

Setting up the Scene: Lessons Learned from PISA 2018 Statistics and Other International Student Assessments



Nuno Crato

Abstract PISA 2018 was the largest large-scale international assessment to date. Its results confirm the improvements of some countries, the challenges other countries face, and the decline observed in a few others. This chapter reflects on the detailed analyses of ten countries policies, constraints, and evolutions. It highlights key factors, such as investment, curriculum, teaching, and student assessment. And it concludes by arguing that curriculum coherence, an emphasis on knowledge, student observable outcomes, assessment, and public transparency are key elements. These elements are crucial both for education success in general and for its reflection on PISA and other international assessments.

1 Sixty-Six Years of International Large-Scale Assessments

Modern international surveys on student knowledge and skills can be traced back to the First International Mathematics Study, FIMS, held in 1964, involving 12 countries and organized by the International Association for the Evaluation of Educational Achievement, IEA. The IEA itself was founded in 1958 at the UNESCO Institute for Education in Hamburg, and since its inception had the ambition of providing reliable assessments of student outcomes.

The IEA further organized the First International Science Study, FISS, in 1970, the Six Subject Survey, in 1970/1971, the second studies in mathematics, the SIMS, in 1980, and the studies in science, the SISS, in 1983. Along the last two decades of the twentieth century, the IEA launched an additional series of international studies. These studies focused on subjects as diverse as civic education (1971) and written

The author was partially supported by the Project CEMAPRE/REM - UIDB/05069/2020 - financed by FCT/MCTES through national funds.

The original version of this chapter was revised: The belated corrections has been incorporated. The correction to this chapter is available at https://doi.org/10.1007/978-3-030-59031-4_13

N. Crato (✉)

Cemapre/REM, ISEG, University of Lisbon, Lisbon, Portugal

e-mail: ncrato@iseg.ulisboa.pt; ncrato@gmail.com

© The Author(s) 2021, corrected publication 2021

N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_1

composition (1984). However, the two most successful waves of international studies this Association organized were the TIMSS—with its acronym which could stand for the third wave of studies, but now denotes Trends in International Mathematics and Science Study—, and the PIRLS, Progress in International Reading Literacy Study.

TIMSS has been held every four years, starting in 1995, and PIRLS every five years, starting in 2001. At this time, the IEA further organizes the ICCS, International Civic and Citizenship Study, held every seven years, and the ICILS, International Computer and Information Literacy Study, held every five years. The last ICSS was done in 2016 and the last ICILS in 2018¹.

In 2000, the Organization for Economic Co-operation and Development, OECD, started the Program for International Student Assessment, PISA, which has become the best known of all international student surveys.

PISA is held every three years and encompasses three core domains: reading, mathematics, and science. Every wave or cycle of PISA is focused on one of these three domains, following thus a cycle of nine years. When PISA was designed, mandatory schooling in most OECD countries ended when students were about 15 years old. Thus, this survey was naturally geared towards assessing all students, those that continued their schooling, and those likely to soon enter the labour force. It was important to assess how prepared they were for this new stage in life.

In addition to PISA, OECD organizes, *inter alia*, PIAAC, a survey of adult skills, and TALIS, Teaching and Learning International Survey, a study directed to teachers and school principals with detailed questions regarding their beliefs and practices.

PISA, TIMSS and all these studies have been labelled as International Large-Scale Assessment studies, ILSA studies, and have a set of common characteristics. Country participation is voluntary, each country pays for the costs and organizes the application of the surveys, following common rules supervised by the promoting organization. Students are selected by a multi-stage random sampling method. Most test questions are confidential, in order to allow for its reuse across surveys for longitudinal calibration purposes.

Although each survey focuses on specific cognitive skills, each provides data on a large variety of issues, such as teaching methods, students' perception of their abilities, and social and economic students' background.

Two main differences between PISA, on one side, and TIMSS and PIRLS, on the other, are the selection of students and the intended measurements. While PISA is age-based, surveying 15-year-old student regardless of their grade and type of program they are following, TIMSS and PIRLS are grade-based—TIMSS is applied to 4th and 8th grade students and PIRLS to 4th grade students. While PISA tries to assess applied knowledge and skills, or literacy, in a generic sense, TIMSS aims to be curriculum-sensitive, and so tries to measure achievement based on an internationally agreed basic curriculum knowledge. While the OECD organizes PISA with specific ideas of what should be measured and specific ideas about the aims of education,

¹For the history of IEA and these studies see IEA (2018).

IEA organizes TIMSS to measure what each school system is achieving, taking into consideration each nation's curriculum and aims.

A few countries have been participating in some of these international tests for decades, thus having a series of results that allow for assessing progress over time and estimate the impact of educational policy measures that have been introduced. A large number of countries have participated consistently in PISA surveys, providing a moderately-long multivariate time series and a set of very rich contextual data that helps understand each country's evolution.

