different. The IEA developed the initial surveys, sampling all students in each classroom and focusing on specific grades. TIMSS (Trends in International Mathematics and Science Study) has monitored the performance of students in grade 4 and 8 in mathematics and science every four years since 1995. PIRLS (Progress in International Reading Literacy Study) has monitored trends in reading achievement at the fourth grade since 2001 and it takes place every five years. Finally, the OECD developed PISA (Programme for International Student Assessment) which samples 15-year-olds in different grades (8, 9, 10 and 11th grades), started in 2000 and has 3-year cycles. While PIRLS and TIMSS have been designed to analyse the extent to which students have acquired curriculum-based content (Mullis et al. 2016, 2017), PISA claims to analyse how the knowledge and skills acquired are applied to solve problems in unfamiliar settings (OECD 2019a). PISA also claims to be more policy-oriented and in fact PISA publications include many analyses to try to identify which good practices distinguish good performing countries (OECD 2016a, b, 2019c, d).

According to PISA, Spain has scored below the OECD average until 2015 when Spain reached OECD levels. The performance of Spain in 2015 was significantly below that of 18 OECD countries, and substantially below top performers such as Singapore. Thus, there seems to be ample room for improvement (Fig. 1).

When the three domains are considered separately, in 2015 Spain performed at the same level as the OECD in science and reading, but below the OECD average in maths. Both in science and reading Spain has a smaller proportion of both low performing and top performing students than the OECD average. However, in maths the proportion of low performing students is similar to the OECD average, while Spain has a substantially lower proportion of top performing students. Thus, the main reason why Spanish students tend to perform worse in maths is because such a small proportion are top performers. More generally, it can be concluded that one of



**Fig. 1** PISA 2015 scores for the main domain (science)

the weaknesses of the Spanish education system is that it does not allow the potential of top performing students to develop.

It is important to take into account the fact that grade repetition is high in Spain compared to other countries (2015: 36.1% in Spain *vs* 13% OECD average). Since the PISA survey includes in the sample 15-year-olds irrespective of the grades in which they are, the % of 15-year-olds in Spain which are in 10th grade is 67.9%, while 23.4% are one year behind and 8.6% two years behind (OECD 2016a). Students who repeat a grade have 99 less points in PISA. Thus, it seems likely that 15-year-olds who have not repeated any grades (i.e. only those in 10th grade) would have a substantially higher score. The fact that grade repetition explains to a large extent the overall PISA scores for Spain, as well as differences between regions, has not received enough attention.

Spain has only participated in the PIRLS and TIMSS surveys for 4th grade. Thus, there are no data for the 8th grade which is generally treated as targeting a sample of students broadly comparable to those included in the PISA survey. However, the sample of 4th grade students provides useful information on the performance of students in primary (Martin et al. 2016; Mullis et al. 2016; 2017). The evidence from TIMSS 2015 shows that Spain performs slightly below the OECD in science and much lower in maths. In addition, evidence from PIRLS 2016 shows that Spanish students perform slightly below the OECD in reading.

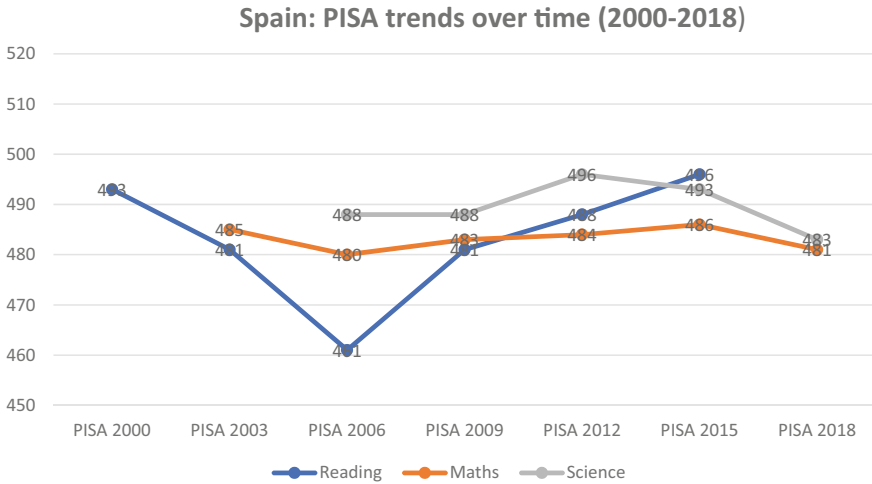
Thus, taking together all the evidence from PISA, PIRLS and TIMSS, it shows that Spanish students have levels of performance similar or only slightly below OECD averages in reading and science, but considerably lower in maths both in primary and secondary. The main deficiency of the education system that explains these results is the small proportion of top performing students. The three surveys also show that Spain performs below around 20 OECD countries and much lower than top performers in Asia such as Singapore and Japan.

## ***2.2 What ILSAs Tell Us About Trends Over Time***

According to PISA in Spain there has been no significant improvement in reading (2000 versus 2015), mathematics (2003–2015) or science (2006–2015). Apparently, there is a modest decline in 2018 for mathematics and science, but these data should be treated with caution since results from the main domain (reading) have been withdrawn due to inconsistencies.

However, the trends seem different for each domain. Over time reading seems to have experienced a decline until 2006, followed by a steady recovery afterwards. Science experiences a slight improvement in 2012 which remains in 2015 and mathematics shows a flat shape (Fig. 2).

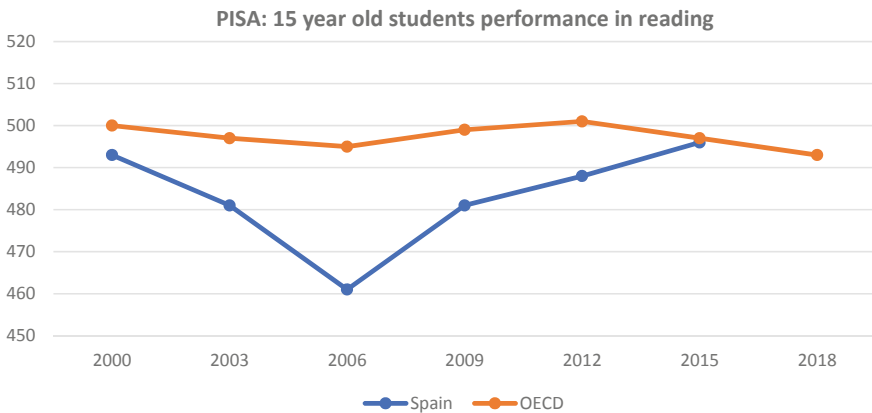
When trends over time are compared to those of the OECD average it emerges that OECD countries have not experienced much change in reading, showing first a slight decline until 2006, followed by a modest recovery until 2012. Spain showed lower values in most cycles and followed a similar trend over time, but the changes in each



