

**DIRE STRAITS-EDUCATION REFORMS**  
**IDEOLOGY, VESTED INTERESTS**  
**AND EVIDENCE**

**MONTSERRAT GOMENDIO**  
**AND JOSE IGNACIO WERT**





<https://www.openbookpublishers.com>

© 2023 Montserrat Gomendio and José Ignacio Wert



This work is licensed under an Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). This license allows you to share, copy, distribute and transmit the text; to adapt the text for non-commercial purposes of the text providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

Montserrat Gomendio and José Ignacio Wert, *Dire Straits: Education Reforms, Ideology, Vested Interests, and Evidence*. Cambridge, UK: Open Book Publishers, 2023, <https://doi.org/10.11647/OBP.0332>

Further details about the CC BY-NC license are available at <http://creativecommons.org/licenses/by-nc/4.0/>

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at <https://archive.org/web>

Digital material and resources associated with this volume are available at <https://doi.org/10.11647/OBP.0332#resources>

ISBN Paperback: 978-1-80064-930-9

ISBN Hardback: 978-1-80064-931-6

ISBN Digital (PDF): 978-1-80064-932-3

ISBN Digital ebook (EPUB): 978-1-80064-933-0

ISBN Digital ebook (AZW3): 978-1-80064-934-7

ISBN XML: 978-1-80064-935-4

ISBN HTML: 978-1-80064-936-1

DOI: 10.11647/OBP.0332

Cover image: Kimberly Farmer, A collection of books (2017), <https://unsplash.com/photos/IUaaKCUANVI>. Cover design: Jeevanjot Kaur Nagpal.

# 5. Does the Evidence Count?

---

## 5.1. What Can We Learn from Top-Performing Countries?

When the results from ILSAs are made public, some countries seem surprised to find themselves as top performers, others shocked to realise that they perform worse than they expected, and others still seem oblivious. We will try to understand the different reactions and the lessons learned both from top and low performers. We will start by considering what top performers have taught the rest of the world (and themselves).

### Finland

The very first PISA cycle took place in 2000 and included a relatively small number (thirty-one) of mostly OECD countries; thus, the magnitude of the differences found between countries was smaller than in later cycles when a larger number of more diverse countries participated. Finland emerged as the top performer in reading (the main domain), but not in mathematics and science, domains in which Asian countries, such as Japan and Korea, were already top performers (OECD, 2001).

Since then, Finland has become a legend in educational circles, with international organisations, academics, policymakers and unions trying to understand what aspects of its education system have led to such high performance among students. The interest in learning from Finland's success has grown to such an extent that it has become a common destination for 'education-tourism', with policymakers flocking in to see for themselves how this miracle was achieved. But the truth is that Finland did not expect such high levels of student performance and had

difficulty interpreting why it was so successful. A report co-sponsored by the OECD and the Institute for Educational Research at the University of Jyväskylä noted that the PISA results have been both a source of “great joy” and a “somewhat puzzling experience” (Grubb *et al.*, 2005). Thus, all interpretations focused on analysing Finland *after* its success was revealed.

Even the OECD sent teams to Finland to understand what made the education system so successful. The main conclusions from these first visits identified several factors, the most salient of which were comprehensive schooling, teacher quality, school autonomy and commitment to equality (Grubb *et al.*, 2005; Valijarvi *et al.*, 2002). In this way, high levels of equity were seen from the very beginning as a key to success in terms of high student performance (i.e. quality), rather than as a different dimension of the education system. This is somewhat surprising since, according to PISA 2000 results, Finland performed worse in terms of equity than countries such as Japan, Spain, Mexico or Korea (OECD, 2001). We will come back to the links between these two dimensions of the education system, i.e. quality and equity, later.

Most analyses highlighted the fact that teachers in Finland go through a highly selective process to enter university and receive demanding, high-quality training; as a consequence, the teaching profession is highly regarded, teachers are given independence to innovate, and there is a well-functioning system of professional development, although salaries are not high in relation to other countries (Grubb *et al.*, 2005; OECD, 2014a). The fact that teachers in Finland enjoy more independence than in other countries has rapidly led to the idea that “the secret to Finnish education is trust”, a conclusion that the OECD maintains to this day, despite Finland’s decline in performance after the first PISA cycle (Schleicher, 2018 and 2020).

The teachers’ unions rapidly capitalised on this narrative, which became very popular. The Finnish unions argued that the success of the education system was due to the high skills of teachers thanks to the quality of their university education, their level of autonomy and a presumed student-centred approach. They also used the fact that levels of investment and teacher salaries were lower than in other European countries to demand more resources (Rautalin and Pertti, 2007). The emphasis on the quality of teachers obviously minimises the impact of

education reforms, as well as the role of families in supporting their children. Thus, while unions were happy to acknowledge the high quality of teachers, they ignored other aspects and used the fact that levels of investment were comparatively low to reinforce their demands, rather than acknowledging that the good results most likely revealed an efficient use of resources.

Over time the narrative about the success of Finland crystallised in an ideal of education systems that most unions supported and that has had a major impact worldwide. This ideal claims first and foremost that teachers should be *trusted*, an attractive turn of phrase which actually means that teachers should not be evaluated for their work and should be given autonomy to decide how and what to teach in exchange for no accountability. In the most extreme cases, it is argued that students should not be evaluated either, except continuously by their own teachers, because there is a risk that student performance could be used to evaluate teachers indirectly. This ideal also encompasses the entire range of so-called 'comprehensive policies', which goes much further than rejecting tracking until the age of sixteen, by denouncing all forms of diversification as segregation. This overstretched interpretation of Finland's success brought the OECD and the unions closer together, since they found common ground in articulating a narrative which saw trust in teachers as paramount for education reforms to prosper. In a book which attempts to give a personal view of the main PISA findings, Andreas Schleicher states:

policy makers need to build strong support about the aims of education reform and engage stakeholders, especially teachers, in formulating and implementing policy responses [...] many of the countries with the strongest student performance also have strong teachers' unions [...] [...] the higher a country ranks on the PISA league tables, the more likely it is that the country works constructively with its teachers' organizations and treats its teachers as trusted professional partners. (Schleicher, 2018)

The rationale that the most effective way to prevent unions from derailing reforms is to give them a relevant role in designing those reforms is simple and clear, but in our view misses the point entirely, because the vested interests of teacher unions are often not aligned with what is required to improve student performance. The claims of a link between student performance and the strength of unions are not

supported by data. Quite the opposite (Moe and Wiborg, 2017). While it is true that myths tend to drift further and further away from hard evidence, because the strength of the narrative at some point becomes more powerful than the evidence itself, it is still surprising that the OECD has fed this narrative in the absence of robust data.

A whole school of thought (and industry) has been created around this notion, so the literature is vast. We will quote just one well-known supporter of this interpretation, who makes crystal-clear the political underpinnings of the trust doctrine:

this book does not suggest that tougher competition, more data, abolishing teacher unions, opening more charter schools, or employing corporate-world management models in education systems will bring about a resolution to these [education] crises—quite the opposite. The main message of this book is that there is another way to improve education systems, one that is different from the market-based reform ideology [...] it includes improving the teaching force, limiting student testing to a necessary minimum, placing responsibility and trust before accountability, investing in equity in education, and handing over school- and district-level leadership to experienced education professionals (Sahlberg, 2021)

Soon after Finland was identified by PISA as an education “superpower”, its performance started to decline, a process which has steadily continued up to the last PISA cycle. Two mutually non-exclusive hypotheses may help us explain the inconsistency in the construction of a powerful narrative around Finland as a role model for the rest of the world, as its actual performance was deteriorating. One possibility is that the explanations of Finland’s success in 2000 are either wrong or have gaps because, although the country has continued to implement what were identified as ‘good practices’, its performance has worsened, mainly due to an increase in the proportion of students performing poorly in all domains, and an increased impact of socio-economic background on student outcomes (Ahonen, 2021; Rautalin, 2018).

Another possibility is that the policies which were effective in a very specific context, i.e. when Finland had a rather uniform population, a deeply egalitarian society and a network of small schools, did not work as the Finnish population became more diverse with the arrival of immigrants (Harju-Luukkainen and McElvany, 2018). If this is the case, then sticking to supposedly good practices while basking in its

popularity as an education superpower may have prevented Finland from making the necessary reforms (Rautalin and Pertti, 2007).

One of the problems with the interpretations of causal relationships is that they often do not take into account “time lags” (Oates, 2010). Since PISA assesses fifteen-year-olds, the impact of changes which affect the whole education system (for example, improved teacher training) on student performance at this age may go back as far as ten years which is more or less the amount of time that students have spent in school, although other policies may have a visible impact within a shorter timeframe. Thus, in some cases student performance according to PISA has more to do with the education system that has been in place for the last ten years or longer, than with current models. In Finland the high standards for teaching qualifications were set in 1979, when all teachers (including primary teachers) were required to have a master’s degree, in addition to a undergraduate degree (Ahonen, 2021; Oates, 2015; Sahlgren, 2015). But the roots of the Finnish success in reading may extend further back. In Finland improvements in student performance were mostly the result of reforms that took place in the 1960s, which were centralised, at a time when strong control by the state was exerted over accountability, a detailed national curriculum and the inspection system (Frassinelli, 2006; Oates, 2010 and 2015; Sahlgren, 2015; Simola, 2005). Relaxation of these measures only took place once curriculum coherence and a highly skilled teaching force were in place. The impact of previous education models and of early reforms is strongly supported by evidence from international test scores available before PISA started, which clearly show that Finnish students were performing poorly and below other European countries in 1975, and improved rapidly until 2000, the year when the first PISA cycle took place (Hanushek and Woessmann, 2015).

Another common problem with the interpretation of causal relationships using PISA data is that, as we have seen, it cannot disentangle the impact of schooling from that of family and culture, because of how student performance is assessed. Finnish society has traditionally placed great emphasis on literacy and early family learning plays a very important role (Aunio *et al.*, 2006; Oates, 2010 and 2015; Sahlgren, 2015). In addition, historical and cultural factors have also contributed to the teaching profession’s high social status, since teachers

have played an important role in the national project to create a Finnish-speaking culture (Sahlgren, 2015).

The inconsistencies in the interpretations of the Finnish success may have to do with the fact that it was a top performer in the very first PISA cycle, when no conclusions had yet been elaborated about good practices. However, the growing disconnect between the country's declining student performance over time and the increasing impact of a narrative focused on comprehensive policies and equity, as well as autonomy and a lack of accountability for teachers, is worrying. To be more precise, although Finland remained a top performer among OECD countries for several cycles, its success was ultimately eclipsed by East Asian countries.

### East Asian Countries

As its performance continued to decline, Finland eventually ranked below other OECD countries as well (in PISA 2015 and 2018 Finland ranked fourth and seventh respectively in reading, fifth and sixth respectively in science and thirteenth and sixteenth respectively in mathematics) (OECD, 2016b, 2019c). Given the extraordinary and consistent performance of some regions of China, and countries such as Hong Kong, Japan, Korea, Singapore and Taiwan in all ILSAs, what lessons have been drawn from them, and to what extent are these compatible with the narrative about Finland?

Surprisingly, no comparable narrative to that of Finland has been built around the outstanding and sustained success of this group of countries. Since most of these countries emerged as top performers in other ILSAs well before they joined PISA (see Chapter 4), it is obvious that (as was the case in Finland) they did not improve by adopting the good practices recommended by the OECD.

Some of them, such as Singapore and Korea, do share some historical background with Finland: these countries had few natural resources, so they decided early on to focus on developing their human capital in order to improve their economies and societies. In addition, they were at a turning point in their histories when building a national identity through education was crucial. Their ability to plan in the long term, to implement consistent policies over time, and to adapt them as their



education systems gradually improved led to the most successful cases in the history of education reforms.

As the data from the survey of adult skills (PIAAC) shows, adult populations in both Singapore and Korea had very low levels of skills around sixty years ago compared to all OECD countries (OECD, 2016d). However, rapid improvements in their education systems led to the acquisition of increasingly higher levels of skills over time, until younger generations became top performers in those ILSAs in which they participated. Thus, in about fifty years Singapore and Korea had evolved from being the underdogs to becoming the best players in the world. From then on, this group of East Asian countries was seemingly in another galaxy, with student performance continuing to improve with no apparent ceiling while most OECD countries remained stagnant or declined. As expected by educational reformists, improved student outcomes led to such exceptional economic growth and social prosperity that the phenomenon is commonly referred to as the “East Asian Miracle”.

