



## FEATURE

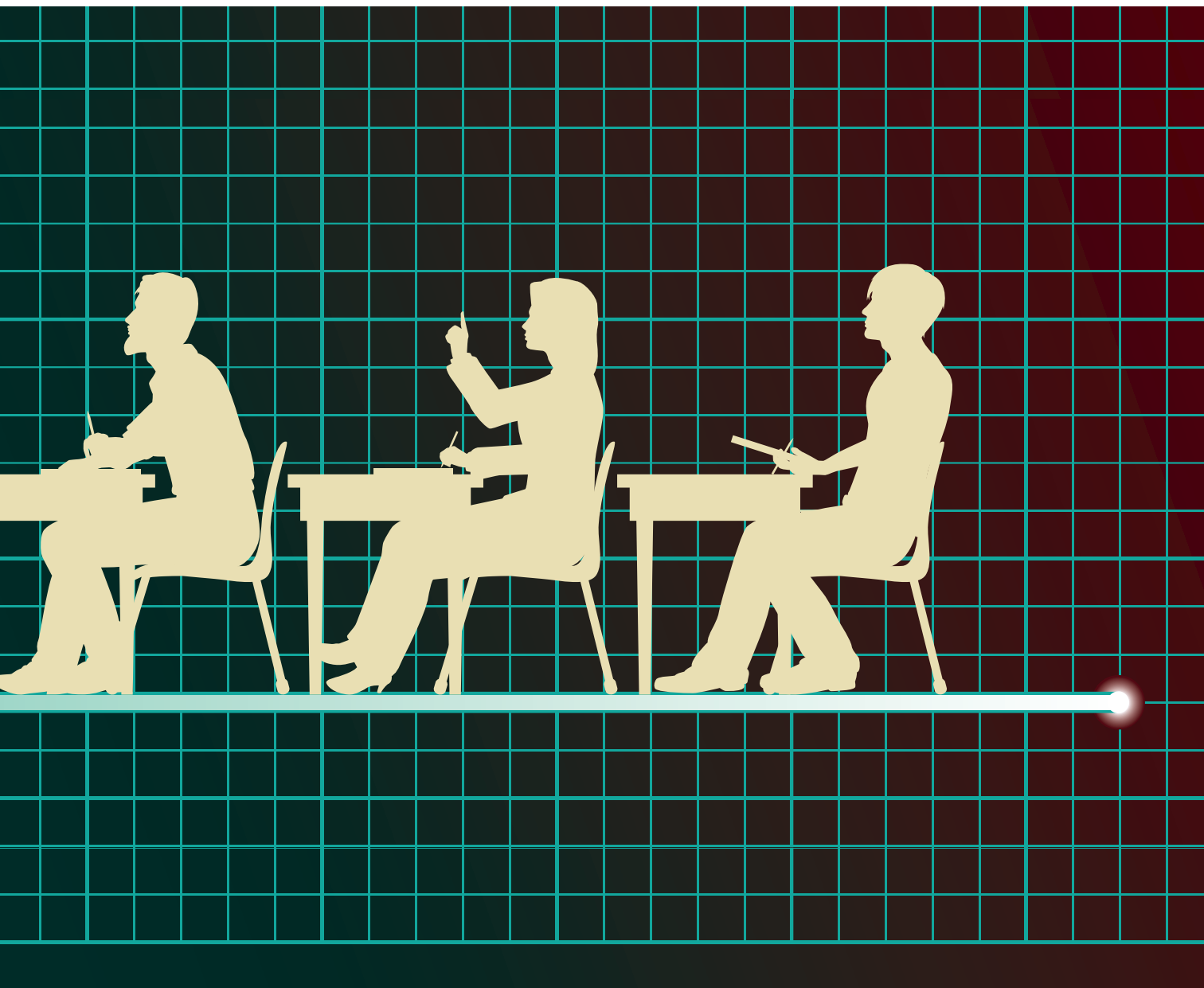
# PISA: Mission Failure

*With so much evidence from student testing, why do education systems continue to struggle?*

By MONTSE GOMENDIO

**I**N THE CONTENTIOUS WORLD of education, nearly every proposed reform has its detractors and supporters. Yet common sense might indicate that a policy backed by solid evidence would foster agreement between policymakers, governments, political parties, and education stakeholders. Shouldn't objective data override ideological divides and political bickering?