Although PISA and TIMSS have been criticised from diverse perspectives², the data they provide are so rich that they allow for various descriptive and correlational studies which shed light on many educational issues.

PISA and TIMSS data also allow for the observation and discussion of policy measures impact. Given the complexity of intervening factors, causality is always difficult to establish. But the time series are now longer than political cycles (usually four or five years) and longer than student's compulsory schooling life (usually nine to twelve years), and this allows the analysis of the impact of educational policies.

One excellent example is a study performed by one of the contributors to this volume and his co-authors; this study shows the impact of standardized testing on student cognitive skills³. Taking advantage of the panel data structure of countries and countries' performance across six PISA waves, from 2000 to 2015, authors show that "standardized testing with external comparison, both school-based and student-based, is associated with improvements in student achievement". They also reveal that such effect is stronger in low-performing countries and that relying on internal testing without a standardized external comparison doesn't lead to improvement in student achievement.

2 Pisa 2018

So far, the largest and most comprehensive of all ILSA studies has been PISA 2018. About 710 000 students from 79 participating countries and economies representing more than 31 million 15-year-old students performed the two-hour test⁴. This time, most of the students answered the questions on computer. The core domain was reading literacy, although the survey also covered the other two domains, mathematics and science⁵.

Having as a reference the cycle in which each domain was for the first time the major one and using results from the then participating OECD countries, PISA normalized initial scores by fitting approximately a Gaussian distribution with mean

²See, e.g. Araujo et al. (2017), Goldstein (2017), Sjöberg (2018), and Zhao (2020); and Hopfenbeck et al. (2018) and the references therein.

³Bergbauer et al. (2019).

⁴OECD (2019d), p. 12.

⁵For a quick overview, essential data are reported at OECD (2019c).

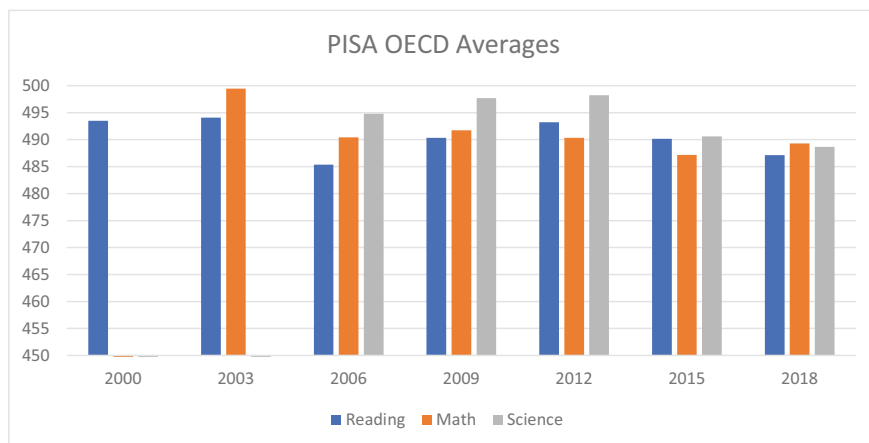


Fig. 1 Evolution of PISA Results for OECD Countries. PISA OECD countries averages include countries that have participate in all PISA waves. *Source* OECD IDE reports with recomputed updated data <https://nces.ed.gov/surveys/pisa/idepisa/report.aspx>

500 and standard deviation of 100 points for each domain. Now, the OECD mean scores are 487, 489, and 489, for reading, mathematics, and science, respectively. OECD countries results have been declining slightly, but in a steady way after 2009, as it can be seen in Fig. 1. Decreases are noticeable for Mathematics since 2003.

As Montserrat Gomendio discusses in this book in her Chapter on Spain, this is a worrisome fact.

Although it is difficult to translate PISA scores into years of schooling in order to estimate effect size of differences, various studies have suggested that a difference in 40 score points is roughly equivalent to a difference between two adjacent year grades. This estimate is an average across countries (OECD 2019a, p. 44)⁶.

If we use this estimate, we find noticeable changes between some waves, even if we only take into consideration OECD countries. For instance, the difference between the Math average scores in 2003 and 2012 amounts to a loss of about a quarter of a school year.

In order to simplify the interpretation of results, PISA scale is categorized into six ordinal proficiency levels. The minimum level is 1, although students can still score below the lower threshold of level 1. The maximum level is 6, with no ceiling. Mean scores are included in level 3.

Students scoring below level 2 are considered low-performers and those scoring above level 4 are considered high-performers. In 2009, recognizing the worrisome number of low performers in reading and the need to better discriminate those students, PISA has subdivided level 1 in 1.a and 1.b (OECD 2016a). In 2018, PISA introduced an additional third level, 1.c.

⁶In 2009 OECD estimated that 0.3 standard-deviation of the PISA scale was roughly equivalent to one school year (OECD 2009 p. 23).