The capacity to plan in the long term is due to several factors. Some of these countries are semi-democracies where the same political party has been in power for decades, such as Singapore, or authoritarian regimes without real political competition, such as China. Others are full democracies with different political parties alternating in power, but there seems to be a high degree of consensus ranking education as a top priority, and a shared understanding of which policies lead to better outcomes, with no major ideological divisions on education. In most of these countries, teachers’ unions either do not exist or have no real power to veto reforms, shut down schools or mobilise masses against them. The reforms which led to rapid improvements of these education systems started recently and from a pragmatic perspective: an understanding of the power of education to transform societies, as well as to propel economic growth and enhance wellbeing.

The undeniable success of these countries makes it clear that governments can and do get it right, and that the lack of capacity of teachers’ unions to veto reforms in order to defend their vested interests greatly facilitates the improvement of education systems. In this respect, the conclusions from countries in East Asia come into conflict with the narrative built around Finland.

Countries in Europe and North America were shocked at first, but then found solace in the idea that differences in success were mainly due to cultural differences, with families in Asia putting undue pressure on their children to achieve high levels of performance through endless hours of rote learning, leading to high levels of stress (Takayama, 2017). This explanation fails to take into account the fact that students in these countries outperform others not just in tasks that require memorising knowledge, but in complex problem-solving, critical thinking and collaborative tasks as well (OECD 2016b, 2019c).

Although it is widely recognised that families in Asia consider their children's education a high priority, the levels of stress differ substantially between countries. In Korea, access to a few prestigious universities is regarded as crucial for professional success, so there is a huge bottleneck for this level of education, and families respond by sending their children to academies (*hagwon*) (OECD, 2014d) after school hours to enhance their learning opportunities. The government is aware of the pressure on students and has devised a number of policies to try to ameliorate it. However, these are unlikely to work as long as the bottleneck to enter the few top universities continues. In contrast, children in Singapore, Hong Kong or Taiwan are not subject to such high levels of stress and still achieve excellent results. This success seems to come in each instance from a very efficient school system, in addition to the high value that families place on good academic results, which they instil in their children, and the responsibility that students assume for their own performance.

This group of countries clearly shows that education systems can improve very fast when societies as a whole identify education as *the* priority for the future of their country, families become involved in inculcating the value of good educational outcomes in their children, and governments have the capacity to plan in the long term, because the chances that reforms will be blocked are minimised due to the lack of veto power from unions and the absence of profound ideological divides. But the question still remains: given that in this context governments have greater leeway in making decisions, what are the policies that they have put in place?

Education systems in these countries show a great degree of variation, so we will concentrate on the commonalities which are known to have

a positive impact on student outcomes. We will also analyse whether the policies that were identified as having a relevant role in the case of Finland have been implemented by countries in East Asia.

We will start with a brief historical account of the education reforms in Singapore and South Korea. This choice is based on the fact that both education systems have improved very rapidly over the last fifty years and successive governments have made the roadmap for each stage explicit, providing a rare and unique opportunity for us to understand which policies were implemented and how they changed over time.

### Singapore

In Singapore, the Ministry of Education was first established in 1955 by the colonial government before this city-state became independent (Doraisamy, 1969; Norrudin, 2018). The first step was to provide six years of free primary education to all children, which required a substantial investment in the construction of schools, their equipment, as well as the training and hiring of teachers. Before then, the education system consisted mostly of private schools teaching in Chinese, Tamil and Malay, besides the government-aided mission schools that taught in English. In the second half of the 1950s, the aim was to establish a national education system where all schools would follow the same curriculum, use the same textbooks, have national exams, and be taught via a bilingual model of instruction (English and the mother tongue). To achieve this goal, schools would only receive government funding if they complied with a set of regulations, a rule which was initially rejected by the Chinese schools.

After independence, in the 1970s, new solutions were sought to address the main weaknesses of the system: high drop-out rates in primary and secondary education, and low levels of literacy (Lee, 2008; Norrudin, 2018; Soon, 1988; Turnbull, 2009; Wilson, 1978). The New Education System (NES), introduced in 1979, implemented three tracks both in primary and secondary schools according to students' academic ability, so that they could learn at different paces according to their needs, and maximise their chances of reaching their full potential and obtaining jobs.

This streaming system generated much controversy and was debated in parliament for four days; some feared that children in the low-performing tracks would suffer from reduced access to university, but others supported the measure because they felt it would minimise drop-out rates, which are far worse outcomes. In order to balance the risks and advantages, the system was designed flexibly so that streaming started three to four years after the beginning of primary school, and students who performed better or worse than expected would be moved to other tracks. Two tracks were mostly academic, and the third was entirely vocational. This streaming system was implemented in secondary schools in 1980 and led to huge improvements: the proportion of students who passed the national exams increased from 60% to 90% and drop-out rates decreased both in secondary education (from 36% to 6%) and in primary education (from 29% to 8%). The streaming system has subsequently been modified to ensure that all students reach their full potential. In 2017 it was replaced in primary education with ability grouping. But tracks remain in place today in lower-secondary education, with students following different tracks from the age of twelve, depending on the grades they obtain in the Primary School Leaving Examination (PSLE). The success of this model in an international context is reflected in the fact that Singapore emerged as a top performer in TIMSS in 1995 (Harmon *et al.*, 1997) and has remained a top performer in different ILSAs ever since.

From the very beginning, teacher quality was identified as a priority. The Teachers's Training College was established in 1950 to train primary teachers but later expanded to secondary teachers (Loh and Hu, 2019; Norrudin, 2018). In 1973, it was transformed into the Institute of Education and later became the National Institute of Education, which offers high-quality undergraduate and diploma programmes and professional development for teachers, and carries out research on efficient teaching practices. Professional development has been elaborated to the extent that teachers can opt for different trajectories depending on the career paths available to them: teachers, mentors, principals or policymakers. Such importance is placed on the lifelong learning of teachers that teachers are entitled to 100 paid hours of professional development each year (Bautista *et al.*, 2015). The emphasis on teachers' professional development was taken one step further

in 2010, when the Ministry of Education introduced the requirement that every school becomes a 'professional learning community' (PLC), with every teacher participating in learning activities with colleagues (Academy of Singapore Teachers, 2012; Dimmock and Tan, 2015). As a consequence, teachers spend less time in the classroom and more time engaged in collaborative activities with other teachers. In order to free up time for all these activities, class size is larger than in most OECD countries. In other words, there is a clear trade-off between time spent in professional development and class size.

Teachers enjoy high social prestige and receive competitive salaries. The main contribution of this highly skilled teaching force to the design of education policies takes place through direct collaboration with the government: teachers with experience in the classroom may take up positions at the Ministry of Education and, in this way, they participate in improving the education system. Despite the high skills of teachers and principals, they do not enjoy high levels of autonomy in schools and classrooms, in terms of resources, curriculum or assessments, which remain under central control (Dimmock, 2011; Dimmock and Tan, 2015; Gopinathan and Deng, 2006).

Assessments are national, high-stakes examinations at the end of key learning stages (primary fourth and sixth grades, GCE "O" levels in lower-secondary school, and GCE "A" levels in upper-secondary school), and the results are used to place students in different tracks from the age of twelve (Tan, Chow and Goh, 2008). The fact that students follow different pathways from an early age does not prevent them from attaining high levels of performance.

## South Korea

The recent history of South Korea is the other main example of an education system which has managed to evolve in a few decades from what was basically an illiterate population to a consistent top performer in all international surveys. In 1945 only 22% of adults were literate, by 1970 adult literacy had increased to 87% and by the late 1980s it was estimated at 93%. South Korea has emerged as a top performer in all international surveys since 1975 (Hanushek and Woessmann, 2015).

In the 1950s, after the Korean War, South Korea was one of the poorest countries in the world, with an economy based largely on agriculture. Today it is the twelfth largest economy and is an advanced, high-tech nation. This amazing rate of economic development was clearly associated with the incredible success of its education system, as is the case in Singapore. The main difference is that while Singapore is a small city-state, South Korea is a large country. Thus, the latter shows that fast improvements in education can also be accomplished at a large scale. Furthermore, rapid improvements in student performance can occur alongside fast expansion of access to education for large numbers of students (Tucker, 2019).

In the 1950s, the decision-making power was taken away from local boards and education became centralised. The Ministry of Education became responsible for the management of schools, allocation of resources, curriculum development, textbook guidelines and teachers, thus concentrating most resources and decision-making power centrally (KEDI, 2015; KICE, 2012; Ministry of Education, 2017 and 2020). Under the influence of the US occupation, South Korea adopted a similar structural design: six years of elementary education, six years of secondary education and three years of high school with different tracks from the age of fifteen. In the 1950s, elementary education was made compulsory for all children. Later, in 1985, compulsory education was extended by a further three years, and was eventually extended until the end of lower-secondary education (fifteen years). The expansion of access to schools took place so rapidly that the number of high schools increased from 640 in 1960 to 2,218 in 2007, while the number of students enrolled increased from 273,434 in 1960 to 2.3 million in 1990 (data provided by Ministry of Education Korea, 2017 and 2020).

Teaching is tightly regulated by the government, which sets high standards (Kang and Hong, 2008). Entrance into the teaching profession is highly competitive. As in other East Asian top-performing countries, teachers are recruited from the top third of each cohort of graduates (top 5% in Korea, top 30% in Singapore and Japan) (Barber and Mourshed, 2007). They receive high-quality training at university and must take a very selective Teacher Employment Examination to obtain a tenured position in a public school. Teachers enjoy high salaries and the social prestige that they deserve. Professional development for teachers

is well-developed and of a high quality, and good performance at in-service training is linked to promotion and pay rates. After three years on the job, teachers must complete an additional training programme to earn a higher-level Grade I Teacher Certificate. Teachers' unions were illegal until 1999, when a new law allowed their existence as part of the membership negotiations held between South Korea and the OECD.

The national curriculum is updated every ten years by government and there are strong accountability measures in place. All schools' performances in Korean national assessments are publicly available since 2008 (OECD, 2012a, 2014d and 2016b).

Competition over admission into a few top universities is fierce, so families put pressure on their children to achieve high levels of performance and are eager to invest in their children attending the *hagwons* (private academies) (Kim and Lee, 2001 and 2010). When we visited Korea, we learned that the government had implemented curfews at ten p.m. to prevent students from spending long hours studying at night, although students in *hagwons* became quite effective at hiding when a patrol turned up. Another measure introduced by the government in a further attempt to lower levels of stress and promote 'happy education' is the 'free semester', when students do not have to take exams. In order to relieve the pressure on students, in the 1980s the Ministry of Education implemented reforms aimed at increasing university enrolment. As a result, although secondary schools do offer vocational tracks, the high return of attendance top universities led to a dramatic expansion of admissions and a high rate of university enrolment (Park and Jang, 2014). Between the early 1980s and the mid-2000s the tertiary gross enrolment ratio increased fivefold, with the number of students increasing from 539,000 in 1980 to 3.3 million in 2015.

However, this strategy backfired, when the financial crisis of 2008 led to large numbers of young graduates losing their jobs. According to a survey from the Korea Research Institute for Vocational Education and Training, in 2013 nearly four in every ten young workers were overeducated. As the returns of a university degree declined, so did the number of students attending university, while the government started to promote vocational education and training as well as apprenticeships.

## Finland and East Asian Countries: Similarities and Differences

The examples of Singapore and South Korea reveal how education systems can improve, rapidly transforming illiterate populations into the world's top-performing students. It is beyond the scope of this book to provide a detailed description of the education systems in other top-performing East Asian countries, which have a longer and more complex history. From what we have seen so far, these countries share one clear feature with Finland: an outstanding teaching force, which is the result of highly selective processes to ensure that the best-performing students enter university to study education, that they receive high-quality training (on subject content as well as pedagogy), that they are evaluated for their performance, and that they continue to upskill and reskill through professional development. This conclusion can be safely expanded to other East Asian countries. As is the case in Finland, in this group of countries teachers enjoy high social prestige. But this is where the similarities seemingly end.

Contrary to the myth of Finnish success, trust in teachers in East Asia does not mean that the system is blind to what teachers do, or that most responsibilities are transferred. In fact, in all East Asian countries teachers are evaluated based on the results of their students in assessments, to a much larger extent than in most OECD countries (OECD, 2019d). Thus, teachers are trusted because they are highly skilled and their teaching practices are very effective. In addition, high curricular standards are set by government, as well as demanding assessments for students, both of which tend to be under strong central control.