Many reformers have looked to assessment and accountability, both within countries and internationally, as a means of encouraging consensus. On the global scene, their hope was that the evidence generated by international assessments could contribute to our common understanding of what works in different countries, since comparative data can identify which policies have boosted student achievement in top-performing nations.



Unfortunately, these expectations have not been met.

Since 2000, the Programme for International Student Assessment, or PISA, has tested 15-years-olds throughout the world in reading, math, and science. Developed by the Organization for Economic Cooperation and Development, or OECD, and administered every three years, PISA is designed to yield evidence for governments on which education policies deliver better learning outcomes as students approach the end of secondary school. The OECD is a member-led organization of nations that provides policy advice to governments and encourages peer learning between countries. Initially, PISA testing involved only the rather homogeneous group of OECD member countries, but its ambition grew. From the first cycle (2000) to the last (2018), the number of participating countries increased from 32 to 79, owing largely to the addition

of many low- and middle-income countries. At this point the OECD asserted that “PISA has become the world’s premier yardstick for evaluating the quality, equity and efficiency of school systems, and an influential force for education reform.”

And yet, according to PISA’s own data, after almost two decades of testing, student outcomes have not improved overall in OECD nations or most other participating countries. Of course, that same time period saw a global recession, the rise of social media, and other developments that may have served as headwinds for school-improvement efforts. Even so, PISA’s failure to achieve its mission has led to some blame games. In an effort to explain the flatness of student outcomes over PISA’s lifetime, the OECD asserted in a report on the 2018 test results that PISA “has helped policy makers lower the cost of political action by backing difficult decisions with

evidence—but it has also raised the political cost of inaction by exposing areas where policy and practice are unsatisfactory.” The OECD was essentially pointing the finger at its own members and other countries participating in PISA, accusing them of not following PISA’s policy advice.

This finger pointing is based on two assumptions: first and foremost, that PISA policy recommendations are sound, and second, that the evidence provided by PISA data is itself enough to reduce the political costs associated with implementing education reforms.

Both assumptions are seriously flawed. My professional experience as an academic and national education minister allows me to look at this issue from a unique vantage point. When I served as Spain’s secretary of state for education, I became keenly aware of the political pushback that education reforms face, how and why that pushback remains hidden from public debate, and the helplessness policymakers feel when they try to ameliorate differences of opinion by bringing objective evidence to the table. As deputy director for education at the OECD and later head of its Centre for Skills,

outrage. This happened in Germany in the first PISA cycle, and, as the OECD wrote in a 2011 report, “the uproar in the press reflected a very strong reaction to the PISA results. . . . Politicians who ignored it risked their careers.”

Politicians around the world do view PISA as a high-stakes exam that leads to intense media scrutiny and political blame games. But surely the only measure that truly reflects PISA’s success is its ability to shape reforms that improve student outcomes. As we have seen, trends over time reveal a flat line, so what went wrong?

### Quality and Equity

Policy recommendations from PISA are based on a combination of two different approaches: 1) quantitative analyses that search for links between student outcomes and a range of features of education systems and 2) qualitative analyses of low- and top-performing countries. Many critics have noted that PISA’s quantitative analyses cannot be used to draw causal inferences, mainly because of the cross-sectional nature of the samples and the almost-exclusive use of correlations.

**According to PISA’s own data, after almost two decades  
of testing, student outcomes have not improved  
overall in OECD nations or most other participating countries.**

I enjoyed the privilege of providing advice to governments all over the world, which allowed me to observe how much the success of specific policies and the magnitude of the political costs associated with implementing those policies differ between countries.