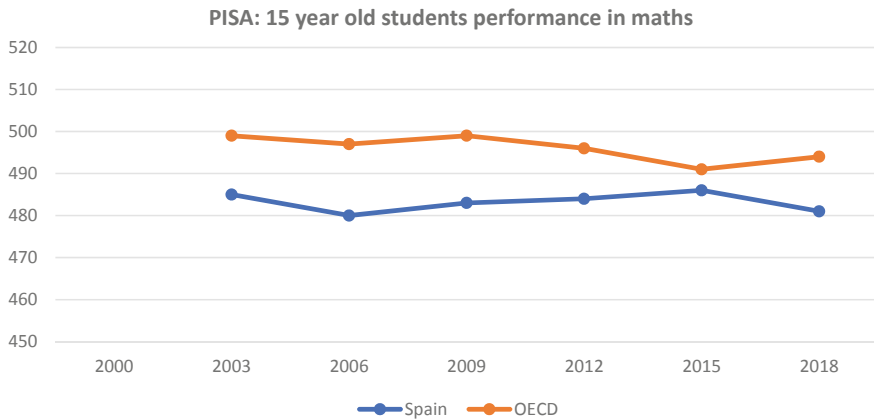
**Fig. 2** PISA scores for Spain from 2000 until 2018

cycle are much more dramatic; the difference between the two became particularly large in 2006 when Spanish students performed at their lowest levels. In 2015 the OECD average declined and continued to drop thereafter, reaching the lowest value of the whole series in 2018. In contrast, in Spain student performance improved between 2012 and 2015. This seems to be mainly the result of a decrease in the proportion of low performing students in 2015. As a result of the opposing trends between 2012 and 2015, in the latter Spain reached the same level of performance as the OECD (Fig. 3).

In mathematics Spain has shown lower values than the average for the OECD in all cycles, except in 2015 when it converged with the OECD average. The poor



**Fig. 3** Spain versus OECD: reading performance over time (PISA)



**Fig. 4** Spain versus OECD: maths performance over time (PISA)

performance of Spain seems to be mainly due to the low proportion of top performing students in maths. Similarly to the trend for reading, OECD countries have not experienced major changes over time in maths: there is a slight decline from 2009 until 2015, followed by a weak recovery in 2018. In all subsequent cycles OECD averages have been lower than the first cycle (2003). Similarly, Spain shows only slight changes, with an initial decline in 2006 followed by a weak recovery until 2015 (Fig. 4).

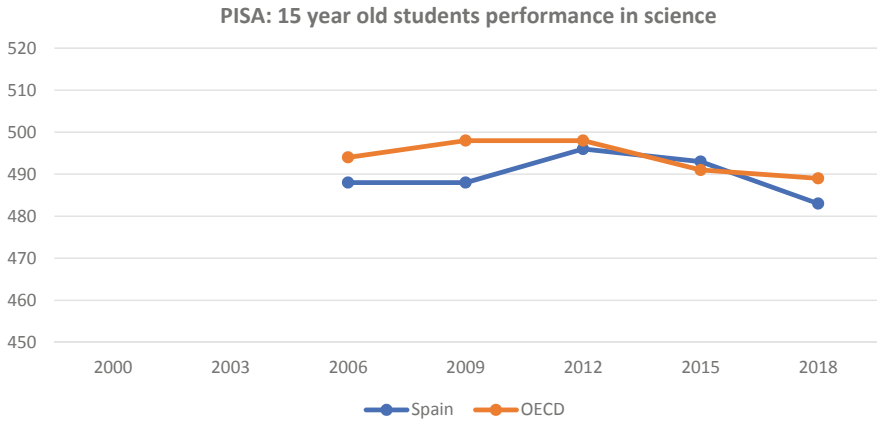
In science Spain has performed slightly below OECD averages in the first two cycles and reached similar values from 2012 onwards. Over time Spain shows a moderate improvement in 2012 and then declines following a similar trend than the OECD. Once again in this domain OECD countries seem to show only slight changes and a decline since 2012. The lower values for Spain seem to arise due to the smaller proportion of top performers, and the convergence experienced in 2012 and 2015 could be explained by the fact that Spain has a smaller proportion of low performing students than the OECD (Fig. 5).

The more limited evidence available for Spain from PIRLS and TIMSS seems to show greater improvements than PISA. Spain improves from 2011 until 2015/2016 reaching values similar to OECD averages. However, it remains below more than 20 OECD countries.

After a lack of progression between 2006 and 2011 in reading, Spain experienced a considerable improvement in 2016. This is due mainly to a decrease in the proportion of low performing students (28–20%). In contrast, the OECD showed only marginal improvements (Fig. 6).

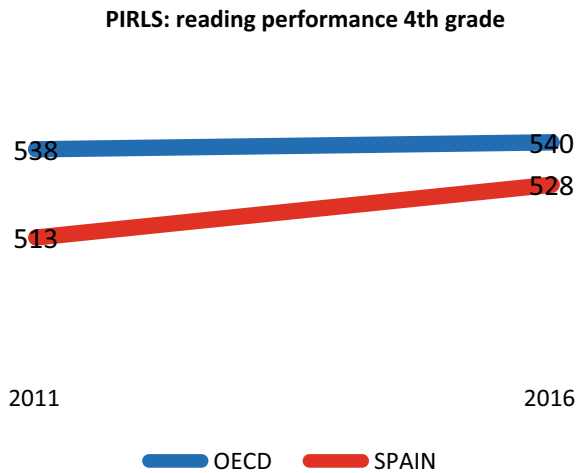
The lowest level of performance of Spain in comparison to the OECD is in maths, even after the substantial improvement experienced in 2015 in Spain and the lack of progress for OECD countries as a whole. This seems to be mainly due to the small proportion of top performing students in Spain (Fig. 7).





**Fig. 5** Spain versus OECD: science performance over time (PISA)

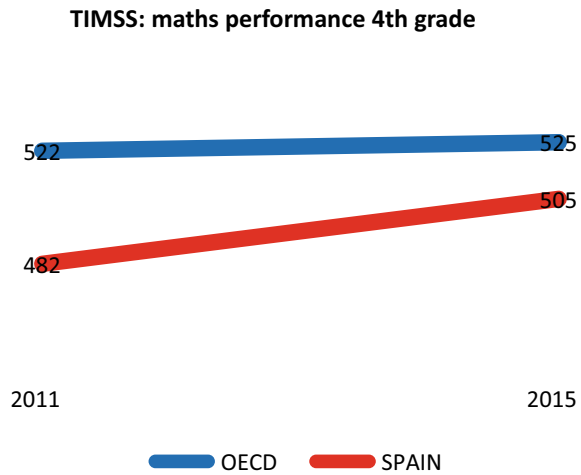
**Fig. 6** Spain versus OECD: reading performance over time (PIRLS)



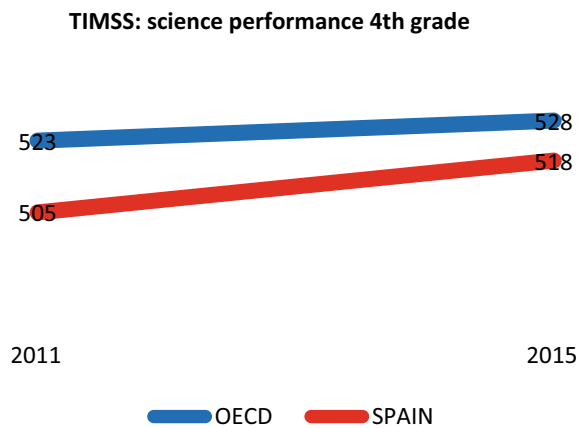
Finally, Spain showed more similar levels of performance to the OECD in science in 2011, which improved in 2015 reaching similar values to the average of the OECD (Fig. 8).