In 2009, the European Union's strategic framework for co-operation in education and training set as goal for 2020 that "the share of low-achieving 15-year-olds in reading, mathematics and science should be less than 15%" (European Council 2009, pp. C 119/2-10). Low-achievers are de facto defined by the European offices as students scoring below level 2 in the PISA scale. This goal is far from achieved and is not in sight: the share of low performers in the European Union has been slightly increasing and in 2018 reached 21.7%, 22.4%, and 21.6%, in reading, mathematics and sciences, respectively.

In 2015, the United Nations defined in their Sustainable Development Goals for 2030 a "minimum level of proficiency" that all children should acquire in reading and mathematics by the end of secondary education (United Nations Statistics Division 2019, goal 4.1.1.). As the Pisa 2018 report indicates, this minimum level corresponds to proficiency level 2 (OECD 2019a, p. 105).

3 The Measurement Changes the Measured

To some extent, almost all participating countries have been affected by PISA, TIMSS and other ILSA studies. When the first cycle results appeared, some countries were shocked by seeing themselves in a relative mediocre position. Others were less surprised or less concerned. But with successive cycles of ILSA studies, every participant country started paying more attention to the results and to their country's comparative position.

Nowadays, the public disclosure of the results is carefully prepared by the ministries and authorities of each country; discussions follow in the press, at conferences, and in parliaments. Some try to minimize negative results portraying them as a product of biased measuring instruments. Some try to diffuse the negative results portraying them as consequences of general socio-economic problems or historical cultural handicaps. At the same time, a number of countries have been elated by their excellent results or praised for their relative improvement. Politicians try to get credit for the successes and educational analysts try to interpret results in the light of their ideological views. Serious researchers try to make sense of the results. No participant country has been completely indifferent to ILSA studies.

This phenomenon is clearly seen in each of the chapters that follow. Coming from countries as diverse as Chile, Taiwan and Portugal, Ema Lagos, Sue Lin and João Marôco describe how their countries have been awakened by poor results and how people started realizing the need for improvement.

In their Chapter on Chile, Ema Lagos explains how PISA studies were important to awake Chile to a recognition of its poor results, to the high disparity of scores in the country, and to the need to attain a general increase in school quality. These two authors also explain how PISA and TIMSS studies have helped modernize both the curricula and the national assessment system.

In her chapter about Spain, Montserrat Gomendio argues that the media impact of PISA is larger in Spain than in most other countries. The likely reason is that no national examinations exists in her country and so ILSAs are the only instrument

available to measure student performance in the whole country and to compare performance across regions.

This contrasts with Tim Oates' perspective on the context in England. With no longitudinal structure in PISA and only a quasi-longitudinal structure in TIMSS, the ILSAs are of secondary interest to policy makers in England, since the country maintains a high quality and comprehensive National Pupil Database (NPD). This contains school and pupil level data; including for each pupil the outcomes of national tests and examinations at primary and secondary levels. Nevertheless, PISA results receive public attention, as a consequence of the international comparison they provide, and the global prominence the results now possess.

4 Time Delay

When tested in PISA, youngsters have been in formal schooling for about 10 years of their lives. Their knowledge, skills, and conduct have been shaped by many teachers, curricula, tests, textbooks and other school factors. Most likely, successive governments and ministers have been in power and a few legislative and administrative settings have changed. Furthermore, the social and economic status of students and their peers, parents' education and many other factors have influenced students' results measured in PISA.

All this means that it is extremely difficult to disentangle the impact of educational policy changes from a very complex set of factors that have been put in place at different points in time. A hotly debated topic is the timeframe that should be adopted to try to measure the impact of specific policy changes⁷.

On one extreme, one can argue that any measure takes a long time to bring changes in education. Social-economic status and parents' education level are known as some of the most important factors explaining the variability of students' outcomes⁸. These factors certainly take generations to change, but they can be reversed by dynamic educational systems, as the spectacular improvement of some Asian countries has shown.

Apart from these generational slow changes, some education policy measures also take an incredible amount of time to impact education. Think, for instance on legislative changes on teachers' initial training requirements. Assume they are decided at year zero. They will impact students' choices through their selection of the high school appropriate courses in order to enter a chosen college program. Suppose the new prospective teachers enter college three years later, take five years to graduate and serve one year of an experimental contract before being hired as fully independent teachers. If these newly trained teachers start their careers teaching grade level 5, PISA results reflect this new teacher training requirements when students are at grade 10, i.e. 11 years after the legislative act.

⁷See e.g. Crato (2020).

⁸PISA 2018 confirms the importance of these factors. Main data syntheses are in volume II of the PISA report, OECD (2019b).

This example is not purely theoretical. As Arto Ahonen explains in his chapter on Finland, his country set a new high standard for teaching qualifications in 1979 when it “set a master’s degree as a qualification for all teachers, also at the primary”. Most analysts point to this measure as an important factor for subsequent Finnish successes.

When looking at 2018 PISA results, one is really looking at the impact of various generations’ education, plus the impact of decades of policy changes. Yes, in education some things take a long time to change.