PISA has proven to be a successful metric for comparing education systems, a challenge that many thought impossible. The fact that the PISA ranking of countries by student performance is similar to the rankings generated by other international assessments has been used both to argue that PISA is robust and to question the need for another test. But PISA is *different*, mainly because, within the OECD framework, its role was predefined as a tool for policy advice, and it enjoys the privilege of direct communication channels with governments. Unlike the sponsors of other assessments, PISA officials work tirelessly to enhance the program’s media impact, a strategy that has two closely linked objectives: to magnify PISA’s visibility and to put pressure on governments to follow its recommendations. Clearly, PISA has a better chance of achieving these goals when exposed weaknesses in an education system provoke a media furor. Program officials seem particularly proud of the “PISA shock” that occurs when unexpectedly poor results in a country lead to media

Meanwhile, its qualitative analyses also suffer from serious drawbacks such as cherry-picking. While these issues are well known, others have gone largely unnoticed.

PISA seeks to measure two complementary dimensions of education systems: quality and equity. While quality is typically measured in a straightforward way—that is, in terms of average student test scores—equity is a multidimensional concept that PISA measures using metrics such as the relationship between socioeconomic status and student performance, the degree of differences in student performance within and between schools, and many others. The problem is that none of these variables tell the full story, each of them leads to different conclusions, and PISA’s prism on equity is ultimately too narrow.

To illustrate this point, I turn to my own country, Spain. From the very first cycle, PISA has hailed the Spanish school system as a paragon of equity. In fact, the praise has gone as far as to suggest that Spain has prioritized equity over excellence, a choice that PISA officials have applauded and domestic policymakers have used as an alibi to downplay the poor overall performance of Spanish students. PISA deems Spanish education to be equitable based on the finding that most of the variance in student performance in the country occurs within

rather than between schools, a result it interprets as revealing no major differences between neighborhoods based on wealth or between schools based on their selectivity. But there is an alternative interpretation: The equity metric that PISA has chosen to highlight is not appropriate in a country with high rates of grade repetition. Variation within schools is large because PISA tests 15-year-olds irrespective of their grade level. That means that Spain tests a large proportion of students who are one or several grades behind because they have repeated grades at least once. The additional problem is that focusing on a single variable while ignoring the bigger picture leads to mistaken conclusions. Grade repetition in Spain is a reliable proxy for early school leaving, which, in turn, leads to a high rate of youth unemployment and a large number of individuals who are not in school, the workforce, or training.

Unfortunately, in Spain the dropout rate has hovered around 30 percent for decades, and when I became secretary of state for education in 2012, at the peak of the financial crisis, the rate of youth unemployment was above 50 percent. It is simply wrong to define as equitable an education system where nearly one in every three students (most of them disadvantaged students or migrants) drops out of school without a minimum level of knowledge and skills.

Labeling Spain's school system as equitable is not an isolated case of misdiagnosis, since PISA also defines as equitable the education systems in countries such as Brazil, China, Mexico, and Vietnam, where a substantial proportion of 15-year-olds do not attend school, either because they never did or because they dropped out. It is mistaken to suggest that lessons about equity can be drawn from these countries.

### Wrongheaded Recommendations

These mistakes mean PISA incorrectly identifies the countries that should serve as role models, but what really matters is the policy recommendations PISA develops after comparing many countries. In a nutshell, out of concern for equity, the program warns against the implementation of any measures that could lead to segregation, such as ability grouping, school choice, and early tracking. This advice seems to be influenced more by ideology than evidence, since none of PISA's own statistical analyses justify such recommendations.

Consider the case of vocational education and training. PISA's conclusion is that it lowers student performance in the subjects tested by the program—reading, math, and science;

thus, PISA's recommendation is to postpone vocational education until upper secondary school to minimize the harm. However, the vast majority of participating countries already follow this practice, stipulating that students cannot choose vocational education until the age of 16. Since PISA assesses 15-year-olds, the number of vocational students it tests in most countries is zero. In those few countries where students follow different tracks at younger ages, the results do



LUC BOEGLY

*The Organisation for Economic Cooperation and Development (OECD) headquarters in Paris is where PISA was developed to assess students in reading, math, and science.*

not always support the conclusion that vocational students perform less well. Thus, PISA is poorly positioned to provide policy recommendations on this topic.