While individual teachers can play a significant role in collaborating with government to continuously improve education systems, unions play either little or no role. This is also the case in Japan, perhaps the East Asian country where unions have been present for longest, although they face stricter constraints than in Western democracies. As local public officials, they are denied the right to strike and the right to collective bargaining (Araki, 2002; Aspinall, 2017). It is therefore a fact that in East Asian countries education systems have rapidly improved their performance in the absence of major opposition from unions, but with the collaboration of teachers.



If we analyse in detail other aspects of the Finnish myth, we find further discrepancies. According to this narrative, one of the main elements in the Finnish success is school autonomy. However, teachers and principals in Singapore and China have low levels of autonomy in the use of resources, curricula, assessment and appointment of teachers, and Korea and Japan also have low levels of autonomy in terms of use of resources; in contrast, Hong Kong and Taiwan have high levels of autonomy in all of the above aspects, which is probably because, for historical reasons, many of the schools are privately managed (OECD, 2016c; Tan, 1997). Thus, high levels of school and teacher autonomy do not seem to be a necessary prerequisite for becoming a top performer, even in a country with an exceedingly high-skilled teaching force, like Singapore.

The other major element of the Finnish success narrative is comprehensive education and equity. The case of Singapore shows that ability grouping in primary education and early tracking into different academic/vocational trajectories from the start of lower-secondary education has not in any way prevented rapid educational progress, to the extent that the country has become a top performer in just a few decades. In fact, the case of Singapore shows that on this journey, streaming has been very efficient in preventing drop-out rates and allowing all students to progress at their own pace, while avoiding grade repetition, which is non-existent or very rare in all East Asian top performers.

This case clearly shows that the potential risks associated with early tracking can be prevented by achieving a good balance between the different tracks and student performance, allowing the flexibility to move to other tracks when a student's level of performance changes, and ensuring that all tracks equip students with solid foundations. In fact, when early tracking is designed in this way, the data from Singapore obtained by PISA show that it does allow all students to reach their full potential, leading to a high proportion of top performers and very few low performers.

Finally, the Finnish narrative claims that comprehensive schooling has led not only to excellent student performance, but more importantly to equity: uniformly high student performance. If these factors are linked, then we should expect much lower equity among East Asian countries

and, in particular, Singapore, which follows a radically different model of early tracking. Once again, the data from East Asian countries clash with the Finnish narrative. As we have already discussed, equity has two dimensions: inclusion, which refers to the proportion of students that fail to reach basic levels of attainment, and fairness, which refers to the impact of socio-economic background (or other personal factors) on student performance. In terms of inclusion, according to data from PISA 2015, the proportion of students who fail to achieve Level 2 (defined as the minimum level of achievement required) is equally low in Singapore (10%) and Finland (11%), and is almost half what it is in OECD countries (OECD average: 21%) in most East Asian top performers (OECD, 2016b and 2016c).

Thus, neither the early tracking in Singapore, nor the lower levels of school and teacher autonomy in other East Asian countries, result in a larger proportion of students failing to achieve basic proficiency levels. Quite the opposite, in fact. In terms of fairness, the impact of family background on student performance is slightly higher in Singapore and China than in Finland, but the remaining East Asian countries show similar or lower values than the OECD average (OECD, 2016b and 2016c). Thus, rapid improvements have been achieved without sacrificing equity, but comprehensive policies and teacher autonomy are not significant ingredients in the recipe of success in East Asia.

In the face of clear contradictions between the policies implemented by top-performing countries in East Asia and the narrative built around the comparatively short-lived success of Finland, the OECD has continued to support the narrative on the virtues of comprehensive education, school autonomy and trust in teachers. This seems surprising, given that, according to PISA, Singapore outperforms Finland in some domains by the equivalent of more than one year of schooling, which is no small difference (OECD, 2016b). This stubbornness has led to a lack of transparency on the data from China and other countries, for the sake of preserving a particular narrative.

### China and Vietnam: How to Read the PISA Data

Before we move on to the next section, we feel that a note of caution concerning PISA data from China and Vietnam is needed. China

participated for the first time in 2009, and was only represented by Shanghai, one of the wealthiest cities in the country. The fact that Shanghai-China emerged as the top performer shocked many Western countries, and raised concerns about China outperforming other countries in terms of human capital development similar to the fears that Sputnik had sparked years before about Russia's unexpected success in space technology. The OECD's Secretary General placed a huge emphasis on this result:

the stunning success of Shanghai-China, which tops every league table in this assessment by a clear margin, shows what can be achieved with moderate economic resources and in a diverse social context. In mathematics, more than a quarter of Shanghai-China's 15-year-olds can conceptualise, generalise, and creatively use information based on their own investigations and modelling of complex problem situations. They can apply insight and understanding and develop new approaches and strategies when addressing novel situations. In the OECD area, just 3% of students reach that level of performance (OECD, 2010a)

According to PISA, Shanghai was also a top performer in terms of equity, since their analyses seemed to show that the impact of socio-economic background was small. In fact, nothing seemed to have a negative impact on the performance of Chinese students; the sample had a very large proportion of top-performing students and almost no low-performing students when compared to other countries. The success of Shanghai, which was attributed to China as a whole, continued in PISA 2012, but critics argued that a sample from one of the wealthiest cities was not representative and could not be compared to samples from other countries which included rich and poor regions, as well as rural and urban areas (Loveless, 2013a, 2013b, and 2014).

It was also argued that the PISA sample did not include students whose families had emigrated from rural areas to Shanghai because they were not allowed in secondary schools following the *hukou* system which restricts rural migrants' access to urban social services including education and health (Roberts, 2013; Tao *et al.*, 2013). According to *The Economist* ("China's left-behind", 17 October 2015; "The plight of China's 'left-behind' children", 10 April 2021), around 270 million Chinese workers have moved from their villages to rural areas looking for work; most do not take their children with them. The Chinese call

these children the “left-behind children” and it is estimated that in 2010 there were 61 million children under the age of seventeen left behind in rural areas, with most of them being cared for by grandparents or more distant relatives. In addition, 36 million children had moved with their families to cities, but the *hukou* system denies them access to state schools or health care, so they are also neglected.

As a response to these criticisms, in 2015 the sample was expanded to include Beijing, Shanghai, Jiangsu and Guangdong (B-S-J-G), but the results dropped and China was outperformed by other countries (ranking tenth in science) (OECD, 2016b). For reasons that the OECD has not explained, the sample for China was modified again in the next cycle, since Guangdong was substituted with Zhejiang, a change which propelled China (this time with a different combination of samples from B-S-J-Z) to the top position in the ranking (OECD, 2019c). Once again, the OECD made an effort to highlight the success of “China”:

Among its many findings, our PISA 2018 assessment shows that 15-year-old students in the four provinces/municipalities of China that participated in the study—Beijing, Shanghai, Jiangsu and Zhejiang—outperformed by a large margin their peers from all of the other 78 participating education systems, in mathematics and science. Moreover, the 10% most disadvantaged students in these four jurisdictions also showed better reading skills than those of the average student in OECD countries, as well as skills similar to the 10% most advantaged students in some of these countries (OECD, 2019c)

Of course, the main problem with these results is that, in contrast to what happens in other countries, the PISA sample does not seem representative in any way of the whole country. Furthermore, the lack of transparency has led to serious concerns about whether the samples are even representative of these regions, or whether they leave out immigrant students and students from lower socio-economic backgrounds (Loveless, 2019). To understand the scale of the problem it is worth taking into account that in PISA 2018 the sample size in China included just 12,058 students and 362 schools, which represents less than 1% of the potential sample (a total of 1,221,747 fifteen-year-olds in those regions according to PISA); the sample size was much larger in many countries with comparably tiny population sizes, for instance in Spain the sample size was 35,943 students (7.9% of all fifteen-year-olds).

Furthermore, in PISA 2015 the proportion of fifteen-year-old students in B-S-J-G China enrolled in school was 64%; thus, 36% of students were not enrolled either because they had left school at an earlier age or because they never attended school. It seems reasonable to assume that these students were either from rural areas or disadvantaged backgrounds.

A sample of students with such limitations is not representative of the performance of a country, as we know from the large degree of regional variation that exists in all countries that have measured it. In this case, the sample in China has changed over time in what can only be interpreted as an effort to reach the top position through a focus on cities or regions which enjoy higher levels of wealth than the rest of the country. It seems no coincidence that the provinces included in PISA 2018 are those with the highest levels of GDP per capita; for instance Guangdong was replaced by Zhejiang which has a higher GDP per capita. Thus, PISA is choosing to assess only wealthy coastal regions in China, disregarding the fact that rural areas are much poorer, to the extent that children suffer from severe health issues (Loveless, 2019; *The Economist*, 2015 and 2021). But even in this biased and privileged sample, more than a third of fifteen-year-olds were not enrolled in school. Thus, to draw any conclusions about China being a top-performing country and to try to embellish the story by adding that it also enjoys high levels of equity seems totally unfounded. The insistence on stating that the supposedly “most disadvantaged children in China” outperform advantaged students in OECD countries shows a craving for headlines which is incompatible with robust evidence. Similarly, the OECD’s Secretary General, speaking on Vietnam in the 2015 cycle, stated:

the data also show that the world is no longer divided between rich and well-educated nations and poor and badly educated ones: the 10% most disadvantaged students in Vietnam compare favourably to the average student in the OECD area (PISA, 2016a).

Here again the problem is that more than 50% of fifteen-year-olds were not enrolled in school, strongly suggesting that the most disadvantaged students were not included in the PISA sample (OECD, 2016b). To add to the lack of transparency, data from Vietnam were not made public by the OECD in the next cycle (2018), although the authorities did receive the data from the OECD and published them in national media.

The OECD has gone to great lengths in its effort to try to build a positive narrative around the Chinese educational system: in a Country Note summarising the results for China in PISA 2015, it praises the fact that “Admission to upper secondary school is not based on a single entrance exam. More emphasis has been placed on comprehensive evaluations, including students’ *‘ideological and moral qualities’* (OECD, 2016f). While many countries have decided to balance the results of high-stakes exams with the work that students have carried out during the previous year(s), it is in our view inadmissible that the OECD would regard as good practice the Chinese authorities’ emphasis on ideological obedience to an authoritarian regime as a prerequisite for education.

## 5.2. What Can We Learn from Low Performers? The Latin American Story

Latin America has followed the opposite path to that of East Asia. In 1960 the region had higher schooling levels and the average income exceeded that in East Asia (Hanushek and Woessmann, 2015). It seems shocking that, despite this clear advantage, Latin America today lags so far behind East Asia in terms of human capital, economic growth rates and per capita income (Hanushek and Woessmann, 2015).

The answer to this puzzle is that while Latin America has made a huge effort in terms of increasing enrolment rates, years of schooling and even access to tertiary education, the quality of education systems in the region remains very poor. Economic growth (measured as regional annual growth rates) is closely linked to knowledge capital but only when measured as student performance in ILSAs, which is a reliable indicator of the level of knowledge and skills in the population. However, economic growth is unrelated to years of schooling, because it is not a good index of human capital in countries where education systems are of poor quality. Since students in Latin America perform badly in ILSAs, the failure to develop its human capital explains why Latin America went from being a relatively rich region fifty years ago to a relatively poor one today (Hanushek and Woessmann, 2015).

When the average performance in different ILSAs is calculated for seventy-seven countries for the period stretching from 1964 to 2003, the sixteen Latin American countries included in the sample consistently

underperform, not only compared to top-performing East Asian countries, but also to European and Commonwealth OECD countries (Hanushek and Woessmann, 2015). According to this analysis, the top-performing country in Latin America is Uruguay, which performs well below the lowest-performing countries in Europe (Greece and Portugal). From 2000 onwards, an increasing number of Latin American countries has joined PISA. Eight countries from the region participated in PISA 2015, all of them performing well below the OECD average (OECD, 2016b). The gap between the top-performing country in the region (Chile) and the OECD average is equivalent to more than one year of schooling; the difference between the top regional performer (Chile) and the lowest performer in the region (Dominican Republic) is equivalent to three years of schooling, and the abysmal difference between the lowest-performing country (Dominican Republic) and the top performer (Singapore) in this PISA cycle is equivalent to seven years of schooling, which is close to the average number of years of compulsory schooling in many countries in Latin America (OECD 2016b).

In short, students in Latin America learn very little, because the education systems are very inefficient. While the share of low-performing students is very high, there are almost no high-performing students. Despite this poor level of performance, most countries in Latin America have not improved over the last decades. The exceptions are Peru, where student performance has improved substantially, and Chile and Colombia, which have also shown learning gains (OECD, 2019c)

As a consequence of the low quality of education systems in Latin America, despite the huge effort made in terms of expanding access to education, most of the adult population has very low levels of literacy and numeracy (OECD, 2016e). The extent to which higher levels of educational attainment in Latin America have not had the expected returns of improved knowledge and skills is illustrated by the fact that the level of basic skills acquired by tertiary graduates in Peru, Mexico and Ecuador is lower than that of people who have not attained an upper-secondary education in most OECD participating countries in the survey of adult skills (PIAAC) (OECD, 2016e).