*Taken together the findings from these international surveys seem to suggest the following. From 2000 until 2012 Spanish students perform below the OECD average and remain stagnated over time. The first signals of improvement appear in 2015 when primary students in science, and to a lesser extent in maths, perform better than in previous cycles (TIMSS 2015). One year later, primary students show a clear boost in reading (PIRLS 2016). Among secondary students, weaker improvements were also seen among 15-year-old students in reading, and to a lesser extent in science*

**Fig. 7** Spain versus OECD: maths performance over time (TIMSS)



**Fig. 8** Spain versus OECD: reading performance over time (TIMSS)



and maths (PISA 2015). As a result, in 2015 Spanish 15-year-old students reached a similar level of performance than the OECD average in reading and science, but remained below in maths.

### 2.3 What PISA Reveals About Regional Differences

In the context of the European Union, the Spanish education systems is quite unique in that there are no evaluations of student performance at the national level. In addition, regions have failed to agree on common standards to measure student performance and even on whether or when should student evaluations take place. Thus, many regions do not have evaluations at the regional level. Those regions which do have

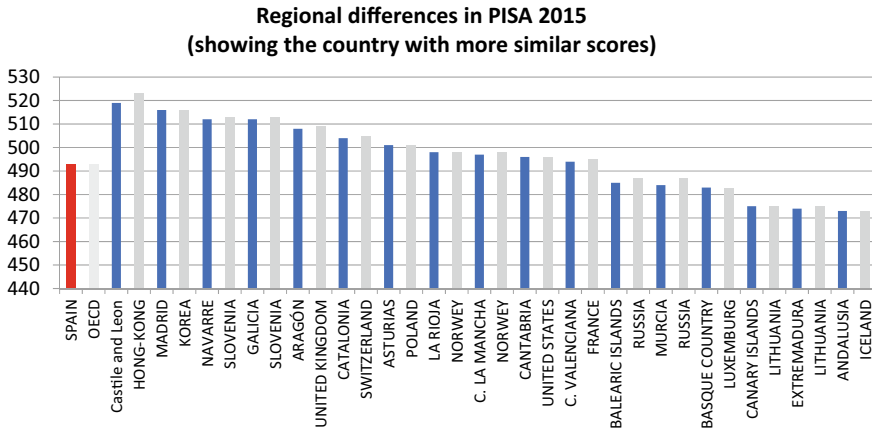


Fig. 9 Regional differences in Spain according to PISA 2015

evaluations tend to include only a limited sample of the students. However, regions have been willing to fund larger sample sizes in PISA surveys in order to get statistically meaningful scores at this level, which clearly reflects an interest in using common metrics that allow direct comparisons between regions, as well as trends over time.

Data at the regional level show that the PISA average for Spain hides major differences between regions (OECD 2015). Thus, in PISA 2015 the difference between the top performing region in science (Castilla y León) and the lowest performing region (Andalucía) is the equivalent of more than 1.5 years of schooling. Of the 17 regions, 11 perform above the OECD average and 6 below.

The distribution of students of different levels of performance between regions shows that the proportion of low performing students varies from 11 to 25%, while the proportion of top performing students fluctuates from 3 to 9% (Fig. 9).

### 2.4 Differences in Levels of Investment Do not Explain Trends Over Time nor Regional Differences

It is important to try to understand what the reasons are underlying the lack of progress in student performance over such a long period of time, as well as the huge disparities between regions. Comparing the Spanish regions also provides an opportunity to compare systems that operate under the same institutional structure and the same basic laws, i.e. same age of school entry, duration of compulsory schooling, basic curricula, and existence of alternative pathways (academic vs vocational education and training).

In Spain the political debate around education focuses almost exclusively on two issues: levels of investment (i.e. input variables) and ideological topics which

contribute to the polarization of the debate. Very little attention is paid to understand which factors contribute to improve student outcomes.

The political debate assumes that increases in levels of investment automatically result in improvements in student outcomes and the other way around. As we will see, this is not the case. It is important to clarify first a few general issues about how funds are raised, distributed and spent in the Spanish education system.

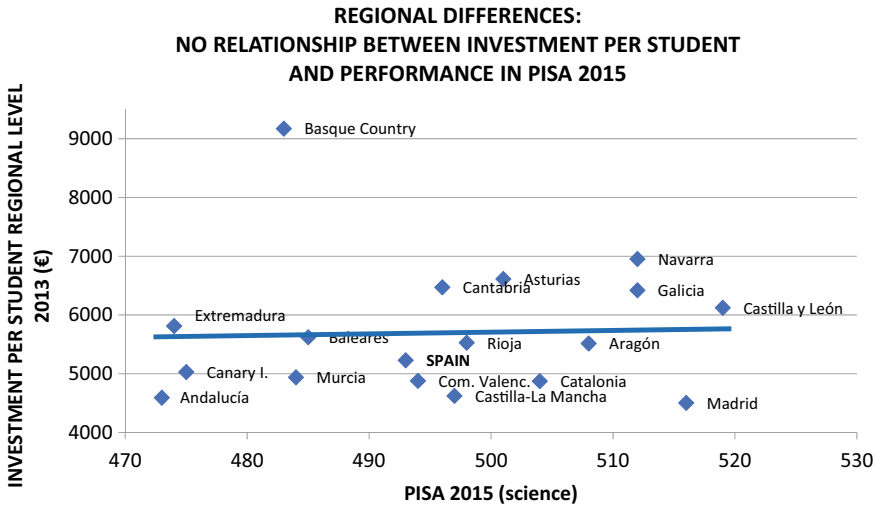
In Spain it is the responsibility of the national government to raise most of the public funds through taxes. Funds assigned to education, health and social affairs are then transferred as a package to regions, following an agreed formula which allocates funds according to population size, demographic factors and degree of dispersion; to some extent this formula is also designed to redistribute funds from wealthier to poorer regions. It is the responsibility of regions to decide how much to invest in each of these “social” policies. After the economic crisis of 2008 regions had to make decisions about where to implement the budget cuts and, as a consequence, levels of investment in education were reduced to a much larger extent than health or social affairs.

Given that the national government transfers most of the funds allocated for social policies to regions, around 83% of the funds that are invested in education are managed by regions. However, accountability mechanisms are lacking to the extent that there is little information available on student performance.

As in most countries, in Spain investment in staff represents more than 60% of the funding allocated to education. Thus, the overall level of resources assigned to education is mainly the result of two factors: the number of teachers (which is, in turn, the product of the number of students and the ratio students per teacher) and the salary of teachers.

Overall investment in education in Spain increased substantially from 2000 until 2009 (2000: 27.000 M euros, 2009: 53.000 M euros), when a peak was reached, and decreased thereafter due to the economic crisis. As we have seen with the evidence provided by the ILSAs, there were no improvements in student performance during the period in which levels of investment increased. On the contrary, levels of performance remained stubbornly low. This suggests that the additional resources were allocated to variables which had no impact on student outcomes. Against all expectations, improvements in student performance were detected by international surveys in 2015 after substantial reductions in investment on education were implemented by regions. Obviously, the budget cuts per se cannot be responsible for the improvements in student outcomes, but this evidence suggests that (a) the system became more efficient in the use of resources, and (b) other changes in policy could be responsible (see below).