On another extreme and in contrast to these long timeframes, some educational measures take a very short time to impact student’s performance. If, in September, a national mathematics test for 9th graders scheduled to May is abolished, it is conceivable that seven months later, in April, at the time of a PISA test, students would be more relaxed regarding their mathematics performance.

Indeed, on his chapter on Portugal, João Marôco points out that in 2016 the devaluation of external high-stakes assessments and the suggestion for trimming of learning targets may have reduced the effort and engagement of the Portuguese students with immediately subsequent low-stakes ILSA tests. In Portugal, significantly more students reported putting less effort on the PISA test than the OECD average.

João Marôco discusses further the evolution of Portugal and shows a very interesting graph, in which he displays a sequence of policy decisions taken since 2000 in parallel with the evolution of PISA scores. This gives very rich food for thought regarding the impact of policy measures.

In her chapter, Gunda Tire discusses the stunning successes of Estonia and explains that this country has not adapted its educational system to boost PISA outcomes, but rather that PISA results have helped to support policy measures this country has taken. She presents a very interesting table in which we clearly see how a sequence of policy measures parallels the results seen in PISA and TALIS.

In the chapter on Poland, Maciej Jakubowski explains that the evolution of scores from 2000 to 2003 was taken as a measure of success of the reform introduced in 1999. Then he proceeds to show how changes in curricula were followed by changes in students’ scores along these 18 years.

In the chapter on England, Tim Oates describes in detail his country’s education policy measures since 2010 and explains how these changes take time to be reflected in PISA results. Major changes took place in 2014, and they did not impact the PISA 2018 cohort.

5 Money Matters, Sometimes...

This is one of the most contentious topics in education. When one talks about investing in education, most likely one means, and is understood as meaning, financial investment. This is so common and pervasive that it almost sounds like a heresy to admit that additional funds may not be the central factor for improving education.

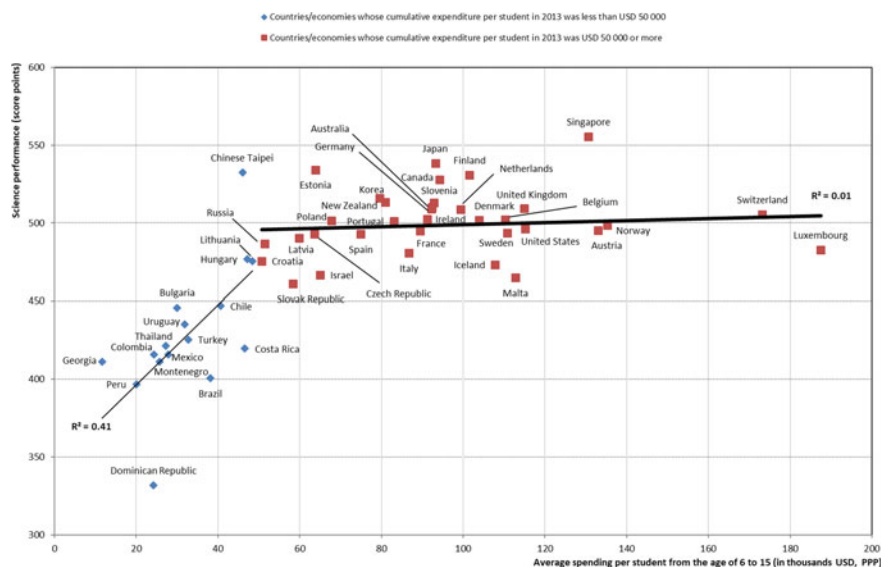


Fig. 2 Student scores in Sciences and spending in education. Only countries and economies with available data are shown. A significant relationship ($p < 0.10$) is shown by the thin line. A non-significant relationship ($p > 0.10$) is shown by the thick line. Source OECD (2016b), Figure II.6.2, p. 186; PISA 2015 Database, Tables I.2.3 and II.6.58. <http://dx.doi.org/10.1787/888933436215>

PISA and other international comparison studies have shown that reality is a bit more complex. Although always welcome, money is not essential for some important and beneficial improvements; the funding discussion obscures the real issues about education quality.

PISA 2015 was centred on sciences and it showed a graph that has circulated in educational circles and surprised many people. This graph is reproduced in Fig. 2. It plots student scores in sciences against cumulative educational expenditure per pupil adjusted for purchasing power parity (PPP). It clearly shows that spending is correlated with education quality until a certain spending point ($R^2 = 0.41$), after which it has a very weak nonsignificant correlation ($R^2 = 0.01$) with spending.