There seems to be widespread recognition that the low quality of teachers in Latin America is the major constraint on the region's education progress and that the major obstacle to raising teacher

quality is neither economic nor technical, but rather political, i.e. the opposition of teachers' unions which are very powerful and active stakeholders (Bruns and Luque, 2015; Bruns *et al.*, 2019; Chambers-Ju and Finger, 2017; Grindle, 2004). Although the proportion of teachers with university degrees has increased over the years (in Brazil from 19% in 1995 to 62% in 2010), the students who enter education degrees are academically weaker than the overall pool of tertiary students in those countries according to their grades in university entrance exams (Bruns and Luque, 2015). In addition, teacher education degrees do not seem to equip future teachers with the knowledge required, since there seems to be a large mismatch between teachers' formal credentials and their real cognitive skills (Tatto, 2014; TEDS-M, 2008). Thus, in Latin America the lack of appropriate selection and training mechanisms means that future teachers tend to be low performers, which is precisely the opposite of East Asian countries.

At odds with the policies implemented by top performers, teacher salaries tend to be flat with no incentives linked to performance, and there are few accountability mechanisms, meaning that teacher absenteeism is common in many countries. Despite many attempts to implement them, no teacher evaluations are in place in most countries, and there is no clear career structure (Bruns and Luque, 2015). Thus, most teachers have a job for life, which trades potentially higher earnings for stability. These features of teachers' working conditions are mostly the result of huge pressures from the unions in Latin America, which are very powerful by global standards (Bruns and Luque, 2015; Bruns *et al.*, 2019; Chambers-Ju and Finger, 2017; Grindle, 2004). This is because unions in the region have large memberships which facilitate coordinated actions, such as long strikes and school closures, as well as large public demonstrations in the streets (Bruns and Luque, 2015; Bruns *et al.*, 2019; Chambers-Ju and Finger, 2017; Corrales, 2005; Grindle, 2004; Kaufman and Nelson, 2004; Palamidessi and Legarralde, 2006). While this is not unique to Latin America, since teachers' unions tend to be the largest in the public sector in most countries due to the sheer size of the teaching force, Latin American unions have gained more power by developing strong alliances with political parties (particularly on the left) and have penetrated governments by occupying key positions in the structure of Ministries of Education.



The most extreme example is Mexico, where the unions control key positions both at the federal and state levels and many education ministers were former state-level union heads (Chambers-Ju and Finger, 2017). Unions in Latin America have opposed changes necessary to improve the quality of teachers to a larger extent than other regions/countries, such as measures seeking higher standards of entry into teaching programmes, better education training programmes, incentives linked to performance, evaluation of teachers and standardised student assessments.

The fierce rejection to these policies is the result of major conflicts of interest between what unions regard as threats to their power and to the favourable working conditions of their members, and the policies necessary to improve teacher quality and enhance student performance. In other words, in countries where teacher quality is low and unions defend the interests of the existing workforce, any measure that may threaten job stability, reduce the size of the teaching workforce, make jobs more demanding or risk the loss of benefits as a consequence of unfavorable evaluations on performance, are vigorously opposed. The deepest conflicts tend to arise in relation to policies concerning the evaluation of teachers. Although reforms addressing this aspect have established attractive incentives for good performers and offered support for teachers who need to improve, the risk that teachers who consistently underperform may be sanctioned or even dismissed seems to override any benefits for high-performing teachers or learning gains for students. Attempts to implement evaluation systems have elicited strong responses from unions in many countries, including violent strikes in Ecuador and Peru (Bruns and Luque, 2015).

In contrast, unions exert strong political pressure in favour of reforms which aim to reduce class size. The benefits for unions are huge, since this means hiring more teachers and therefore expanding the number of members, which is an attractive option for teachers because it facilitates their work and is popular with parents. However, when unions put pressure on governments to reduce class size, they strategically hide these benefits for their members and instead argue that smaller class sizes will facilitate individualised attention to students and therefore increase the quality of teaching.

This narrative is false because, as we have seen, all the available evidence shows that reducing class size *per se* does not have a positive impact on student performance (Bruns and Luque, 2015; Hanushek, 2002; Hanushek and Woessmann, 2015). However, parents are easily convinced that a small class size leads to higher quality of schooling. Thus, governments face huge pressures that have led to a decrease in class size in many Latin American countries, to the extent that nine countries in the region (Argentina, Brazil, Chile, Costa Rica, Cuba, Ecuador, Panama, Paraguay, and Uruguay) have lower class sizes than many better-performing OECD countries, despite this measure being very costly in economic terms (Bruns and Luque, 2015).

But the costs of investing resources in reducing class size are profound and long-term because there are trade-offs involved, which do not seem apparent to other stakeholders. As we have seen in the case of top-performing countries in East Asia, these countries have very large class sizes because they choose to invest the resources in teacher selection, training and professional development. Thus, investment goes to a smaller, more selective, better-trained and higher-paid teaching force, which achieves much better student outcomes. Governments must choose between investing resources in high-quality teachers or decreasing class size. To put it in simple terms: high quality teachers can achieve rapid learning gains in a class with many students but decreasing class size, while keeping the quality of the teaching force low, will not improve student performance. The appeal of small classes to parents and the huge incentives that unions have to reduce class size imply that governments are caught in a double bind between the high political costs of refusing a popular measure and the long-term costs to the quality of the education system. So far, where the political costs seem insurmountable, governments have had little choice but to reduce class size, a trend which is not only present in Latin America but in most OECD countries.

Education systems in Latin America are characterised by their large degrees of inequality. In terms of *inclusion*, the results from PISA reveal a terrible tragedy: between 40% and 60% of the fifteen-year-olds at school in Uruguay, Costa Rica, Mexico, Colombia, Brazil and Peru are low performers, in Chile the proportion is slightly lower (35%) but in the Dominican Republic it amounts to over 80% of students. Compared

to the OECD average of 21% of students, this represents a huge failure of education systems (OECD, 2016b, 2016c and 2018a). But the picture is even bleaker if we take into account the fact that drop-out rates are high, so in most countries in the region between 25% and 40% of fifteen-year-olds are no longer in school; it seems reasonable to assume that those students who have dropped out of school have even lower levels of performance, leading to the conclusion that the overall level of performance among the whole population of fifteen-year-olds is even worse than PISA data suggest.

In terms of *fairness* the impact of family socio-economic background on student performance is higher than the average for the OECD in most Latin American countries, but not all. It should be taken into account that the impact of socio-economic background tends to be smaller when most students have low levels of performance, as is the case in Mexico, but it is more challenging to minimise the impact of family background in countries with higher performance levels, such as Chile.

To measure equity, PISA often compares the variation in student performance between schools with the variation found within schools. The assumption is that larger levels of between school variation reveal larger inequities, because differences in student performance are assumed to be strongly associated with students attending different types of schools and/or schools located in neighbourhoods which differ in family wealth, thus influencing the socio-economic composition of students and resource allocation between schools. Contrary to expectations, this index of inequity developed by PISA shows that in most Latin American countries, between-school variation is considerably smaller than the OECD average (OECD, 2018a). Since PISA has repeatedly shown that messy data should not be allowed to get in the way of a good story, the OECD surprisingly concludes that “the variation is largely due to differences in performance between schools”, unlike “comprehensive education systems—those which do not sort students by programme or school based on their ability—which often tend to have small between school variations in performance”. It also concludes that the supposedly “large” between-school variation in Latin America is due to the “segregated nature of their societies” (OECD, 2018a). This in sharp contrast to the way this same index is used again and again to conclude that, because between-school variation is low in Spain (as in most Latin

American countries), the education system is equitable. The problem, of course, lies with the index.

The reason why this presumed index of equity gives results which are very different to other measures of equity is that in countries where grade repetition is high (as in Latin America and Spain), there is much greater variation within schools, because a large proportion of fifteen-year-old students surveyed by PISA are one or two grades below their modal grade. In all Latin American countries grade repetition is high, with more than 25% of students in most countries repeating a grade at least once, a figure which increases to over 40% of students in Colombia. Thus, in the context of high rates of grade repetition, this index is not a reliable indicator of levels of equity, whether high or low. In the case of Latin America, detailed analyses in Chile using the results of students from the same grade in national exams show that there is a strong correlation with the area in which schools are located (Fontaine and Urzúa, 2018).

As we have seen, PISA recommends “comprehensive” policies in general, and for Latin America in particular, but it also concludes that grade repetition is “often unfair and always costly” (OECD, 2018a). The problem is that in societies with high levels of economic and social inequity, as shown by the Gini index (World Bank, 2021) and where parents have huge differences in their levels of skills (OECD, 2019e), adopting the recommendation to follow “comprehensive policies” (i.e., no streaming or grouping according to student ability) seems to go hand-in-hand with grade repetition. It seems likely that among societies with high levels of inequity, children begin school with very different starting points in terms of basic skills and receive very different levels of support from their families throughout school. In this context, ‘comprehensive’ policies which are based on the principle of treating all students equally, probably end up leaving some students so far behind that at some point the only alternative is for them to repeat a grade. As we have seen, experimental studies have shown that when there are major differences between students, streaming may give low performers a much better chance of achieving their potential (Duflo *et al.*, 2011; Kremer *et al.*, 2013). The case of Singapore clearly shows that early tracking is an efficient policy with which to fight early school leaving when rates are high. Even after Singapore became a top performer in

PISA, early tracking in lower-secondary education led to high levels of student performance without compromising equity.

Historically in Latin America, families and students have actively defended education as an engine of social mobility, leading to powerful social movements against private education and in support of free access to university. The student strikes in Chile in 2011 illustrate this point clearly. As a consequence of this huge political pressure, in Latin America government-funded and privately run schools are almost non-existent, except in Chile, where they represent almost half of schools and have played a major role in improving student performance (Fontaine and Urzúa, 2018). These social movements consider the so-called 'comprehensive' policies to be an essential element of equity, very much in line with the recommendations from the OECD. As a consequence, there are no different trajectories until the age of fifteen or sixteen and very few students choose vocational education at upper-secondary level (OECD, 2016b, and 2016c), since families regard university as the only way to ensure a good job for their children.

Latin America is unfortunately the prime example of what goes wrong when education policies that work well in societies that were already egalitarian (such as Nordic countries) are implemented in non-egalitarian societies in the belief that they will achieve similar results in terms of quality and, more importantly, equity. 'Comprehensive' policies in Latin American countries imply a monolithic path that is the only option available to a very diverse student population, creating the illusion that such an education system offers all students the same opportunities to succeed.

Latin America represents the laboratory for investigation into the outcome of such policies in societies with high levels of inequity: high drop-out rates, high rates of grade repetition, an almost non-existent VET system and low standards in order to maximise the number of students progressing through the education system despite their low levels of performance. It is likely that the low quality of teachers makes it even more difficult to achieve good levels of student performance when a diverse student population only has one path and, contrary to expectations, makes equitable outcomes impossible. Such poor outcomes lead to intense levels of frustration even among students who achieve high levels of education attainment formally, since they have low

levels of knowledge and skills and therefore do not obtain the expected returns. When education systems offer big promises through high levels of enrolment, but fail to deliver on quality, they breed resentment, which is reflected in the massive student protests that periodically sweep through Latin America. Such a system not only fails those who drop out. It also fails those who remain.

### 5.3. What Have Countries Learned from the PISA Shock?

From time to time, some countries have worse results than expected in PISA and this generates an intense policy debate about education that puts pressure on governments to implement reforms. This strong reaction to disappointing results in PISA is known as the 'PISA shock'. The better-known cases are Germany, Denmark and Japan, which are often cited as examples of PISA's strong impact on education policy (Baird *et al.*, 2016; Breakspear, 2014; Egelund, 2008; Ertl, 2006; Martens and Niemann, 2010; Martens *et al.*, 2016; Ringarp, 2016; Takayama, 2008; Waldow, 2009).

However, it is striking that after more than twenty years of PISA cycles there are so few countries where PISA results have led to the implementation of education reforms. It is also important to consider whether those education reforms were aligned with the policy recommendations from PISA and, most importantly, whether they improved student performance.