Another line of evidence which strongly supports the view that it is wrong to assume that levels of investment in education are directly related to the quality of the system (i.e. levels of student performance) comes from a comparison between regions. Levels of investment per student show large variation between regions: the Basque Country invests twice as much than Madrid or Andalucía. However, there is no relationship whatsoever between investment per student and the level of student performance according to PISA. In fact the two regions at the extremes of the range



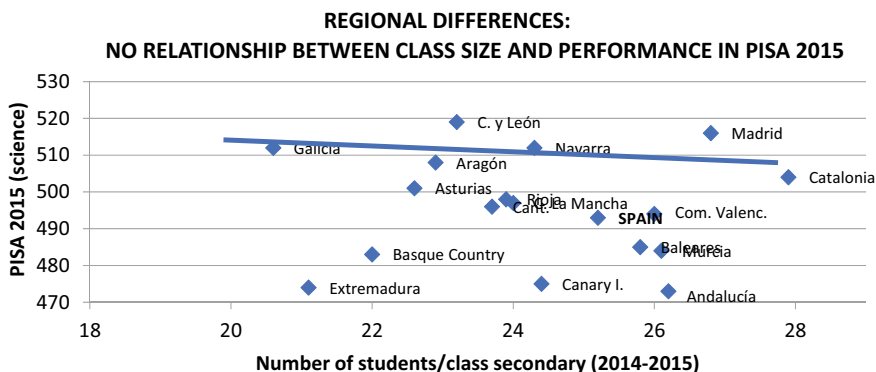
**Fig. 10** Relationship between investment per student in each region and student performance according to PISA 2015 (modified from Wert 2019)

of investment levels are clear outliers: students in the Basque Country have poor levels of performance despite the fact that this region shows the highest levels of investment per student by far, and students in Madrid are among the highest performing students despite the low levels of investment per student (Fig. 10).

Perhaps the second most widespread assumption is that the ratio of students per teacher is associated with student outcomes. Many families use class size as a proxy for quality; thus, in the political debate decreasing class size is regarded as an assurance of improved outcomes and increasing it as a major threat to the quality of the system. The evidence also shows that this assumption is wrong.

It is important to realize that the belief that class size is a proxy for quality is so strong in Spain, that over the years a growing share of the resources has been devoted to decreasing class size. As a consequence, Spain has a smaller ratio of students per teacher than most EU and OECD countries. Even after a small increase in class size during the economic crisis, Spain in 2014 had a smaller ratio of students per teacher in public schools than the OECD (11 versus 13) and slightly larger in private schools (15 versus 12) (OECD 2016c). Despite all the resources invested in reducing the ratio, no improvements in student outcomes were detected and Spain continued to perform below the OECD average before 2015, and much worse than countries in Asia which have very large class sizes.

There are large differences between regions in class size, with Galicia being close to 20 students per class and Cataluña close to 28. Among PISA participating countries the range is much larger since top performing countries in Asia tend to have much larger class sizes than countries in Europe. However, an analysis of the impact of class size between regions in Spain may be more meaningful since it clearly excludes many of the confounding factors that cannot be accounted for when PISA



**Fig. 11** Relationship between class size in each region and student performance according to PISA 2015 (modified from Wert 2019)

participating countries are compared. At the regional level, there is no relationship whatsoever between class size and student performance in PISA (Fig. 11).

A third widespread assumption is that teacher salary has a positive impact on student outcomes, because good candidates can only be attracted into the teaching profession if the salaries are high enough. Unfortunately, Spain is a clear example that teacher salaries per se are unrelated to student performance. Teacher salary is higher in Spain than the average for the EU and the OECD at all stages, but particularly the starting salary (OECD 2017a). However, as we have seen, student outcomes are poor. Probably the reason is that salaries are not linked to teacher performance, University educational degrees are not demanding, and the requirements to become a teacher give too much weight to seniority and too little to merit.

Since the variables that have to do with the input of resources into the education system do not seem to be able to explain either trends over time in student performance, nor differences between regions (see also Cordero et al. 2013; Villar 2009), education reforms and changes in education policies should be considered.

## 2.5 The Impact of Education Policies

The debate about educational policies in Spain rests on the assumption that there have been too many legislative changes and that the root of the problem lies partly in the instability created by so many changes. Quite the opposite. The LOGSE in 1990 established the architecture and rules of the game of a “comprehensive” system which remained essentially the same until 2013 when a partial reform of this law was approved (LOMCE). Since the educational laws approved between 1990 and 2013 did not imply major changes, the education system in Spain did not change in any substantial way for 23 years.

The LOGSE extended compulsory education to the age of 16 and increased the number of teachers by 35%, which led to a marked decrease in the ratio of students per teacher. This required a substantial increase in the investment in the education system which increased until 2009, when the economic crisis led to the first budget cuts in education. The LOGSE implemented a “comprehensive” education system following a rather extreme interpretation. It was designed to treat all students equally under the belief that this was the only way to achieve the major goal: equity. Thus, until the end of below secondary (16 years) students could not receive differential treatment according to the level of performance, be grouped according to their ability, nor have the flexibility to choose among different trajectories.

The lack of national (and standardized regional) evaluations was a key element, since it was regarded as a way to avoid segregation and stress among students. Thus, the system was blind since no national metrics and assessments were developed to evaluate student performance. As a consequence, students who were lagging behind could not be identified early enough and did not get the additional support that they needed, and students who had the potential to become top performers were not given the opportunity to do so.

The rigidity of the educational system and the fact that it was blind to the performance of students, led to the emergence of two problems which have remained the main deficiencies of the Spanish education system ever since. First, the level of grade repetition increased, since low performing students had no other choice. In 2011, the rate of grade repetition in Spain was almost 40% (3 times that of the OECD); no progress had been made since at least 2000 when the same level of grade repetition was observed (INEE 2014). It is well known that grade repetition is an inefficient strategy, both for students and for the system as a whole (Ikeda and García 2014, Jacob and Lefgren 2004, Manacorda 2012). Students who repeat grades are much more likely to become early school leavers. In addition, the cost of grade repetition represented 8% of the total investment in education, obviously a very inefficient way to invest resources. Second, the level of early school leaving remained astonishingly high for decades (around 30%). A large proportion of these students left the education system with no secondary degree and, given their low levels of knowledge and skills, they faced high levels of unemployment (youth unemployment reached almost 50% in 2011). Most of the early school leavers came from disadvantaged and migrant backgrounds. Thus, a model which was designed in theory to promote equity, led to the worst type of inequality: the expulsion of students from an education system which was blind to their performance and insensitive to their needs.

The lack of national standards also led to major differences between regions in the rates of grade repetition which are closely associated with the rates of early school leaving. As we can see in Fig. 12 while the Basque Country has low rates of grade repetition and low rates of early school leaving, at the other extreme there is a large group of regions with rates of grade repetition around 40–45% and rates of early school leaving between 30–35%. The latter suffer from high rates of NEETs and youth unemployment.