For some reason, PISA 2018 report discusses the same issue with a slightly different functional approach. Figure 3 is directly reproduced from the PISA report (Figure 1.4.4, OECD 2019a). This time, instead of a piecewise linear regression, the report adjusts a logarithmic function, which by nature is always monotonically increasing. A visual observation of data reveals essentially the same reality. Up to a certain level situated around OECD average (89,092 US Dollars per student), the increase in expenditure roughly parallels the improvement in reading results. After this level, there is no visible association. Again, Portugal and Poland outperform the Netherlands, Austria, and Luxembourg with one-third of the spending of the latter country. The example of Estonia is even more striking: it outperforms almost all countries that have a higher education expenditure.

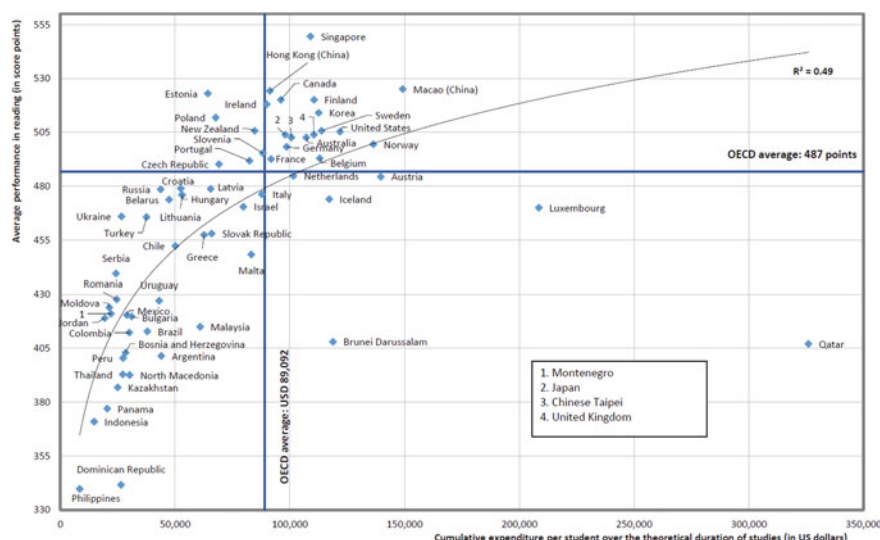


Fig. 3 Student scores in Reading and spending on education *Source* OECD (2019a), Figure I.4.3, p. 65; PISA 2018 Database, Tables I.B1.4 and B3.1.1. <https://doi.org/10.1787/888934028406>

All this means that a nuanced approach should be adopted as we discuss education spending. As Ema Lagos explains in her chapter on Chile, expenditure in education in their country is right on the expected level of the adjusted logarithmic function. And they correctly point out that other countries with a similar level of expenditure attain lower reading scores. As they also highlight, there are other “principles of action that could be beneficial to raise student performance”, such as “employing better qualified teachers and establishing educational outcomes as a main target.”

A similar point is made by Eric Hanushek in his chapter on the United States, a country that is at the extreme regarding expenditure: real spending per pupil more than quadrupled between 1960 and 2016 and student achievements registered little or no change over this long period of time.

In Portugal and Spain, the situation is even more revealing: in recent years, past improvements in PISA scores have been parallel to a decrease in public spending on education. It’s clear that other factors are at play.

In Spain, one may compare spending and scores both longitudinally and cross-sectionally, as there are many regions with different spending and different mean scores. As Montserrat Gomendio shows in her chapter, both analyses reveal no significant relationship between the two variables.

In the chapter on Australia, Sue Thomson argues that the problem is the lack of funding for the areas and schools that need more resources. This sets the problem at a completely different level and shows how education outcomes and spending need to be analysed beyond the macro level.

6 Performance and Inequality—Two Nonconflicting Poles

Another most controversial topic in education is the relation between performance and equity. Everyone agrees that educational policy “aims to maximize educational excellence and reduce inequity” (Parker et al. 2018). But there are different approaches to achieve this.

In reality, none of the aims make sense without the other. For a statistician this is trivial – location and dispersion are the ABC of statistical analysis. Excellence can increase in mean, while low performers get worse results. By the same token, inequality can be reduced at the expense of lowering everybody’s attainment.

However, it is very common to hear people debating either excellence or inequity. At first, people debated excellence. But lately, inequalities seem to be the sole priority.

In the following chapters, the authors debate these two sides of educational improvement. Some cases are worth mentioning.

The chapter on Australia offers a detailed view of the gaps between high performers and low performers and the gaps between various socioeconomic and ethnic groups. Sue Tomson describes her country’s decline in overall results and looks in detail at various asymmetries that contribute to the average results. She shows how some disadvantaged areas are additionally suffering with teacher absenteeism and a high percentage of inadequately or poorly qualified teaching staff. She doesn’t rejoice with the simple narrowing of gaps, recognizing that some are due to the “larger decline in the scores of high achieving students”.

The chapter on Chile has a very interesting discussion of related points. The authors describe both Chilean struggle against the dramatic lack of quality of the system (in PISA 2018, 1/3 of students performed below level 2 in Reading) and the correlation between the social-economic status and differences in cognitive scores.