Another questionable policy recommendation from PISA concerns school choice, about which the OECD concludes that, after correcting for socioeconomic status, students do not perform better in private schools than in public schools. These analyses, however, lump private schools together with government-funded, privately managed charter schools, thus making it impossible to draw separate conclusions about charter schools, which in many countries are the real target of controversy. More elaborate analyses using data from many international assessments, as well as other studies, have concluded that school choice often does lead to better student outcomes without necessarily generating segregation and that some of the few countries with early tracking show little (if any)

differences in student performance and employability rates for vocational-education students. PISA needs to pay more attention to academic research and look at the broader picture.

PISA's qualitative analyses rely heavily on differences between Nordic countries and others. In particular, the sharp contrast in PISA's first cycle between the unexpected success of Finland and the unexpected poor performance of Germany has crystallized into an influential legend: that inclusive policies in place in Finland at the time led to both quality and equity and should be emulated, while the heavily tracked system in Germany led to inequity and should be avoided. Nordic societies were egalitarian long before PISA started, however. The alternative explanation is that in egalitarian societies teachers deal with a rather uniform student population, and therefore these countries can, without much risk, implement inclusive policies that tend to treat all students similarly. In contrast, less-egalitarian societies may require differentiated approaches and policies to meet the challenges that come with a heterogeneous student population. A number of comparative analyses show a correlation between the degree of economic inequality and the extent of disparities in student outcomes. Thus, the right question to ask is: To what

Spain implemented a rather radical comprehensive reform that delayed the start of the vocational-education track by two years (moving the starting age from 14 to 16) and avoided any differential treatment of students until the age of 16. This system was designed, as the OECD recognizes, for the sake of equity. But it failed: early school leaving increased as 14-year-olds no longer had a vocational option.

Latin America is also a region where levels of inequality are very high. Most countries there follow the egalitarian rules (no early tracking, no ability grouping, almost-nonexistent vocational education), leading to poor educational outcomes: low student achievement and high rates of grade repetition and early school leaving. In these countries, more than 70 percent of teachers and principals report that broad heterogeneity in students' ability levels within classrooms is the main barrier to learning.

### Political Pushback

These examples point to a broader conclusion: policy recommendations cannot be universal, because what works in egalitarian societies may lead to bad outcomes in societies with high levels of inequity. Education systems should instead follow a sequence of steps as they mature. Singapore shows the

**It is wrong to define as equitable an education system where nearly one in every three students drops out of school without a minimum level of knowledge and skills.**

extent can education systems compensate for large social, economic, and skills inequalities, and how?

I will return briefly to Spain which, compared to Nordic countries, is a rather inequitable society, not just economically but also in terms of skills. According to the OECD's Survey of Adult Skills, adults in Spain have low skill levels compared to their counterparts in most European countries. What's more, because in Spain universal access to education came about relatively late and the dropout rate has been high for decades, older Spaniards have very low skill levels, as do the relatively large proportion of adults of all ages who dropped out of school early. Among populations with such a lopsided distribution of skills, children entering school have very different starting points, levels of support at home, and access to resources. For teachers to be effective, it may be necessary to adopt practices that reduce student heterogeneity through the use of ability grouping or, in more extreme cases, different tracks. If these measures are not implemented early enough, students who are behind when they start school may not be able to catch up with their peers and, as they lag farther and farther behind, may end up repeating grades. In the 1990s,

way. A few decades ago, Singapore had an illiterate population and very few natural resources. The country made a decision to invest in human capital as the engine of economic growth and prosperity, and, in a few decades, it became the top performer in all international assessment programs, thanks to an excellent and evolving education system (see Figure 1). But PISA does not draw any lessons from the fact that Singapore started improving by implementing tracking in primary school in an effort to decrease its high dropout rate. Once this was achieved, the country delayed tracking until the end of primary school. Even today, however, Singapore remains one of the few countries with early tracking, along with Austria, Germany, the Netherlands, and Switzerland.