In 2000 PISA offered the first diagnosis of the performance of Spanish students in comparison to other countries: poor level of performance, which is explained mainly

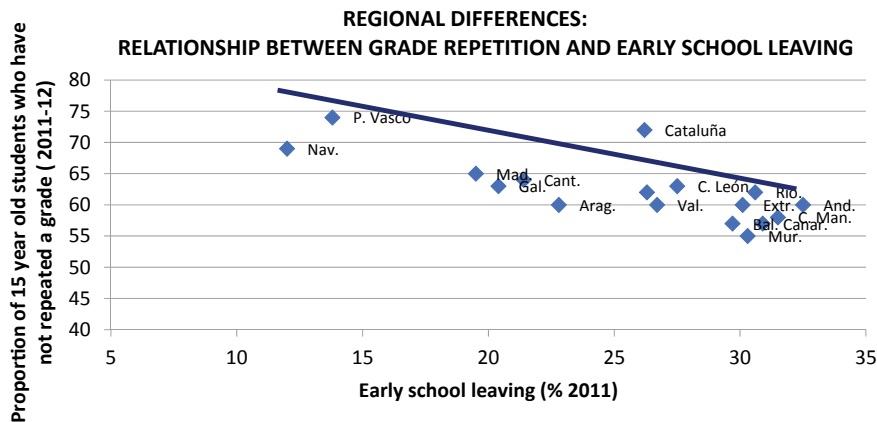


Fig. 12 Relationship between grade repetition and early school leaving (modified from Wert 2019)

by the small proportion of top performing students. Furthermore, this comparatively low level of performance in relation to the OECD remained until 2015 when similar levels of performance were achieved. It should be noted that PISA consistently defines the Spanish education system as equitable (OECD 2016a, 2019c, d), thus reinforcing the legend. This interpretation is based on the fact that fewer differences are found between schools than within, but completely ignores the fact that the high rates of grade repetition found at the age of 15 is a major source of inequalities leading to the eventual expulsion of around 1 in 4 students from the education system without having acquired the basic knowledge and skills.

In 2013 an education reform (LOMCE) was approved to address these deficiencies. Implementation started in primary in the following academic year (2014/15). The reform addressed 5 main pillars: (1) implementation of flexible pathways which included the modernization and development of vocational education and training in order to lower the high rates of early school leaving which had been for a long time a major source of inequality; (2) the modernization of curricula and the definition of evaluation standards to promote the acquisition of both knowledge and competences instead of the prevalent model which required almost exclusively the memorization of contents; (3) the re-definition of areas of the curricula that would be defined by the state and the regions; (4) enhancement of the level of autonomy of schools and the leadership role of principals, and (5) the establishment of national evaluations would allow the detection of students lagging behind early on to provide the support required to catch up, and would signal the knowledge and competences required to obtain the degrees at the end of each educational stage, so that students, teachers and families were aware of the standards required. These evaluations were also conceived as a potent signal that effort and progress, both from students and teachers, would be promoted and rewarded. The national evaluations also aimed to help ameliorate the major differences found between regions that were the root of differences in the rate of NEETs and youth unemployment. In this way, the national government would be



able to ensure minimum levels of equity among regions, so that all Spanish students could achieve similar levels of knowledge and skills.

These changes in educational policies led to clear and rapid improvements in the following: an increasing proportion of students enrolled in vocational education and training, leading to a historic decline in early school leaving between 2011 and 2015 (26.3–19.9%), and the rate of grade repetition declined (Wert 2019). From the very first year of its implementation, the LOMCE provided additional funding to the regions to offer a growing number of places in vocational education and training, and to modernize their qualifications.

However, the national evaluations that represented one of the main pillars of the reform, were never fully implemented due to the intensity of the political pressures against them. In 2014/2015 the new curricular contents were implemented, as well as the national evaluations in primary. In the following academic year, the full implementation of the calendar designed for evaluations at the end of lower secondary and upper secondary was interrupted. This concession was made to facilitate a national consensus on education. However, no progress has been made on reaching a consensus.

Thus, interpretations about the impact of this education reform on student performance must remain speculative. It seems reasonable to argue that, since implementation of the reform started in primary (including curricular content and the introduction of evaluation standards, as well as the first national evaluations), the improvements detected by TIMSS in science in 2015 may represent a first signal of a positive impact; the fact that primary students improved substantially their performance in reading in 2016 (i.e. 2 years after implementation started in primary) supports the view that consistent improvements in student performance were already taking place.

The evidence from PISA seems less clear, since 2015 may have been too early to detect any changes among 15-year-olds, although the decrease in low performing students in reading seems consistent with the evidence from other international surveys.

Unfortunately, it will be difficult to evaluate any further the impact of this education reform on student performance, since subsequent governments paralyzed important aspects of the implementation of the reform. In addition, no PISA results were released for Spain in the main domain in 2018 (reading).

### **3 What Happened in PISA 2018: A Broken Thermometer?**

At the official launch in December 2019 of the PISA results, the data for Spain in the main domain, i.e. reading, were not released. Despite uncertainties about the reliability of the results for maths and science, these were published. The OECD press release and the explanation provided in the PISA publication (OECD 2019c, Annex A9) reads as follows: “Spain’s data met PISA 2018 Technical Standards. However, some data show implausible response behaviour amongst students”.

The problem lies in the new section on “reading-fluency”. According to PISA (OECD 2019c, page 270) the reading expert group recommended including a new measure of reading fluency to better assess and understand the reading skills of students in the lower proficiency levels. These items come from the PISA for Development framework (OECD 2017) which was developed to measure low levels of performance among 15-year-olds (in and out of school) in low- and middle-income countries. This section had the easiest items in the reading-literacy assessment. Items in this section seem designed to assess whether students had the cognitive skills to distinguish if short sentences make sense or not. Examples include “airplanes are made of dogs” or “the window sang the song loudly” which do not make sense but are grammatically correct, along with others such as “the red car had a flat tire” which are supposed to make sense and are also grammatically correct.

Any problems with this initial section may have had major implications on the whole assessment because in 2018 PISA introduced another major change. PISA 2018 was designed for the first time as an “adaptive test”, meaning that students were assigned to comparatively easy or comparatively difficult stages later on, depending on how they performed on previous stages. This contrasts with PISA 2015 and previous cycles, when the test form did not change over the course of the assessment depending on how students performed in previous stages. It is also worth mentioning that this adaptive testing cannot be used in the paper-based assessments. Thus, any anomalies in this first section labelled as “reading fluency” may have led, not only to low scores, but more importantly to mistakes in how students were assigned to easy or difficult tests for the rest of the assessment.

According to the OECD a “large number” of Spanish students responded in a way that was not representative of their true reading competency (OECD 2019c, Annex A9). Apparently, these students spent a very short time on these test items and gave patterned responses (all *yes* or all *no*), but then continued onto more difficult items and responded according to their level of proficiency. Although the section on Spain claims that this problem is unique to this country (OECD 2019c, page 208), in a different section the OECD reports that this pattern of behaviour (“*straightlining*”) was also present in over 2% of the *high performing students* in at least 7 other countries (including top performers such as Korea) and even higher in countries such as Kazakhstan and the Dominican Republic (OECD 2019c, page 202). No data are provided on the prevalence of straightlining behaviour among all students. The OECD recognizes that it is possible that some students “did not read the instructions carefully” or that “the unusual response format of the reading fluency tasks triggered disengaged response behaviour”.