They present some clear examples of an undesirable reduction of inequalities that have been observed in Math and Sciences. Firstly, the authors compared performance differences in PISA for the different economic, social and cultural status of students’ parents (ESCS)⁹. These differences have been narrowed from 2006 to 2018, but at a high cost: results worsened for all levels of ESCS and reduced more rapidly for higher levels. Secondly, the authors show that gaps between genders in Math and Sciences have been reduced, also at a high cost: in Math, boys decreased their performance by 11 points, while girls improved theirs by one point only; in Sciences, boys decreased their performance by nine points, while girls improved theirs by two points only.

The chapter on Finland shows that problems exist even in the developed educational systems. Arto Ahonen discusses his country’s evolution and shows that the gender gap in reading literacy has consistently been one of the highest in the PISA

⁹ESCS is the PISA index of economic, social and cultural status, built by weighting the International Socio-Economic Index of Occupational Status (ISEI), the highest level of education of the student’s parents, converted into years of schooling, the PISA index of family wealth, the PISA index of home educational resources, and the PISA index of possessions related to “classical” culture in the family home.

participating countries. He also shows that the link between socioeconomic background and students' performance has increased since 2009. Discussing the general decline of Finish results, he shows that the phenomenon is essentially due to the "increase in the numbers of weak performers in all assessment areas", although the level of high performers also declined in Mathematics and Sciences. He also reveals that the gap between the highest and the lowest decile has widened in all areas, especially in Reading and Sciences.

To put the Finish evolution into perspective it may help to know that the country usually reviews the curriculum approximately every ten years. The last revisions went into effect in 2004 and 2016.

The case of Portugal is also interesting, as discussed in this country's chapter. Up to 2015, the nation was able to steadily increase the academic levels of those at the bottom of the scale at the same time it was developing a demanding and well-structured education. In 2018, about three years after a coalition vote in parliament abolished national exams for some school grades and the ministry started pressing for curricular flexibility and less knowledge-goal-oriented education, overall results stalled and even registered slight decreases. Simultaneously, the estimated fraction of low performers increased a bit in Sciences (2.8 pp¹⁰) and Reading (2.4 pp) and decreased slightly in Mathematics (0.5 pp). In parallel, the estimated fraction of top performers decreased in Science (1.8 pp) and oscillated very slightly in Reading (−0.2 pp) and Mathematics (+0.2 pp).

The evolution of Taiwan that Su-Wei Lin, Huey-Ing Tzou, I-Chung Lu, and Pi-Hsia Hung describe in their chapter gives us hope. Although still performing at a very high international level, Taiwanese are worried about some declines in their performance, namely in Mathematics and Science. As the authors explain, top performance helps to develop a country's talent pool. So, "increasing the proportion of top students in reading and science and maintaining Taiwanese students' high performance in mathematical literacy are critical for Taiwanese education system."

In parallel to this concern, Taiwanese have a policy of "actively assisting students with low performance". This is more than necessary given the worrisome level of low performers, namely in Reading. In order to change this reality, Taiwan is developing programs for both teaching and assessment related to literacy. In line with modern research on curriculum coherence, it is good to see teaching and assessment equally stressed.

Su-Wei Lin and her co-authors also explain that some gaps have narrowed in a desirable way. The gender gap in Reading narrowed "because male students' reading performance improved, and female students' reading performance" remained the same." Contrarily to many countries, Taiwan was also able to reduce the correlation between socioeconomic status and scores.

¹⁰Percent points.

In his chapter on England, Tim Oates reports that the gender gap is significantly lower than the OECD average, but “equity remains challenging”. There has been an increase amongst higher performing students, but the low achievers’ scores have remained unchanged. He highlights the importance of the post-2010 emphasis on reading, which is a foundational domain for students’ progress in all subjects.

7 Grade Repetition

As it happens with the false dichotomy between performance and inequality, many discussions about grade repetition stress a false dichotomy between performance and grade advancement. Simply put: Some traditional currents of thought stress the need to nudge students to attain a minimal level of performance by retaining them until they attain such minimal level, while some progressive schools of thought stress the discomfiture to students and the perpetuation of socioeconomic inequalities implied by retaining low achievers. In particular, they stress that low achievement is correlated with socioeconomic status.

Grade repetition is a calamity in many countries, as it reaches a high fraction of students. The OECD average for repetition is about 13% in primary and secondary education, but some countries display a much higher rate. France, for instance, has a retention rate of about 14% at the primary and 20% at the secondary level.

Repetition can be viewed as a measure of failure of the education system and an economic burden for the countries. In Chap. 7, Montserrat Gomendio estimates that repetition represents 8% of education expenditure, for a fraction of about 40% repetition in Spain.

Sometimes, the solution seems to be to eschew repetition, or even to ban it. In many cases this may postpone failure to a higher grade-level, and students may drag their difficulties throughout mandatory schooling until they may drop out of school altogether. At the end, school still fails these students; it just postpones failure.