Singapore is one of the education superpowers of East Asia, a group that also includes South Korea, Japan, Taiwan, Hong Kong, and certain regions within China. While Finland was PISA's top performer in reading in the first cycle (when a small number of countries was compared), student outcomes in that country have since declined. In contrast, these East Asian countries consistently outperform other nations—particularly in math and science—and their extraordinary



outcomes continue to improve. The comparison between this group and the low-performing countries in Latin America (that is, countries on the opposite poles in the PISA rankings) is useful in examining PISA's second assumption: that the evidence provided by PISA data is itself enough to minimize the political costs associated with implementing education reforms.

Teacher quality is widely recognized as key to both the success of East Asian countries and the failure of Latin American countries. In East Asia, only top-performing students can enter education-degree programs, and, throughout their careers, teachers continue to develop their skills via demanding professional-development pathways. This emphasis on teachers' lifelong learning means that they spend less time in the classroom, a trade-off that leads to large class sizes. In contrast, in Latin American countries, students in education-degree programs are academically weak, selection mechanisms to enter the profession are ineffective, and accountability mechanisms are almost nonexistent. As a result, teachers tend to have high levels of skills in East Asian countries and weaker skills in Latin American countries.

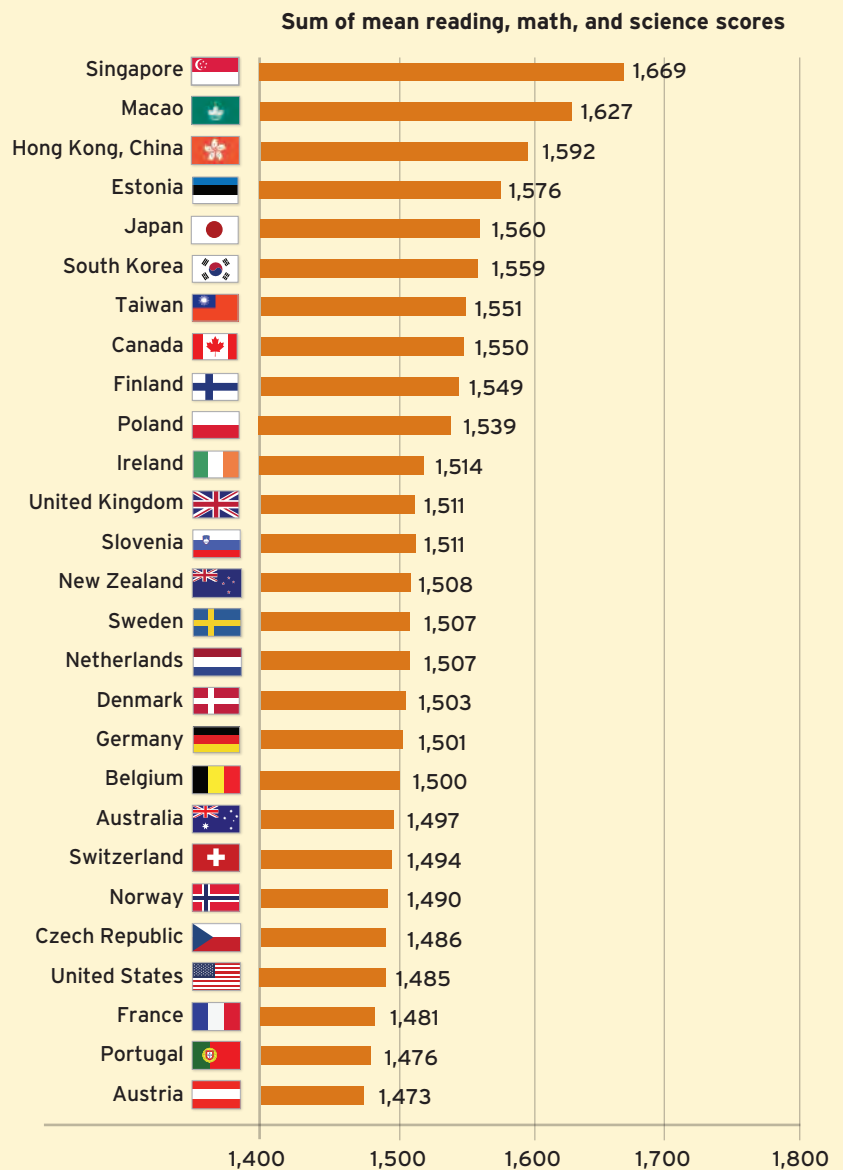
There is widespread recognition that the main constraint to raising teacher quality in Latin America is political. Unions in the region are very powerful by global standards, and they put huge pressures on governments to defend their interests, among which small class size is prominent. Smaller classes mean more teachers and more union members. A larger membership results in greater monetary resources and the increased power that comes with them. In contrast, union power in top-performing East Asian countries is very weak. This crucial difference is what makes the implementation of certain policies (such as large class sizes or rigorous teacher training and stricter selection mechanisms) very costly in political terms in Latin America, while such political costs barely exist among top-performing countries in East Asia.