Germany apparently expected to be recognised as a top performer in the first PISA cycle, but ultimately performed below the OECD average and worse than many European countries (OECD, 2001). PISA data also showed that student socio-economic background and immigrant status had a large impact on student outcomes, raising concerns of low levels of equity in the German education system. The fact that these poor results were a shock is remarkable and highlights an important feature of the impact strategy developed by PISA. From the beginning, PISA has developed a very proactive communications strategy that targets media in all participating countries and provides them in advance (under embargo) with data, as well as the most important policy messages it is developing for each country. This collaboration with the media also

includes interviews with OECD staff by major media, targeting those countries where the impact is greatest. The media effort is such that PISA has developed launches at different levels: global, regional, and national launches. In relation to the latter, it works closely with governments to prepare joint launches at press conferences once the main conclusions have been discussed. The public presentation of results hand-in-hand with governments all over the world greatly facilitates the internal policy discussions that PISA intends to promote within countries, since it places current governments at the centre of the debate and elevates the OECD to the role of honest broker.

As a result, its impact is achieved mainly through the media, who give greater salience to the findings by highlighting poor results. This explains at least partly what happened in Germany when the results of PISA 2000 were launched, since the low levels of performance could not possibly have come as a surprise to policymakers, who already knew that the results of Germany in TIMSS 1995 (Harmon *et al.*, 1997) were very poor and that the data available from other international surveys showed that student performance in Germany was already below most other European countries in 1975, and continued to decline further until 2000 (Hanushek and Woessmann, 2015; Martens and Niemann, 2010). But the results of previous international surveys did not have a similar impact on the media and were therefore not widely known. By generating a media scandal, PISA does attract the attention of the public and other stakeholders, which may put pressure on governments to take action.

In an OECD report which claims that PISA was the trigger in Germany for a package of reforms that led to rapid improvement in student performance, the fact that this approach aimed to create a significant media impact is openly acknowledged as the following quotes show:

Whilst the TIMSS results had hardly been reported, major newspapers ran four, five and six-page special sections on the PISA results. The news and discussions of the results were all over the radio and television. The news about Germany's poor results got far more coverage in Germany than the surprise news that Finland had topped the PISA league tables got in Finland [...] Each (political) side had been effectively blocked by the other for years, producing gridlock on educational policy change. But the "PISA shock" changed all that. Now, for the first time in years real change was possible on a surprisingly large scale. The uproar in the

press reflected a very strong reaction to the PISA results from the public. Politicians who ignored it risked their careers. (OECD, 2011)

These quotes also reflect a unique attribute among international surveys that PISA proudly emphasises: by amplifying poor results in the press, it intensifies the pressure from the public and stakeholders such that governments end up caught between a rock and a hard place. As a consequence, they need to be seen to act swiftly. Although PISA also claims that it facilitates governments' jobs by providing the right solutions, this is an overstatement to say the least. In any case, it seems contradictory that PISA boasts about the costs it inflicts on governments to the extent of claiming that policymakers may risk their careers, since it is those same governments who have decided, funded and implemented the participation in PISA. Furthermore, forcing governments to respond quickly to public outcry does not seem the best strategy to ensure an in-depth analysis and the search for appropriate policy solutions that each country needs. On the one hand, this 'awe and shock' strategy is likely to prompt the wrong quick-fix response in governments. On the other hand, it does not seem appropriate for a member-led organisation like the OECD to play this name-and-shame game against its own members.

The same OECD report cites Edelgard Bulmahn, who was the German Minister for Education at the time, claiming that:

the tripartite system of secondary schools was a mirror image of the feudal system, a system that only needed a small number with high qualifications, a few with the middle range of education and the rest with only a basic education. (OECD, 2011)

Andreas Schleicher (widely recognised as the father of PISA) has always supported this view, but in this report he goes further and states that after the fall of the Berlin Wall,

the West German system was implemented in the East. Lost to East Germany was their more equitable, de-tracked education system along with their excellent early childhood system (OECD, 2011)

Here, Schleicher clearly highlights the contrast between what he calls "feudal" and real "communist" education systems and shows a clear ideological preference for the latter in a misguided interpretation of



“equity”, despite the clamorous failure of the Soviet Union (OECD, 2011).

While in the case of Finland, PISA developed its policy recommendations gradually after the results of the first cycle, in the case of Germany the conclusion that the most pernicious element of its education system was early tracking and the co-existence of at least three tracks from lower-secondary education was adopted immediately. The fact that Andreas Schleicher is German may have contributed in a positive and a negative way, through a better understanding of the education system and perhaps the prejudices commonly associated with personal, direct experience.

Whatever the reasons, PISA attributed both the low levels of student performance and the low levels of equity to early tracking and a well-developed vocational education and training system which attracts a significant proportion of students (OECD, 2011). The contrast between Finland’s unexpected success and Germany’s unexpected failure was used to create a strong narrative that remains powerful today: comprehensive policies in Finland were associated with high levels of equity, which led in turn to high levels of quality, while early tracking in Germany was linked to low levels of student performance and large inequalities. But this strong recommendation was never followed by Germany, nor by other European countries with early tracking such as Switzerland and the Netherlands.

Germany is a federal republic with sixteen states (*Länder*) which have full responsibility over education and coordinate their activities through the Standing Conference of the Ministers of Education (*Kultusministerkonferenz, KMK*). The ‘PISA shock’ generated an intense policy debate in Germany that led to important reforms. These included the establishment of national educational standards in 2004 that defined curricular elements for core subjects, as well as common objectives for all states; each state is then required to develop a full compulsory curriculum which is aligned with these objectives (Ertl, 2006; OECD, 2011; Tarelli *et al.*, 2012). The reforms also led to the creation of the Institute for Educational Quality Improvement in 2003, which evaluates whether advances are being made towards these standards.

In June 2006, the Standing Conference of the Ministers of Education adopted a strategy for educational monitoring which included

national student assessments, designed independently by the Institute for Educational Quality Improvement, as well as participation in international comparative surveys. Teacher unions ensured that student performance in the new exams would not be linked to teachers' pay or influence their promotion or retention (OECD, 2011).

Since Germany has not followed the consistent and generalised recommendation from PISA to delay vocational education and training until upper-secondary level, early tracking remains today, as well as several differentiated academic and vocational tracks (Tarelli *et al.*, 2012).

Thus, in lower-secondary schoolchildren are assigned to different tracks according to their ability level. States differ in the age at which children are assigned to different tracks (which varies from ten to twelve years) and in the options that they offer, although the three traditional tracks are the most common: *Hauptschulbildungsgang* which allows students to proceed to vocational training, *Realschulbildungsgang* which leads to vocationally oriented upper-secondary school and *Gymnasialer Bildungsgang* which allows students to obtain the *Abitur*, which qualifies them for university. These secondary school tracks may be offered by different types of schools known as *Hauptschule*, *Realschule* and *Gymnasium*.

In the 2012–2013 academic year, about 14% of German students in eighth grade attended a *Hauptschule*, about 23% attended a *Realschule*, and about 36% attended a *Gymnasium* (Tarelli *et al.*, 2012). An increasing number of schools offer several of these tracks and some states have re-designed the offer so that schools combine at least two tracks (*Gesamtschule*).

Upper-secondary education offers different pathways that cover a broad range from full-time general education to dual vocational education and training, which integrates on-the-job-learning (Tarelli *et al.*, 2012). Since more students are attracted to vocational education and training than in most other European countries, a lower proportion of students attend university, a feature that has been repeatedly criticised by the OECD (OECD, 2020d).

Despite the decision by Germany to maintain early tracking, according to PISA equity improved from 2000 to 2015 (OECD, 2018c). The impact of socio-economic background on student performance

is slightly higher than the OECD average, but much lower than the same impact in some countries with no early tracking, such as France. Furthermore, 32% of disadvantaged students in Germany perform at high levels of proficiency, which is higher than the OECD average (25%) and higher than most European countries. Thus, although family background has a stronger impact, it is also true that a much larger proportion of disadvantaged students reach high levels of performance.

These findings suggest that early tracking *per se* does not necessarily have a negative impact on quality or equity, but it depends entirely on whether students in each track acquire a good level of knowledge and skills, whether the assignment of students to tracks is based on a fair assessment of student ability rather than personal factors that may lead to discrimination (such as family socio-economic background or immigrant status), whether there is flexibility to move between tracks when levels of performance change and whether mobility between each track to higher levels of education is allowed. Other examples in Europe include Switzerland and the Netherlands and we have already discussed Singapore, a top performer with early tracking. Furthermore, the development of vocational education and training, its attractiveness to students and its close links to firms (through on-the-job training), has led to high rates of youth employment in Germany during economic crises compared to other countries in Europe, since many VET students get jobs that require middle and high levels of skills (OECD, 2020d and 2020e).

In the European context, Germany has received many refugees and migrants, particularly as a consequence of the Syrian crisis in 2014 and 2015, and it has granted asylum to the largest number of applicants of any European country. Consequently, a large number of children and teenagers have arrived in Germany who do not speak the language. The proportion of immigrant students increased from 18% in 2009 to 22% in 2018 (OECD, 2019d and 2020c). In order to facilitate these students' integration, several measures have been implemented to support immigrants, including basic programmes to learn German (Bergseng *et al.*, 2019; Fazekas and Field, 2013; OECD, 2020f). However, this has proven difficult, as illustrated by the fact that the performance of foreign-born students deteriorated from 2009 to 2018, and that they face higher rates of early school leaving (OECD 2020f). In this particular case, VET

has proven to be a pathway that allows immigrant students to acquire strong foundation and practical skills more rapidly, thus facilitating their entry to the labour market (Bergsen *et al.*, 2019; Fazekas and Field, 2013; OECD, 2020f).

Undoubtedly the most important question is: was the 'PISA shock' effective in promoting reforms that led to enhanced student performance? Although PISA widely claims that this is the case, in our view data from PISA and other international surveys show that it is debatable (see Martens and Niemann, 2010; Martens *et al.*, 2016). According to PISA, while the performance of fifteen-year-olds in reading improved slowly and gradually until 2012, it has remained at levels around the OECD average since then. German students had better starting points in mathematics (OECD, 2004) and science (OECD, 2007), which remained more or less stable until 2012 and declined thereafter.

Other international surveys show that, among primary students, performance in reading was lower than that of other European countries and remained stable from 2001 until 2016 (Mullis *et al.*, 2017), while performance in science and mathematics remained stable from 2007 onwards and was also lower than the performance of students in many European countries (Mullis *et al.*, 2020). Thus, the data from international surveys do not show significant improvements over time, although the large increases in the share of immigrant students may have contributed to the declines observed in recent years.

If we analyse other countries which experienced the so-called 'PISA shock', we realise that the policy reactions are quite different, but none of them have had a significant impact on student performance (Niemann and Martens, 2018). Denmark was also shocked by its first PISA results, which were not only worse than expected but also below the performance level of its Scandinavian neighbors, particularly Finland (OECD, 2001). This was surprising in a country with high levels of investment in education.

However, the government did not implement reforms until an in-depth international review was able to analyse in detail the strengths and weaknesses of the education system, and policy recommendations were developed on the basis of several studies (Breakspear, 2014; Egelund, 2008). The recommendations included developing student assessments, improving teacher pre-service training, which was

perceived as too generalist in its approach, and taking advantage of the high levels of investment to develop in-service training for teachers. Most of the reforms in Denmark aimed at implementing national assessments and providing support for disadvantaged students (Egelund, 2008). Despite this more elaborate approach, the performance of fifteen-year-old Danish students has not improved significantly over time according to PISA (2000–2018) (OECD, 2019f). Their performance in reading and science has followed a flat trajectory, with scores similar to the OECD average, while their performance in mathematics declined until 2012 and then returned back to the levels observed in the first cycles.

Data from TIMSS and PIRLS support the conclusion that Denmark has not improved over time, in this case by looking at the performance of primary students: no significant changes have been observed between 2007 and 2019 in mathematics and science, and no changes have taken place in reading between 2006 and 2016 (Mullis *et al.*, 2017 and 2020). In all domains, Danish students continue to be outperformed by Norway, Finland and Sweden, providing additional support to the idea that investment *per se* is not a guarantee of better student outcomes.

Japan is also often cited as an example of PISA shock, but in this case it seems to be the reaction of a top-performer to a slight decline in the performance in reading in PISA 2003. Japan had been a top performer since at least 1975 according to previous international surveys, and it continued to improve until the year 2000 (Hanushek and Woessmann, 2015). When the first PISA cycle took place in 2000, all of the media attention focused on Finland, which became the top performer in reading (the main domain in the initial cycle), while Japan, the top performer in mathematics and “joint top-performer in science” with Korea (OECD, 2001), generated no interest.