Arguing against repetition, some currents of thought argue that repetition does not help students, that they do not learn more just by repeating a grade. But reality here is nuanced, and evidence is mixed. A well-known extensive meta-analysis by Chiharu Allen and co-authors (Allen et al. 2009) couldn’t find overall negative effects in retention. A more recent survey published by the OECD (Ikeda and García 2014) also reports mixed results, suggesting that postponing retention to middle-secondary school may be beneficial. Similarly, rigorous localized counterfactual studies (see e.g. Nunes et al. 2018 and Schwerdt, West, and Winters 2017) point to positive effects of retention for retained students. In particular, a very recent study with rich and detailed Florida microdata points to immediate and long-run positive effects of grade retention (Figlio and Ozek 2020). In his Chapter on Portugal, João Marôco points to a curious effect: repeaters seem to progress faster in some subjects.

The issue of grade repetition can be looked at from different perspectives.

Firstly, the problem is not only whether a low-performing student improves or not by repeating a grade. The problem is more wide ranging: will the system as a whole

improve if all students are told that repetition will not happen, no matter what level a given student attains?

Secondly, if we consider keeping repetition combined with measures to increase excellence, on one side, and no-repetition plus lenience towards students' low achievement, on the other side, are we setting up the right comparison?

In this volume, authors who discuss repetition take a balanced approach that avoids this false dichotomy: the focus on excellence should be sustained with special support for struggling students.

8 Exams and Assessment

High-stakes and low-stakes tests are also a controversial terrain. The first type of these assessment tools, i.e. exams that have consequences for students' future path are often associated with a conservative view and a ruthless selection of students that predominantly alienates those from more disadvantaged backgrounds. The second type, i.e. formative assessment tests that have no or minimal direct impact on students' path, tend to be associated with a progressive view that cares about inclusion and the progress of students.

This chapter presents a different point of view, arguing that both forms of assessment are necessary. Both monitor the education system, both provide feedback to students, teachers, schools, principals, and parents.

Low-stakes tests are valuable for giving frequent feedback to students, helping them regularly in improving their knowledge and skills. Indeed, one of the most solid results of modern cognitive psychology indicates that testing is one of the most efficient tools for improving knowledge retention and consolidation.¹¹

High-stakes tests or exams are essential to nudge students progresses, to make sure different levels of learning are attained at each step, and to increase greater transparency and efficiency of the educational system as a whole.

Recent research by one of the authors of this volume and his co-authors shows that standardized testing helps to improve countries' educational performance, particularly those testing systems that have "consequential implications". Their results "indicate that accountability systems that use standardized tests to compare outcomes across schools and students produce greater student outcomes. These systems tend to have consequential implications and produce higher student achievement than those that simply report the results of standardized tests". Consequently, "both rewards to schools and rewards to students for better outcomes result in greater student learning" [...] Most interestingly is their finding that testing and accountability are more important for low performing educational systems than for other systems (Bergbauer et al. 2019).

Almost all authors in this volume address the assessment question and it's interesting to see their approaches.

¹¹ See e.g. Roediger and Karpicke (2006) and Roediger, Smith, and Putnam (2011).

In the chapter on Estonia, Gunda Tire explains in some detail the external evaluation system established in 1997 in the country, which includes tests in grades 3 and 6, plus high-stakes exams in grades 9 and 12. She also explains that the Estonian model uses assessment to detect struggling students “early enough” and to support them “while they are with the same age group peers”. Consequently, “grade repetition is not commonly practiced”. She also stresses the fact that with this system the “poorest students in Estonia” perform “better than the top quarter with the most affluent background in many countries”.

In his chapter on Poland, Maciej Jakubowski explains how in Polish external national examinations at the end of every stage of education creates both incentives for teachers and students and social pressure and support for achieving good outcomes. He makes an interesting point by stressing that the external assessment of student outcomes and a large degree of school autonomy constitutes a good mix of freedom and external monitoring.

In the chapter on Portugal, João Marôco points out the impact of the introduction of high-stakes exams for mathematics and the Portuguese language and the PISA score improvement that followed. He also stresses the fact that the removal of high-stakes exams in grades four and six may have had detrimental consequences even on low stakes assessments like PISA.

Most stimulating is also the discussion in the chapter on Spain about repetition and assessment. Montserrat Gomendio explains that the lack of standardized testing delayed the detection of students lagging behind and coexisted with a high level of grade repetition (36% versus 13% OECD average). The author concludes that the system implemented in 1990, with its lack of reliable and uniform assessment, although “designed in theory to promote equality, led to the worst type of inequality: the expulsion of students from an education system which was blind to their performance and insensitive to their needs”.

A related point is made in the chapter on the United States. Eric Hanushek stresses that there have been large policy changes in the U.S., but they have neither led to better average outcomes nor to the consistent narrowing of achievement gaps. Many different programs intended to improve the educational system had funding that was not tied to any specific use and had no requirements to perform an impact evaluation.