The evidence from PISA on class size is one of the most robust results about what does not work in education. Decreasing class size uses up a vast amount of resources and seems to have no impact on student

## Top-Performing Countries on the 2018 Programme for International Student Assessment

(Figure 1)

Southeast Asian countries and Estonia top the list. The United States clocks in at number 24.



NOTE: The OECD did not include Spain's reading-test results in its report. Major changes in the methodology of PISA 2018 made those data unreliable.

SOURCE: OECD, PISA 2018 Results (Volume I): What Students Know and Can Do

## In Latin America, more than 70 percent of teachers and principals report that broad heterogeneity in students' ability levels within classrooms is the main barrier to learning.

performance at the system level, so PISA's policy recommendation has been to increase class size. However, many countries (including OECD members) have not acted on this evidence-based recommendation. They have continued to reduce class size over time because of the huge political costs of not doing so. Most increases in education spending have therefore gone toward a strategy that has no impact on student outcomes. This example suggests that evidence, no matter how robust, is unlikely to diminish the high political costs associated with reforms that result in the redistribution of the vast resources (and power) that education systems command.

PISA seems to misunderstand the nature of the political costs that reformers face. Those who oppose change are not resisting it because they haven't been convinced of the merits of the reforms. Evidence won't change their position. Decreases (or lack of increases) in investment generate a head-on conflict with the vested interests of unions and other stakeholders that will strongly oppose policies that reduce the resources these players receive. These vested interests tend to be hidden in the political debate, since pressures to decrease class size in order to increase the number of teachers are often presented as attempts to improve the quality of education.

### Mistaken Assumptions

In conclusion, PISA's two assumptions—that PISA's policy recommendations are right and that the evidence provided by PISA data is enough to minimize the political costs of attempting education reform—are flawed. First, some of PISA's conclusions are based on weak evidence. The greater problem, though, is that most policy recommendations are strongly context-dependent, and PISA's recommendations may be difficult for policymakers to interpret correctly if they lack precise knowledge of their education system's state

of maturity. Ignoring this fact and making universal policy recommendations has dire consequences for many countries, particularly those most in need. It would be much more helpful for PISA to look at countries that have achieved gains and try to extract lessons for other countries that had similar



GARETH FULLER / AP

*Students in London sit for their PISA exams in 2017. Although the United Kingdom was among the top-scoring countries, Asian nations like China and Singapore performed better.*

starting points when they joined PISA but have not improved.

Policymakers should remain aware, though, that reforms cause intense clashes of interest when resources are redistributed. That is especially the case when powerful unions are among the losers. Evidence has nothing to do with the nature of such conflicts. Those reformers who have tried and failed when confronted with such huge political costs need better advice from PISA, not a reprimand.

*Montse Gomendio is a research professor at the Spanish Research Council. Formerly, she served as Spain's secretary of state for education, as OECD's deputy director for education, and as head of its Centre for Skills. She is co-author of Dire Straits: Education Reforms, Ideology, Vested Interests and Evidence (2023).*