It is a matter of concern that, despite the high incidence of straightlining behaviour among Spanish students, the OECD did send the results for all three domains to the Ministry of Education and to the regions with extended samples assuming that they complied with the so-called “PISA technical standards”. Very soon after receiving the data, some regions detected the problem with straightlining behaviour in the “reading fluency” section, which they claim has a considerable impact on the scores of the overall reading test. In addition, the regions discovered that the unreliable results for reading seem to contaminate the results for the other two domains, since

students who did not perform the science or maths test were given scores that were extrapolated from the reading test. In addition, some regions reported major flaws in the scores given to students in schools that were the responsibility of specific contractors.

Thus, these regions informed both the Spanish Ministry of Education and the OECD requesting an explanation or the correction of what seemed like errors. On the basis of the information provided by these regions, the OECD and the Ministry of Education agreed to withdraw the results for the main domain (reading). Despite doubts raised by the same regions about the reliability of the scores for the other two domains, and the fact that they are less robust statistically than the main domain, the OECD and the Ministry of Education agreed to release data on maths and science at the PISA launch in December 2019. No further explanations have been provided by the OECD.

Trust in international surveys requires accountability. The lack of explanations so far about the irregularities that led to the withdrawal of data have raised serious concerns about the reliability of the survey (El Mundo: “La Comunidad de Madrid pide a la OCDE que retire todo el informe PISA por errores de un calibre considerable: Toda la prueba está contaminada” 29 Nov 2019; El Mundo: “Las sombras de PISA: hay que creerse el informe tras los errores detectados?” 02 December 2019; El País “Madrid pide que no se publique ningún dato de PISA porque todo está contaminado” 30 November 2019; La Razón “Madrid llama chapucera a la OCDE por el informe PISA” 02 December 2019). Until the OECD explains in detail the methodological changes in the PISA 2018 survey it will be difficult to understand fully the implications, both for the comparability between countries and for comparisons with past cycles. In the case of Spain clear explanations should be provided about the irregularities that justified the decision to withdraw the data for reading, and the extent to which science and maths may also be affected by these problems.

The Spanish case illustrates how a substantial change in methodology in PISA 2018 led to serious methodological problems, which seem to have affected other countries. The extent of the problem is not known, since most countries did not question the results from the OECD and therefore did not conduct an independent evaluation of the PISA data provided. It is important to note that the concerns that led to the withdrawal of the results for Spain, reflect a wider issue. ILSAs have two goals: to develop metrics to compare student performance between countries and to measure trends over time. While trends over time can only be accurately estimated with a constant metric (or a set of anchor items which remain constant), meaningful comparisons between countries require metrics which adapt to the changes taking place in most education systems. Different ILSAs seem to have made different choices to address this issue: while PISA places more emphasis on innovation, TIMSS and PIRLS take a more conservative approach.

In the 2018 cycle, PISA incorporated items from PISA for Development and implemented an “adaptive” approach. Presumably, these changes were adopted to make PISA more sensitive at the lower levels of student performance in order to provide more detailed information to the growing number of countries joining the survey, most of them with low levels of performance. This raises the broader issue

as to whether an overemphasis on innovation could lead to a lack of reliability of the comparisons between cycles and trends over time.

When trends over time are compared between PISA and TIMSS and PIRLS, it seems that the former is less sensitive to changes over time, particularly after major changes were introduced in methodology 2015 and 2018. Previous studies comparing how countries perform in both PISA and TIMSS have shown that the averages for countries are strongly correlated both in 2003 (Wu 2010) and 2015 (Klieme 2016). However, when changes over time are analysed for countries participating in both surveys then it becomes clear that since 2015 PISA started to show declines in performance for countries which showed improvements in TIMSS (Klieme 2016). The conclusion from this study is that this is the consequence of a new mode of assessment in PISA 2015.

The results from PISA 2018 seem to support this view, since only 4 countries improve in reading between 2015 and 2018, while 13 decline and 46 remain stable. When longer periods are considered, only 7 countries/economies improve in all 3 domains, 7 decline in all domains, and 12 show no changes in any of the 3 domains. When only OECD countries are considered, PISA detects no major changes between 2000 and 2018. The OECD concludes that the lack of progress detected by PISA is the result of countries not implementing the right policies (OECD 2019c). However, data from PIRLS and TIMSS show clear improvements overall and, more importantly, in many of the same countries over similar periods. This suggests an alternative explanation: that PISA may not be sensitive enough to detect positive trends, particularly after the methodological changes introduced in 2015 and 2018.

It is beyond the scope of this chapter to analyse in detail which of the methodological changes that PISA implements in each cycle may obscure the real changes that are taking place in education systems. The available evidence seems to suggest that changes adopted to improve sensitivity at lower levels of student performance, may have been made at the expense of the consistency required to detect changes over time. Whatever the reason may be, it seems clear that PIRLS and TIMSS seem much more sensitive to the changes that are taking place over time than PISA. Thus, they seem to be more useful for countries as thermometers which can detect meaningful changes in student performance.

Finally, in many countries governments evaluate their education systems through the evidence provided by ILSAs. By doing this they expose themselves to the huge media impact that international surveys generate. This implies that the results will have major implications about the way particular education policies, reforms or governments are perceived by their societies. Thus, the stakes are very high for governments and policy makers. The case of PISA in Spain is a clear example. In this context, ILSAs must remain accountable when the reliability of the results generate reasonable doubts.

## 4 Conclusions

The evidence from the ILSAs shows that student performance in Spain is lower than the OECD average and has failed to show any significant progress at least from 2000 (when Spain joined PISA) until 2011/2012. The performance of Spanish students seems to be particularly low for maths. Both the low levels of performance and the stagnation over time seem to be explained mainly by the low proportion of Spanish students which attain top levels of performance, according to PISA, PIRLS and TIMSS. The averages for Spain hide huge differences between the 17 regions, which are equivalent to more than one year of schooling.

The stagnation in levels of student performance occurred despite substantial changes in levels of investment in education (which increased until the economic crisis and decreased thereafter), declines in the ratio of students per teacher and increases in teacher salaries. Similarly, large differences revealed between regions are unrelated to these “input variables”.

During this long period of stagnation the education system was characterized by having no national (or regional) evaluations and no flexibility to adapt to the different needs of the student population. The fact that the system was blind to the performance of its students, its rigidity and the lack of common standards at the national level gave rise to three major deficiencies: a high rate of grade repetition and early school leaving, and large differences between regions. These features of the Spanish education systems represent major inequities.

The lack of national evaluations implied that the only information available on how Spanish students perform in comparison to other countries, trends over time and divergence between regions, was provided by PISA (Spain joined in 2000 and has participated in every cycle, with a growing number of regions having an extended sample). As a consequence, the media impact of PISA has been huge. In contrast to other countries, the furore over PISA did not lead to education reforms for over a decade. Thus, governments did not pay much attention to the evidence provided by ILSAs in relation to the poor quality of the Spanish education system and international examples of policies that could help overcome the main deficiencies.