It has been argued that biased media reporting on PISA results in 2003 was used as an opportunity by the government to halt an unpopular curriculum reform known as *yutori* (Takayama, 2008). Once again, according to PISA no major changes have been observed in the performance of Japanese students from 2000 until 2015, and over this period the country has remained a top performer particularly in science and mathematics (OECD, 2019g). The results from other international surveys are more positive, and show that, despite being a top performer, Japanese students continued to improve in mathematics and science

both in fourth grade and eighth grade (although no significant changes were observed between 1995 and 2011) (Mullis *et al.*, 2020). Thus, improvements occurred well after the PISA shock and are unlikely to be related to reforms implemented as a reaction to it.

The evidence shows that the first reaction to PISA results consists mainly of media outrage, which supposedly serves as a wake-up call for governments to act. The fact that the media act as a mediator in creating a narrative around the PISA results is a direct consequence of the effort that PISA makes to engage with media all over the world. Although PISA boasts that this forces governments to react, this strategy involves huge risks. The most important is that even when governments do respond to the pressure by hastily implementing reforms, the evidence shows that it has had no impact on student performance. This may be either because (a) governments do follow policy recommendations which happen to be wrong, (b) governments use additional data and analyses to deal with weaknesses that PISA data do not identify and therefore implement reforms which are not aligned with PISA recommendations and have an impact on variables not captured by PISA, or (c) stakeholders with vested interests take advantage of the media uproar and escalate the pressure on governments to implement policies that are beneficial to them but not necessarily to students.

#### 5.4. What Can We Learn from Countries that Improve in Europe?

In the European context, there is a different group of countries which have improved rapidly over time. As is the case with top-performing countries in Asia, such improvements seem mostly unrelated to PISA policy recommendations. We will analyse which policies have proven successful, the extent to which they are aligned with PISA policy recommendations and the extent to which PISA has learned from them.

##### Poland

Poland joined PISA in the first cycle (2000) and performed below the OECD average in reading, mathematics and science. Student performance improved rapidly to the extent that Poland became a

top-performing country within Europe in 2012, but then declined in 2015 (OECD, 2019h). Data from TIMSS broadly support these findings since the performance of primary students in Poland improved from 2011 until 2015 and declined in 2019 (Mullis *et al.*, 2020). In addition, data from PIRLS show that primary students had high levels of performance in 2016 (Mullis *et al.*, 2017).

After the collapse of the communist system in 1989, Poland experienced a rapid transition to a Western-style parliamentary democracy and a fast-growing market economy. The first changes in education were focused on eliminating the indoctrination that the communist regime had introduced in textbooks and curricula. It took almost a decade to prepare for major structural reforms of the education system. The series of reforms started in 1999 with a structural reform, followed by a curriculum and evaluation reform in 2007 and an early education reform which was gradually implemented from 2009 until 2015 (Jakubowski, 2021). Despite their positive impact on student performance, these reforms were largely reversed in 2016 following the arrival to power of right-wing populists, the PiS (Law and Justice) party.

The structural reform of 1999 was very ambitious since it introduced several major changes into the education system at the same time: it extended compulsory education by one year, by replacing eight years of basic primary education with nine years of comprehensive education, which was divided into two stages: six years of primary education and three years of lower-secondary education. The creation of a new school level (lower-secondary) was the most dramatic structural change, which was aligned with the implementation of a new curriculum and opened up the possibility of hiring teachers specialised in particular subjects. Thus, all students had access to one additional year of compulsory education which delayed the beginning of upper-secondary education by one year; the byproduct of these major changes was that the differentiation between academic and vocational tracks in upper-secondary education was delayed by one year until the age of sixteen. In upper-secondary education, students could choose between the *lyceum* (academic, three years), technical school (four years) or vocational school (two to three years).

An important aspect of the reform was to improve teachers' professional development. The Ministry organised a new programme,

“New School” (*Nowa Szkoła*) to train a large number of coaches who then provided professional development programmes for teachers, which included changes in the curriculum, capacity building to assume new responsibilities as a consequence of enhanced school autonomy, the establishment of school assessments, and preparation for external examinations (Wisniewski and Zahorska, 2020). Approximately 70% of all Polish teachers participated in these programmes.

The reform also introduced standardised external national exams, which were implemented in 2002 at the end of every educational stage. The results were evaluated centrally to ensure the same standards. They were available to students and teachers and had academic consequences. The results at the school level were made public based on measures of student progress. The impact on student performance was particularly positive for students lagging behind. Additional measures included changes in governance towards the decentralisation of the education system, which increased school autonomy, and changes in funding, which made the system more efficient (Jakubowski, 2021).

Later in 2008, the curricular reform designed a consistent curriculum from preschool all the way to upper-secondary level, defined learning outcomes as targets for each educational stage, and strengthened the core subjects in vocational schools. Students who followed vocational degrees improved their employability substantially. Finally, the reform of early education which started in 2005 expanded compulsory education by another year by advancing the starting age to five years. Thus, all in all these reforms extended the duration of compulsory education by two years.

This complex package of reforms was aligned with the rapid changes that were taking place in the economy and was successful in achieving the goals that had been defined. A growing number of students enrolled in upper-secondary (either academic or vocational) education, which provides access to tertiary education. As a consequence, Poland experienced one of the largest increases in the proportion of students with a tertiary degree in Europe.

But the most convincing evaluation of the success of such reforms was the drastic improvement in the performance of fifteen-year-olds revealed by PISA. Despite the sheer complexity and magnitude of the reform package, the conclusions that PISA extracted from the



improvement in Poland focused almost exclusively on the fact that vocational education and training was delayed by one year (OECD, 2011). Other authors have placed greater emphasis on the extension of compulsory education by one year to all students, which involved the creation of new lower-secondary schools with new curricula and more autonomy, as well as greater emphasis on core subjects in vocational education (Wisniewski and Zahorska, 2020).

The effort made by the government was huge since the education reform was implemented in parallel with reforms on pensions, health and public administration. The political cost was enormous, partly because the implementation of so many structural reforms simultaneously was a mammoth task, and partly because the reduction of years in primary school, the closure of small rural schools, and higher requirements (university degrees) for teachers of the new lower-secondary level all led to some teachers losing their jobs. In addition, the new professional development framework linked participation to promotion and salary increases.

Some of the reasons behind the rejection and reversal of the reforms are a good example of the extent to which cultural factors can play a key role, and of the low impact of international evidence when it clashes with values held by societies, or segments of them. The fact that school became compulsory for six-year-olds and pre-school for five-year-olds was used to criticise government for supposedly taking young children away from their families, a concern which was stronger among rural families (Wisniewski and Zahorska, 2020). As a consequence, a grassroots social protest movement was organised under the slogan “Save the Toddlers”. Although there is plenty of evidence that early childhood education and care is the best tool enabling children from disadvantaged backgrounds to catch up with others before they start school (OECD, 2015d), once these fears were triggered the policy was rejected and the compulsory school starting age was raised to seven years.

The political cost of implementing such an ambitious package of reforms meant that the government collapsed shortly after, and some elements of the reform were never implemented by succeeding governments (for example, an obligatory standardised mathematics exam at the end of upper-secondary school was postponed for nearly ten years) (Jakubowski, 2021). The high political cost and the fact

that such reforms were implemented despite strong rejection from the unions may have contributed to the popular support for reversing the reforms in 2016 (Jakubowski, 2021). It is uncertain whether such an ambitious package of reforms could have achieved consensus without major concessions and delays; some have argued that this is highly unlikely (Jakubowski, 2021).

## Portugal

In the first PISA cycle (2000) Portuguese students performed well below the OECD average and near the bottom of the ranking (OECD, 2001). This generated no shock, since Portugal had already performed poorly in TIMSS 1995 (Harmon *et al.*, 1997) and it was not afflicted by the superiority complex that other countries shaken by PISA results seemingly were. Quite the opposite: universal access to education, length of compulsory education and completion rates had traditionally lagged behind other European countries, so Portugal did not have high expectations (Crato, 2021). As an example, let us just mention that compulsory schooling was extended from three to four years in 1956 for boys and in 1960 for girls, to six years in 1967, to ninth grade in 1986 and to twelfth grade as recently as 2012 (Crato, 2021).

Despite this unpromising start, Portugal's performance in PISA improved up until 2015, and it is the only OECD country that has shown a positive trend in all three domains (OECD, 2019i; Maroco, 2021). Such progress means that fifteen-year-old students in Portugal have evolved from performing at the equivalent of one school year behind the OECD average, to performing at the same level as the OECD average. This improvement has occurred due to the combination of two positive trends: the proportion of low-performing students has decreased, while the proportion of high-performing students has increased (OECD, 2015; Maroco, 2021). This shows that reforms can improve the quality and equity of education systems simultaneously, by supporting struggling students and developing the potential of top performers (OECD, 2016b; Maroco, 2021).

However, in PISA 2018 Portugal's performance declined, which could be due to the methodological issues that plague the results from this cycle, or to changes in policy. Data from other international

surveys which assessed students in primary education also show rapid improvements from 1995 until 2015, and a drop or stagnation in 2019 in mathematics and science (Mullis *et al.*, 2020), suggesting that policy changes may be responsible for the declines.

The steepest increase in PISA results took place from 2006 to 2009, just after the introduction of standardised ninth grade exams for mathematics and Portuguese in 2005, leading to the conclusion that high-stakes assessment was probably the policy with the single largest impact on student performance (Bergbauer *et al.*, 2018; Maroco, 2021). The fact that several consecutive governments gave continuity to the reforms initiated by previous ones is key to Portugal's success. Further improvements in student performance were achieved despite the deep financial crisis, which led to the international bailout of Portugal from 2010 until 2014. This period was marked by an intense austerity policy, and a significant decrease in the level of investment in education (which reduced from 4.8% invested in education in relation to GDP in 2000 to 3.6% in 2018) (Maroco, 2021).

The education minister, Nuno Crato (2011–2015), implemented a package of reforms which ensured that improvements in student performance continued. These include defining learning targets in the curriculum, reinforcing core subjects and implementing new high-stakes exams (Crato, 2021; Maroco, 2021). The leadership of school principals was strengthened and major reforms were introduced to enhance the quality of the teaching profession. These included higher academic requirements to enter educational programmes, stronger content of STEM subjects in teacher training, improved pre-service training, more demanding selection procedures to enter the profession, evaluation of teachers, and a new framework for teachers' professional development (OECD, 2020g). An important aim was to decrease high rates of grade repetition and associated high rates of early school leaving. The policies included the modernisation and development of two differentiated tracks of vocational education and training in upper-secondary level, including stronger links with industry and more on-the-job-training, as well as the introduction of vocational courses in lower-secondary starting at the age of thirteen (Crato, 2021).

The positive impact of these policies was clear: dropout rates decreased from 43.6% in 2000 to 28.3% in 2010 and 13.7% in 2015. To this

day Nuno Crato claims (pers. comm.) that one of the key elements of the success of such reforms was to increase expectations for disadvantaged students through more demanding assessments and aligned curricular content, instead of lowering standards to make it easier for students with low levels of performance to obtain educational degrees (Crato, 2020). However, student performance dropped after 2016 when the national high-stakes exams for fourth and sixth grades were substituted for diagnostic tests that covered only a sample of students and had no academic consequences; teacher evaluations were also abolished (Crato, 2021; Maroco, 2021).

In conclusion, both in Poland and Portugal, education reforms began before the first PISA cycle and were therefore not the result of PISA policy recommendations. More worryingly, such reforms were undone after PISA data demonstrated that improvements in student performance were substantial. Thus, the evidence from PISA was not used to design the reforms and was not enough to prevent the reversal of successful policies.

## Estonia

This is how Estonian education expert Gunda Tire describes the impact of the launch of PISA 2018 results in Estonia:

PISA 2018 data was released on December 3, 2019 and it turned out to be almost like a national holiday. The press conference led by the minister of education and research was streamed online, all the main media channels were present, and the news spread fast—according to PISA 2018, Estonian education system is the best in Europe and among the best performing systems in the world! The evening news on national TV devoted more than 10 min to covering PISA results, journalists had interviewed students and teachers from different urban and rural schools, and everybody felt that they had personally contributed and were very proud of their achievement. (Tire, 2021)

Estonia joined PISA in 2006 with levels of student performance similar to the OECD average in reading, slightly above the OECD average in mathematics, and considerably better than it in science (OECD, 2019j). Since then, performance in reading has improved steadily, surpassing the OECD average in 2012 and continuing to improve in successive cycles, while improvements in mathematics have been smaller, and

science has shown no improvement from 2006 until 2018. Despite small or no improvements in the latter two domains, student performance has remained well above the OECD average in science and mathematics. It is revealing that PISA elevated Estonia to the status of “PISA star” in 2018, although Estonia did not perform better in all domains in that cycle. In fact, it was in PISA 2012 when the country achieved its steepest improvements and highest levels of performance: a peak in science (which declined thereafter), as well as in mathematics (which remained unchanged in the following cycles), and the largest improvement in reading (which continued to increase slightly).

This points to the importance of the country rankings for PISA conclusions and policy recommendations: Estonia reached its highest levels of performance in science and mathematics in 2012, but in this PISA cycle it performed below other European countries such as Finland, Switzerland or the Netherlands (OECD, 2014c). Although its performance in science declined in 2018, and remained unchanged in mathematics and reading, the fact that it became the top-performing country in Europe in all domains and, in particular, the fact that it outperformed Finland in reading, was used to elevate it for the first time into a success story. This decision was questionable, since Estonia was already performing better than the OECD average in two domains since joining PISA in 2006.

This early success is supported by data from TIMSS: in 2003 Estonia was *the* top performer among European countries in science (performing only below East Asian Countries) and one of the top performers in mathematics (Mullis *et al.*, 2017). Data from the survey of adults (PIAAC) paints a similar picture since Estonia ranks among the top ten participating countries in literacy and only slightly lower in numeracy (OECD, 2019e). The improvement in the level of basic skills acquired by the adult population took place some time ago, since the survey conducted in 2011 shows that the youngest cohort (sixteen to twenty-four years old) has a level of numeracy and literacy which is very similar to that of older cohorts (twenty-five to thirty-four and thirty-five to forty-four years old), and only moderately above the oldest cohorts (forty-five to fifty-four and fifty-five to sixty-five) (OECD, 2019f). This in stark contrast with countries whose education systems have improved in the last decades, such as Singapore, where younger cohorts have

acquired much higher levels of basic skills than older cohorts (OECD, 2019l). Thus, as appealing as it may be to be proclaimed the winner of the European gold medal, the nice story built around Estonia's surprising success in 2018 is misleading, because the roots of the country's success can be found much earlier.

One of the main messages from PISA was that Estonia had achieved the status of top performer despite the fact that it invested 30% less than the OECD average on education, thus making it a prime example of efficient investment (OECD, 2019c and 2019d; Tire, 2021). However, if we take into account that the number of students in Estonia has decreased dramatically over the last years (40% decrease in the last fifteen years) due to emigration and low birth rates (OECD, 2019k), public spending per student as a percentage of GDP per capita in 2015 was similar in Estonia to that in Canada and higher than that in Australia, Germany or the Netherlands (Goss and Cowgill, 2019; OECD, 2020d).

What factors have led to the high level of performance of students in Estonia? According to the Vice-Minister of Education and Research, Mart Laidmets, who was asked this question in an interview after the launch of PISA 2018,

the three pillars of the Estonian education system are the national curriculum, the second is the teachers who are highly qualified—holding master's degrees" and "the third important aspect is parents. It is important that parents recognise that education gives people the chance to move forward and to reach their potential. (*TES Estonia: "PISA's European success story"*, John Roberts, 3 December 2019)

As is the case in Finland, teachers in Estonia have played a significant role in its history and in the creation of a national identity since the nineteenth century. Estonia first became an independent state in 1918 and introduced free, compulsory and public education for all, but in 1940 it was occupied by the Soviet Union. A few years before Estonia regained independence, teachers became particularly vocal in their defence of a curriculum free from communist indoctrination.

Estonia regained independence in 1991 and used this opportunity to implement an ambitious reform of the whole education system. Due to its geographical proximity and cultural and linguistic similarities to Finland, the redesign of the education system borrowed many elements from Finland. Some have argued that the cultural similarities between

the two in fact go much further, since both countries share a history of high levels of literacy. In the second half of the seventeenth century, the Church played a major role in teaching peasant children to read by providing education in the native language, and in 1897 literacy levels among Estonians (80%) were the highest in the Tsardom of Russia (Lees, 2016). In 1920, six grades of education were compulsory, and this was extended to eight grades between 1959 and 1963, and then to nine grades between 1968 and 1988 (Lees, 2016).

After Estonia regained independence, a national education system was created and several laws were adopted during the 1990s to provide the legal framework. The national curriculum implemented in 1996 lays out the learning outcomes that students should achieve at different stages and has been praised for its quality and coherence (Oates, 2010). Thus, the first opportunity that Estonia had to compare itself on the international stage and measure the impact of this new curriculum was six to ten years later, when it joined first TIMSS and then PISA.

During the 1990s, schools were decentralised, transferring more responsibilities to local municipalities and giving more autonomy to principals (who could select teachers and manage the budget) and teachers (who could choose learning materials). In the mid-1990s Estonia started to develop centrally administered assessments and evaluations (Lees, 2016). In 1997 an external evaluation system was implemented to evaluate the extent to which students had achieved the learning goals set in the curriculum (Tire, 2021). At the end of compulsory education (lower-secondary level) and at the end of upper-secondary education, exit exams were introduced for all students. These are centrally designed and are required to enter tertiary education.

As a country, Estonia made a commitment to become a digital economy and society and this has been incredibly successful. Internet voting is used by almost half of the population and 99% of public services are available online (E-estonia, 2019b; OECD, 2020a). Digitalisation also had a big influence in the education sector, since the “Tiger Leap” (a reference to the successful economies in East Asia) programme, initiated in the 1990s, which was developed to ensure Internet connectivity in schools, access to computers for all students, and IT training for teachers (Laanpere, 2002). All schools in Estonia use “e-school solutions”, which is an online school-management and communication system

for pupils, teachers, parents and government bodies (E-estonia, 2019a and 2019b), the “e-schoolbag”, which is an online portal that provides digital learning materials for all education levels (E-estonia, 2019b) and the Examination Information System, which is a tool to carry out and evaluate examinations and assessments (Innove, 2019).

The Estonian Education Information System is a particularly powerful tool. It is a database that collects data on students, schools, examinations and curricula, among other things. It allows the tracking of each student’s development over time, including teachers’ assessments, grades, state exam results and any cases requiring special support. Individuals have access to their own personal data, schools to their own information, aggregated data are available to the public, and parents can advise their children on career decisions since information on employment rates and average income for vocational training courses and university degrees is available. Finally, policymakers make decisions based on these data. This is in our view fundamental, since evidence-based policymaking requires reliable data, which are rarely available. However, the link between this unique asset and high student performance has not been investigated so far.

As part of the general reforms, teachers were required to obtain a master’s degree (Tire, 2021). In 2013, professional standards were defined to assess potential candidates who wished to become teachers, and a programme of teacher professional development was designed (Lees, 2016). However, teachers in Estonia follow a so-called teacher-centred approach, which is considered traditional (Tire, 2021). Despite the excellent student outcomes achieved through this approach, the recent Estonian Lifelong Learning Strategy 2020 has outlined its goal to move to a so-called ‘student-centred approach’.

The education system has high levels of equity, with few differences in student outcomes due to family socio-economic background and only minor gaps between rural and urban areas. However, a major divide seems to persist between students attending Estonian schools and those attending Russian schools, with the latter lagging behind by the equivalent of a year of schooling in reading and science (Tire, 2021). It would be important to analyse the extent to which changes in reading performance in PISA are related to the existence of an education system with two languages taught in different schools.



Finally, the importance that parents place on education is reflected in the high rates of pre-school education, with 91 % of four-year-olds and 87% of three-year-olds enrolled in early-childhood and care institutions, which is well above the OECD average (Lees, 2016). While attending kindergartens before they join school at the age of seven, children follow the national curriculum and learn to read and write.

Despite all these reforms, some have argued that the success of the education system in Estonia is to a larger extent the result of cultural and historical factors. Marc Tucker has stated:

So the fact that Estonia is among the top ten performers on PISA worldwide does not appear to be the result of education policies pursued since Estonia gained its independence, as much as it is the result of hundreds of years of political, social and educational development which ended up supporting a strong, deep and widespread commitment to education as well as a tradition of very high education standards, a very demanding curriculum matched to the standards, high quality examinations built directly on that curriculum, highly educated teachers with masters degrees from research universities, a well designed qualifications system, a strong system of support for families with young children, and most of the other drivers of high performing national education systems that we had found over the years in such systems. (Tucker, 2015)

The quote seems a bit misleading since it refutes the role of education policies in the first sentence and then goes on to identify a number of education policies which are key to Estonia's success. In any case, Estonia has often been used as an example of a country where history, culture and the value placed on education by families, all play a much greater role than specific education policies (*The Economist*: "PISA results can lead policymakers astray—the parable of Finland", 7 December 2019). Whether this is true or not, it raises important issues: to what extent does PISA measure what happens in school or elsewhere? If the latter, to what extent can policymakers improve education systems in cultures where education is not highly valued?

### 5.5. What Can We Learn from Countries where PISA Has a Huge Media Impact but No Policy Reactions?

Spain has a unique relationship with PISA: the media impact in Spain is much greater than in most other countries (Martens and Niemann, 2010), but the policy reactions have been scarce. Contrary to what

happened in countries such as Germany, where the contrast between high expectations and poor results led to media outrage (the so-called PISA shock) and put pressure on policy makers to initiate a number of reforms, in Spain there is no discrepancy between the level of expectations (low) and the results (mediocre). The good adjustment between expectations and actual levels of performance is a trap that has prevented most policymakers from reacting to PISA findings and reveals a clear lack of ambition as a country or, in other words, a lack of trust in the power of education to transform lives, societies and economies. The only consolation which has become a sort of mantra is the argument that Spain has prioritised equity over excellence, a misleading message that has been reinforced by PISA (Gomendio, 2021).

So why does PISA have such a media impact if no one cares? The most likely explanation is that Spain does not have national assessments, nor regional evaluations with common standards. Thus, PISA scores represent the only information available to compare regions using the same metric and to evaluate trends over time. The first issue attracts a lot of attention since there are major differences in performance between regions that give rise to intense political bickering about the causes, which inevitably ends up escalating into conflicts on the mechanisms by which central government transfers funds to regions for education, health and social matters (Gomendio, 2021; Wert, 2019). It seems fair to say that in Spain, PISA is not as valuable as an international benchmark as it is in other countries, since the country (but not all regions) seems resigned to its fate as a mediocre performer. Instead, it is a powerful weapon in the invariably heated political debate surrounding education.

The picture that different ILSAs paint about the performance of Spanish students is consistent in terms of relative performance when compared to other countries, but not in their assessment of trends over time. The three major surveys show that Spain performs below around twenty OECD countries and much lower than top performers in East Asia. Spain joined PISA in its first cycle (2000) and scored below the OECD average until 2015, when it finally reached OECD average levels (the OECD initially withdrew the results for Spain in the launch of PISA 2018 due to their unreliability, and published them months later warning that they are not comparable to previous cycles, as we have discussed before). Spanish students exhibit particularly bad level of

performance in mathematics when compared to other OECD countries, mainly due to the substantially low proportion of top-performing students (Gomendio, 2021; Wert, 2019). Thus, one of the weaknesses of the education system in Spain is that it does not allow top-performing students to develop their potential. The evidence from PIRLS and TIMSS is broadly similar for primary school students: Spanish students perform slightly below the OECD average in science and reading, and much lower than it in mathematics.

But ILSAs draw different conclusions when trends are analysed over time: according to PISA, student performance remained stagnant from 2000 until 2015, while PIRLS and TIMSS show greater improvements. In mathematics and science, Spain improved from 2011 until 2015 and experienced a clear improvement in reading in 2016, after a lack of any progression between 2006 and 2011 (Gomendio 2021; Wert 2019).

In terms of equity, from its very first cycle, PISA identified Spain as a champion of equity, a conclusion that successive cycles have strengthened (OECD, 2001). This conclusion is based on one of the measures which PISA often uses to evaluate equity: the variance explained by between- and within-school variation. PISA assumes that large between-school variation is a reliable indicator of inequity because it reflects the extent to which differences in student performance arise from attendance at different types of schools, either due to early tracking (academic versus vocational schools, which happens in very few countries when students are fifteen years of age), major differences in wealth between neighbourhoods, or different admission policies by schools. Spain consistently ranks as having one of the lowest levels of between-school variation, since most of its variance is explained by within-school variation. An early OECD report investigating levels of equity in the Spanish education system reinforced this conclusion: “the Spanish results point to a low level of inequity that is more centred on differences within the institutions than between institutions” (Calero, 2005). Unfortunately, the conclusion by PISA is wrong on two counts: first, the variable used as an indicator of inequity is not appropriate, and second, there are many other variables that show high levels of inequity which PISA either does not measure or chooses to ignore.