In contrast to other countries, such as Germany, the explanation for the widespread interest in PISA does not seem to lie in the difference between the high expectations and the poor results (the so-called “PISA shock”) (Hopfenbeck et al. 2018, Martens and Niemann 2010). In Spain, the expectations seemed low and better aligned with the PISA results. In fact, PISA results were used to reinforce the misguided view that the Spanish education system prioritized equity over excellence. This seems to be a poor excuse for the mediocre performance of Spanish students, since many countries have shown that improvements in student performance can occur alongside improvements in equity. Furthermore, the high rates of grade repetition, that lead to high rates of early school leaving, represent the most extreme case of inequity that education systems can generate.

This changed in 2013 when an education reform was approved to address the main deficiencies of the Spanish model, including major inequities such as early

school leaving and regional disparities, and also evidence from PISA on the poor levels of performance and the inability of the prevailing model to allow a significant share of top performing students. Implementation started in 2014/15 and had a clear and positive impact on the following: decreased rate of grade repetition, increased enrolment in vocational education and training, and substantial decreases in early school leaving. The impact on student performance is less clear given that the implementation of one of its key elements, i.e. national evaluations, was halted. However, changes in curricular content, the development of evaluation standards and the implementation of evaluations in primary, seem associated with a weak improvement among primary students in maths and science (TIMSS 2015) and a substantial improvement in reading in 2016 (PIRLS 2016).

For secondary students the only information available on the performance on secondary students comes from PISA. In 2015 Spain converged with the OECD average, but this was due to a combination of a weak improvement in the performance of Spain (associated to some extent to the decrease in the rate in grade repetition) and a decline in the performance of the OECD. In 2018 the OECD withdrew the results for reading (main domain) for Spain after some regions complained about anomalies in the PISA scores and the data received. However, the results for science and maths were released despite the concerns raised by the same regions which claimed that they were contaminated by the same problems plaguing the reading scores.

In summary, for a long time PISA received a lot of attention in Spain because it was the only common metric available to compare the performance of Spain with other countries, trends over time and regional differences. However, policy makers did not listen to the evidence on good international practices that could improve student performance and instead became complacent about the poor results obtained by Spanish students hiding behind the excuse of a greater goal: equity. As a consequence, the education system remained substantially unchanged until 2013 when a major reform was approved.

The level of interest and respect that PISA had built in Spain was shaken when the results for main domain in 2018 were withdrawn due to serious inconsistencies and lack of reliability. Since the OECD has provided no explanations so far, the trust on what was considered an international benchmark has been eroded. The available evidence seems to suggest that the underlying causes may have to do with PISA's bet for an innovative approach, including the decision to merge some methodological tools from PISA for Development.

The silence from the OECD has replaced all the noise that has traditionally accompanied each PISA launch. It is important that the trust is re-established so that policy makers can listen to the international evidence that identifies the strengths and weaknesses of education systems, and the good practices that can be applied to each specific context. This can only be accomplished if ILSAs are open and transparent about the potential trade-offs that may occur between innovative approaches which are adopted to capture new dimensions, and the need for consistency over time. In this regard, PISA seems to have followed a riskier approach than TIMSS and PIRLS.

**Authors' ADDENDUM (27 July)**

On the 23rd of July 2020, the OECD published the results for Spain on the main domain (reading) (<https://www.oecd.org/pisa/PISA2018-AnnexA9-Spain.pdf>, retrieved on 27th July 2020). These data were withdrawn before the official launch of PISA in December 2019, due to “implausible response behaviour amongst students” in the new section on reading fluency.

The scores that have been recently published are the same that were sent by the OECD to the Spanish Ministry of Education and all 17 regions after the summer of 2019. What seems surprising is that the OECD still recognizes the “anomalies” in student responses and, more importantly, that the data are not comparable to previous PISA cycles or other countries, since it acknowledges a “possible downward bias in performance results”. It is unclear why the “old” data have been released now, given its major limitations.

The “implausible response behaviour” affects only the new section on reading fluency, but no effort has been made to correct these anomalies or to explain how the *straightlining* behaviour displayed by some students may have affected the whole reading test, given that in 2018 PISA was designed as an adaptive test.

Instead the OECD presents new analyses in an attempt to explain why some students gave patterned responses (all *yes* or all *no*) in the reading fluency section. I quote the main conclusion: “*In 2018, some regions in Spain conducted their high-stakes exams for tenth-grade students earlier in the year than in the past, which resulted in the testing period for these exams coinciding with the end of the PISA testing window. Because of this overlap, a number of students were negatively disposed towards the PISA test and did not try their best to demonstrate their proficiency*”.

It is unclear what the OECD means by “high-stakes exams”. According to Spanish legislation, at the end of lower secondary all regions in Spain have to conduct diagnostic tests, which do not have to conform to national standards. These regional diagnostic tools are based on a limited sample of students and do not have academic effects. Although the degree of overlap of the sample for these end-of-lower secondary tests and the PISA sample is often unknown, some of the regions ensure that no school participates in both. This is the case of Navarra, a region which has suffered one of the most marked declines in reading performance according to PISA. It seems reasonable to conclude that either the degree of overlap is small or non-existent.

As in most countries, students at the end of lower-secondary undertake exams for each subject (during and at the end of the academic year). If the OECD refers to these tests, it is also difficult to understand why the argument focuses exclusively on 10th grade students. The PISA sample includes 15-year-olds, irrespective of the grade. Since grade repetition in Spain is one of the highest of the OECD, in 2018 the sample included 69.9 students in 10th grade, 24.1 in 9th grade, and 5.9 in 8th grade. Thus, 30% of the students in the national sample were not on 10th grade, and did not take any of the tests mentioned, as all analyses assume. Among poor performing regions, the proportion of 15-year-olds who are not in 10<sup>th</sup> grade increases to almost half of students. This well-known fact not only reduces the overlap between the PISA

sample and that of any type of end-of-secondary exams even further, but represents a challenge to all the analyses presented.

The OECD's analyses attempt to link the dates of "high stakes exams" in 10th grade and the date of the PISA test, with the proportion of students who reported doing little effort in the PISA test and the proportion of "reading fluency anomalies". Beyond detailed methodological issues, these analyses focus on the new section on reading fluency where the anomalies were detected. As explained by the OECD, this section is the easiest of the reading test, and students showing *straightlining* behaviour continued onto more difficult items and responded according to their true level of proficiency. Thus, it is unclear why any potential overlaps between PISA and other tests, would have affected the behaviour of students when responding to the first and "easiest" section, but not during the rest of the test which had more demanding questions. Furthermore, unless this first section has a major impact on the whole reading test, the analyses do not address why student performance in reading has apparently declined in Spain.

The OECD recognizes indirectly that there is a problem with the new section on "reading fluency" since it states that "the analysis of Spain's data also reveals how the inclusion of reading fluency items may have strengthened the relationship between test performance and student effort in PISA more generally. The OECD is therefore exploring changes to the administration and scoring of reading fluency items to limit the occurrence of disengaged response behaviour and mitigate its consequences".

In conclusion, it is difficult to understand why data which the OECD defines as unreliable and non-comparable have been published after more than 8 months, with no attempt to correct anomalies. The biases will not make the data useful in Spain. More generally, PISA participating countries deserve a credible explanation about the methodological problems encountered in PISA 2018 with the new section in reading fluency and the adaptive model exported from PISA for Development.

## References

- Adonis, A. (2012). Education. *Reforming England's schools*. Biteback Publishing Ltd, London.
- Ball, S. J. (2013). *The education debate* (2nd ed.). Bristol: The Policy Press.
- Cordero, J. M., Crespo, E. & Pedraja, F. (2013). Educational achievement and determinantes in PISA: A review of literature in Spain. *Revista de Educación*, 362 <https://doi.org/10.4438/1988-592X-RE-2011-362-161>.
- Cordero, J. M., Cristóbal, V., & Santín, D. (2018). Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS. *Journal of Economic Surveys*, 32(3), 878–915.
- Delibes, A. (2006). La gran estafa, el secuestro del sentido común en la educación. Grupo Unisón producciones.
- El Mundo. (2019a). *La Comunidad de Madrid pide a la OCDE que retire todo el informe PISA por errores de un calibre considerable: Toda la prueba está contaminada*. November 29, 2019.
- El Mundo. (2019b). *Las sombras de PISA: ¿hay que creerse el informe tras los errores detectados?*. December 2, 2019.



- El País. (2019). *Madrid pide que no se publique ningún dato de PISA porque todo está contaminado*. November 30, 2019.
- Enkvist, I. (2011). *La buena y la mala educación*. Madrid: Ejemplos internacionales. Ediciones Encuentro S.A.
- Gustafsson, J. E., & Rosen, M. R. (2014). Quality and credibility of international studies. In R. Strietholt, W. Bos, J. E. Gustafsson, & M. Rosen (Eds.), *Educational policy evaluation through international comparative assessments*. Münster: Waxmann.
- Hanushek, E. A. & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the Economics of Education* (Vol. 3, pp. 89–200). North Holland: Amsterdam.
- Hanushek, E. A., & Woessmann, L. (2014). Institutional structures of the education system and student achievement: a review of cross-country economic research. In R. Strietholt, W. Bos, J. E. Gustafsson, & M. Rosen (Eds.), *Educational Policy Evaluation Through International Comparative Assessments*. Münster: Waxmann.
- Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J., & Baird, J. A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353.
- Ikeda, M. & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences OECD. *Journal: Economic Studies* 2013/1 [https://doi.org/10.1787/eco\\_studies-2013-5k3w65mx3hnx](https://doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx).
- INEE. (2014). Sistema estatal de indicadores de la educación. MECD.
- Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics*, 81(1), 226–244.
- Johansson, S. (2016). International large-scale assessments: what uses, what consequences? *Educational Research*, 58, 139–14. <https://doi.org/10.1080/00131881.2016.1165559>.
- Klieme, E. (2013). The role of large-scale assessments in research on educational effectiveness and school development. In M. von Davier, E. Gonzalez, I. Kirsch, & K. Yamamoto (Eds.), *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht: Springer.
- Klieme, E. (2016). TIMSS 2015 and PISA 2015 *How Are They Related On The Country Level?* DIPF Working Paper, December 12 2016.
- La Razón. (2019). Madrid llama chapucera a la OCDE por el informe PISA. December 02, 2019.
- Lockheed, M. E., & Wagemaker, H. (2013). International large-scale assessments: thermometers, whips or useful policy tools? *Research in Comparative and International Education*, 8(3), 296–306. <https://doi.org/10.2304/rcie.2013.8.3.296>.
- LOGSE: Ley Orgánica 1/1990 de Ordenación General del Sistema Educativo (1990) BOE 238 (4 octubre 1990): 28927–28942.
- LOMCE: Ley Orgánica 8/2013 para la mejora de la calidad educativa. BOE 295 (10 diciembre 2013): 12886.
- Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics*, 94(2), 596–606.
- Martens, K. & Niemann, D. (2010). Governance by comparison: How ratings and rankings impact national policy-making in education. TranState Working Papers, 139, University of Bremen, Collaborative Research Center 597: Transformations of the State.
- Martin, M. O., Mullis, I. V. S., Foy, P. & Hooper, M. (2016). *TIMSS 2015 International results in science*. Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2016) *TIMSS 2015 International results in mathematics*. Boston College.
- Mullis, I. V. S., Martin, M. O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 International results in reading*. Boston College.

- OECD. (2016a). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.
- OECD. (2016b). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: OECD Publishing.
- OECD. (2016c). *Education at a Glance*. Paris: OECD Publishing.
- OECD. (2017a). *Education at a Glance*. Paris: OECD Publishing.
- OECD. (2017b). *PISA for Development Assessment and Analytical Framework: Reading, Mathematics and Science*. Paris: OECD Publishing.
- OECD. (2019a). *OECD Skills Outlook 2019: Thriving in a digital world*. Paris: OECD Publishing.
- OECD. (2019b). *OECD Skills Strategy 2019: Skills to Shape a Better Future*. Paris: OECD Publishing.
- OECD. (2019c). *PISA 2018 Results (Vol. 1): What Students Know and Can Do*. Paris: OECD Publishing.
- OECD. (2019d). *PISA 2018 Results (Vol. 2): Where All Students Can Succeed*. Paris: OECD Publishing.
- Striethold, R., Bos, W., Gustafsson, J. E., & Rosen, M. (Eds.). (2014). *Educational Policy Evaluation Through International Comparative Assessments*. Münster: Waxmann.
- Villar, A. (Eds.). (2009). *Educación y Desarrollo. PISA 2009 y el sistema educativo español*. Fundación BBVA.
- Wert, J. I. (2019). *La educación en España: asignatura pendiente*. Sevilla: Almuzara.
- Wu, M. (2010). Comparing the similarities and differences of PISA 2003 and TIMSS. OECD Education Working Papers, No. 32. OECD Publishing, Paris. <http://dx.doi.org/10.1787/5km4psnm13nx-en>.

**Montserrat Gomendio** worked previously at the OECD from 2015–2019, first as Deputy Director for Education and then as Head of the OECD Centre for Skills since it was created in June 2017. The Centre for Skills is a horizontal Centre which carries out interdisciplinary projects linking the impact of megatrends on the changing demands by the labour market with the responsiveness required from lifelong learning systems to equip people with the right bundles of skills.

Ms. Gomendio was Secretary of State for Education, Vocational Training and Universities at the Spanish Ministry of Education, Culture and Sports, from January 2012 to May 2015. During this period of time the Spanish Government undertook an education reform.

Previously she enjoyed a productive career in academia. She worked first as researcher and lecturer at the University of Cambridge (U.K.) and then joined the Spanish Research Council (CSIC) in 1991, where she became a Research Professor. In 1997 she was appointed Director of the Natural History Museum, and from 2003 until 2004 served as Vice-President for Organisation and Institutional Relations of the CSIC.

Ms. Gomendio, a Spanish national, holds a General Management Program degree from the IESE Business School, a PhD from the University of Cambridge (U.K.), and a BSc in Biology from the Complutense University of Madrid.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