As we have discussed in the case of Latin America, the reason why most of the variance is explained by within-school variation is that rates

of grade repetition are high in Spain (from 2000 until 2011 it was almost 40%, three times that of the OECD average). Since PISA assesses fifteen-year-olds irrespective of the grade in which they study, the survey does identify and quantify this problem, showing that 67.9% of fifteen-year-olds were in tenth grade, 23.4% were one year behind and 8.6% were two years behind in 2015 (OECD, 2016b and 2016c). PISA data clearly show that students who repeat a grade score much lower than those who do not. Hence, it is clear that having fifteen-year-old students in the same school performing at such different levels overrides any differences that may exist between schools.<sup>1</sup>

The relevance of this issue goes far beyond the flaws in the use and interpretation of this variable in the context of equity, because PISA fails to recognise both the causes and consequences of grade repetition. The conclusion that PISA draws is oversimplistic: since grade repetition is associated with lower levels of performance, it systematically recommends avoiding this practice. The problem, of course, is that some education systems use grade repetition as a last resort when students are lagging so far behind that they can hardly follow what is being taught in the classroom and the magnitude of the delay is such that it prevents any learning progress.

Thus, grade repetition is a symptom rather than a cause of severe underperformance. It reveals the inability of the education system to allow struggling students to catch up at earlier stages. It also has dire consequences for students, since grade repetition is a reliable proxy of early school leaving, which is an endemic problem in Spain. The level of early school leaving remained incredibly high for decades (around 30%) with many of these students dropping out of school without even a lower-secondary diploma (Gomendio, 2021; Wert, 2019). Most early school leavers come from disadvantaged and migrant backgrounds. Thus, an education system which has been praised by PISA as being a model of equity actually leads to the worst type of inequity: the expulsion of disadvantaged students who have not acquired the most basic levels of knowledge and skills from an education system which

---

1 Actually, in the case of Spain in PISA 2015, students who had not repeated any grade scored ninety points above those who had repeated one or two grades (OECD, 2016a).

basically gave up on them from the very beginning. But PISA seems blind to this kind of inequity.

Huge regional differences are another major source of inequity. In PISA 2015, the difference between the top-performing region and the bottom-performing region was equivalent to almost two years of schooling. Among the seventeen regions, the variation was such that, if they were to be sorted with national entities in PISA rankings, some would be considered top performers (Castille and León and Madrid have similar levels of performance to some East Asian countries such as Hong-Kong and South Korea), while others perform well below the OECD average. Huge regional disparities in levels of student performance (as measured by PISA among fifteen-year-olds) have a huge impact on their lives: students in regions with lower levels of performance overall suffer higher rates of grade repetition, which lead to high rates of early school leaving and very high rates of youth unemployment (Gomendio, 2021; Wert, 2019).

The fact that the Spanish education system generates such terrible outcomes in terms of equity represents a big failure of the education models (explicitly designed to achieve high levels of equity) which have been implemented there during the last decades. So why this profound contradiction between aims and outcomes?

The evaluation of the outcomes leaves only one plausible explanation: despite all of the good intentions, the policies which were implemented were wrong. This is relevant for many other countries, because some of these policies have been associated with equity elsewhere, but they did not work in the Spanish context. We will review the major education policies in Spain, their impact, and the reaction by PISA in the following chapter.

## 5.6 Conclusions

It seems reasonable to argue that robust, reliable and objective evidence about which education policies lead to better student performance should guide policymakers and help them overcome any divisive ideological issues or strong conflicts of interest. Most of the information available comes from ILSAs and, since PISA from its inception adopted a policy advisor role, it boasts about the impact of its policy recommendations

on governments and the reforms that they implement. However, when we look at the impact of PISA on actual education reforms, the evidence suggests that this is small.

There is much consistency between ILSAs in terms of the levels of student performance attributed to countries, and how they compare to each other. Thus, the rankings do provide important and reliable information about which countries should be analysed in order to understand which education policies lead to outstanding levels of student performance, and which should be analysed to understand which combination of policies leads to poor levels of student performance. Over time, more countries have become participants of ILSAs, so care must be taken to understand the precise contexts in which countries are labelled as “top” or “low” performers in different cycles.

A detailed review of top performers reveals that Finland became *the* legend after the first PISA cycle, when it was the top performer in reading (but not in science or mathematics). In this same cycle, Germany experienced the so-called ‘PISA shock’, owing to the huge mismatch between expectations and results, which led to media outrage followed by a number of reforms. Since this was the very first cycle, PISA had not developed any policy recommendations. In our view, the vivid contrast between the unexpected success of a small and humble European nation like Finland and the wounded pride of powerful Germany was crucial to the creation of a narrative which remains almost intact and very influential today. In 2000, the education system in Finland was described as comprehensive (no early tracking), having a high-quality teaching force, high levels of school autonomy and being equitable. In contrast, Germany had early streaming of students into vocational and academic tracks at the age of ten years and a greater impact of family socio-economic background and immigrant status on student performance.

As with all influential narratives, the one that emerged from this comparison was rather simplistic: education reforms should address inequalities before student excellence can be achieved, and the way to do this is by designing comprehensive systems which do not segregate students into academic vs vocational tracks, ability grouping, or different types of schools (e.g. charter vs public). High levels of school autonomy and good-quality teachers were combined in a conclusion

which proved to be far-reaching: teachers should be trusted. Eventually this concept got lost in translation and unions all over the world argued that PISA demonstrated that teachers should not be evaluated, should not be asked to follow a pre-defined curriculum, and should be allowed to innovate. PISA has contributed to this narrative, arguing that teachers should own their profession, establish their own standards and design education reforms.

As this narrative crystallised and became more popular, the data collected over successive cycles clearly challenged some of its conclusions. First, the initial success of Finland was followed by a decline in student performance over time. This suggests that while Finland was basking in the glory of its success it failed to recognise the need for reforms. Thus, PISA may be as influential in promoting reforms as it is in preventing them, a potential role that has been largely ignored. Second, it has been argued that the high performance of fifteen-year-olds in the year 2000 is related to the education system that was in place before then, which was more centralised and gave less autonomy to schools. This suggests that some of the causal relationships established are wrong, because this time lag has not been taken into account. Similarly, a high-quality teaching force had been the focus of much attention long before PISA. Third, there is also plenty of evidence that in Finland reading literacy had historically been high, and society and families contribute substantially to early learning before students start school. This is a good example of the extent to which PISA outcomes may be the result of what happens outside of school. While this is something that PISA recognises, it does not seem to accept the limitations of policy recommendations which address the education system.

In the small group of countries that experienced the PISA shock, such as Germany, the findings of successive cycles also challenge the view that PISA has played a major role in supporting these countries to improve student performance. Such countries, like Germany, did implement reforms, but none of them improved student outcomes substantially over time. Thus, the impact of PISA was obvious in terms of media visibility, the main variable which it uses to evaluate its influence, but unclear in terms of the positive impact on student performance.

The real education superpowers, the countries which have shown outstanding levels of performance and which continue to improve

over time, are countries in East Asia. Despite this indisputable success, no narrative has been developed around them. These countries were top performers from their very first participation in ILSAs, and their journey towards peak performance started decades before, so it cannot be argued that they improved due to PISA's policy recommendations. Perhaps the idea that students perform better because of the pressures that families exert on their children had led to the mistaken conclusion that not much could be learned by education systems in other parts of the world where education is not as highly valued. It is also possible that the fact that some of the policies implemented by these countries clearly contradict the dichotomous narrative (comprehensive vs segregation, trust vs mistrust on teachers) mentioned above has downplayed any lessons than can be learned.

Countries like South Korea and Singapore demonstrate that major improvements can take place much more rapidly than previously assumed. In a few decades, they have evolved from illiterate societies into the best-performing education systems in the world, but this requires long-term vision and consistency. The evolution of these education systems over time clearly shows that a trade-off between teacher quality and class size has been made, and that this delivers excellent results. This has been possible because unions do not have the power to block reforms, so admission processes for university degrees are highly selective, training is of a high standard, entry into the profession is based on merit and highly competitive, teacher professional development is well-structured, and teachers are evaluated on the basis of the performance of their students. Teachers are valued and enjoy high prestige, but the misleading concept of trust does not apply: high curricular standards as well as student assessments are defined centrally, and schools do not enjoy high levels of autonomy. Even more challenging to the dominant narrative is the fact that, by implementing tracking as early as at primary level, Singapore successfully lowered high rates of early school leaving and, in clear opposition to comprehensive policies, has kept early tracking at lower-secondary level to this day. This model does not prevent all students from achieving their full potential and the different pathways represent an efficient way for teachers to successfully manage student diversity.



The region which represents the biggest divergence from East Asia in terms of developing its human potential is Latin America. Despite a better starting point around fifty years ago, at present the performance of its fifteen-year-olds is many years of schooling behind countries in East Asia and well below the OECD average. In a nutshell, the policies implemented are in many ways the opposite of those which have succeeded in East Asia. The effort in terms of increased access to education and higher rates of enrolment in higher levels of educational attainment has been huge, but the returns in terms of the knowledge and skills acquired have been very low. Much investment has gone into decreasing class size, which is supported by unions and parents alike, but does not have any positive impact on student performance. The power of unions in the region is unparalleled, and they have forcefully rejected attempts to improve teacher quality, systematically opposing teacher (and student) evaluations and higher standards. Most education systems are comprehensive, with no early tracking and few students choosing VET. This is probably the result of the hopes that families in societies with high levels of inequity place on education as an engine of social mobility, of the pressures by strong social movements in the region, and of policymakers' desire to be seen as facilitating such aspirations. But these policies have failed.

Obviously good student outcomes cannot be achieved by teachers with low levels of skills, no matter how small the classes are. In addition, in societies with high levels of inequity, education systems need to deal with a very diverse student population. The comprehensive policies, which may work in egalitarian societies with high levels of skills, have proven unable to deal with high levels of diversity in the classroom, leaving those students with difficult starting points without the differentiated support they need, which leads to high rates of grade repetition and early school leaving. In the context of large inequity and diversity, the outcome of comprehensive policies is to expel students who are struggling. This seems the worst form of segregation.

Those countries which have improved after participating in ILSAs may give better insights into which policies contributed to success, even if most of them also started their reforms well before joining ILSAs. These are all European countries which followed (in slightly different ways and at different times) what seem to us elementary good practices:

improve teacher quality, define a coherent curriculum with high standards, implement student assessments which are well-aligned with the curriculum, modernise and develop VET, and give more autonomy to schools in exchange for accountability. While ILSAs have played a crucial role in showing positive trends in student performance over time, this evidence has not prevented most of these reforms from being reversed after they have proven to be successful.

Finally, Spain is another sad case of successful evidence-based reforms followed by reversals. Despite major budget cuts since 2009 due to the financial crisis, in 2015 and 2016 improvements were detected by PISA and to a larger extent by TIMSS and PIRLS among primary students, the educational stage where implementation started.

We believe that is important to emphasise that the Spanish case reveals some important weaknesses of PISA. First, PISA defines the two dimensions which matter in education systems (quality and equity), but while quality is measured straightforwardly (via student outcomes), equity is measured in many different ways, because no single variable can capture all of the layers of this complex dimension. Second, some of the most commonly-used variables developed by PISA as indicators of equity are misleading or interpreted in the wrong way. For example, it is a mistake to use small between-school variation in Spain to conclude that the system is equitable, because high rates of grade repetition make intra-school variation much higher than in other education systems. Third, variables which are either ignored or not measured by PISA clearly show that the Spanish education system suffers from high levels of inequity, because high rates of grade repetition lead to high rates of early school leaving, to the extent that one in every four students is expelled from the education system. Most dropouts are disadvantaged students.

The legend that the Spanish system has prioritised equity over excellence has helped policymakers disguise the system's poor quality and has justified the lack of reforms over decades, which has led to stagnation in student outcomes. Since PISA has supported this mistaken conclusion, it should take responsibility for the stagnation of student performance in Spain. Finally, PISA has refused to give any detailed or convincing explanation for the initial withdrawal and later publication of unreliable results for Spain in 2018, which were used for political

purposes. PISA is a high-stakes exam for policymakers, of the kind it no longer supports for students, but it should be held accountable if it wishes to be regarded as a trusted source of data.