9 Curriculum, Pedagogy, and Learning Outcomes

PISA 2015 reports included results that surprised many policy advisors and policy makers but pleased many cognitive scientists. Those results revealed an association between different teaching practices and outcomes in the Sciences. Unfortunately, no similar graphs were reported for PISA 2018, which has Reading as the major domain.

The first results, summarized in Fig. 4, which is taken from the OECD PISA Report, reveal some widely documented associations between performance and variables such as students’ socio-economic profile, the socio-economic profile of his or

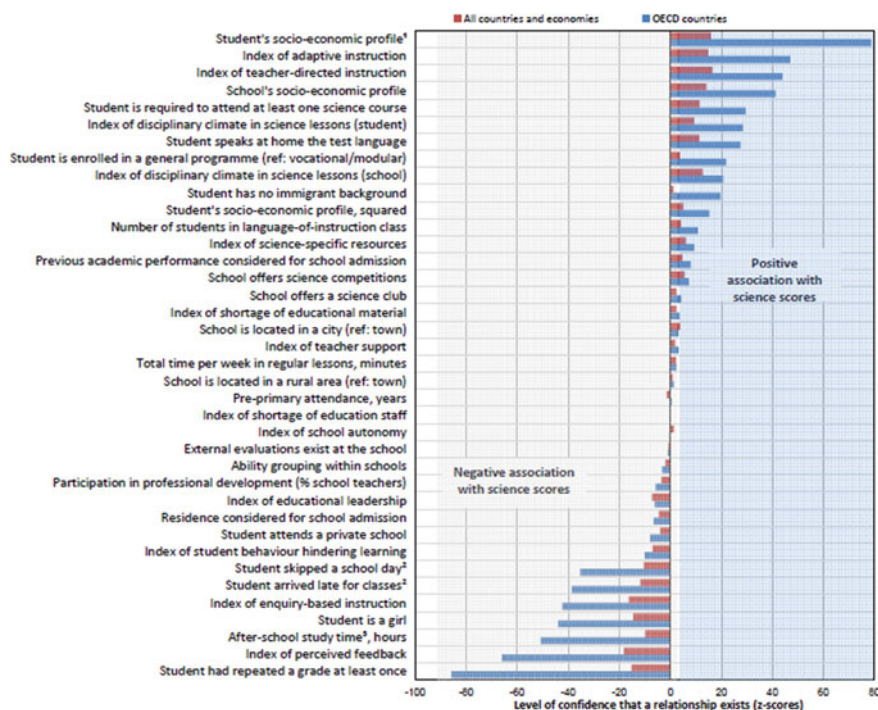


Fig. 4 Factors associated with science performance. *Notes* 1. The socio-economic profile is measured by the PISA index of economic, social and cultural status (ESCS). 2. In the two weeks prior to the PISA test. 3. Includes homework, additional instruction and private study. Factors are ranked in descending order of the z-scores for OECD countries *Source* OECD, PISA 2015 Database. Figure II.7.2 from OECD (2016b). <http://dx.doi.org/10.1787/888933436455>

her school, the language spoken at home, previous retainment, absenteeism, and gender. For these associations, there were no surprises.

However, the PISA 2015 survey introduced additional variables which are often categorized as distinguishing student-centred and teacher-centred teaching approaches. The origin of these designations and this dichotomy are unfortunate as they are deeply ideologically laden.¹² Nowadays, many educationalists chose

¹²If we go back to the origins of the classification, it would be surprisingly difficult to accept the child-centred approach, as it essentially prescribes the abandonment of curricular goals, a completely outdated and non-scientific recapitulationist theory of child development, and a radical Rousseauian view of child freedom for self-development. The main founder of this classification adopted a now completely outdated recapitulationist approach to mind's evolution (Hall 1901). According to this understanding, the child's psychological development would repeat that of the species over evolutionary time. Next, Rugg and Shumaker (1928) developed the idea that education starts and is developed to follow children's interests and development. For a modern practical critique of this recapitulationist approach see the seminal paper of Chi, Feltovich, and Glaser (1981). The concept has since evolved, but its meaning still equates with the child being the one initiating, explaining and testing his or her experiments.

to characterize this dichotomy in a pragmatic way¹³, by listing various teaching approaches as child-centred (active participation, enquire-based instruction, and the sort) or as teacher-centred (lecturing, curricular goal-oriented classes, and the sort). This characterization doesn't do justice to the original distinction and is prone to eclecticism¹⁴.

Debates on this characterization have been going on for the last two or three decades among cognitive scientists, namely experimental psychologists. Based on a long set of observations, experiments, and scientific arguments, John Anderson, John Sweller, Paul Kirschner, David Willingham and many others have made the point that structured and organized teaching is an essential first element of school success and that at different stages different approaches may be necessary¹⁵. Novices need clear directions, and students who are more advanced in a specific area benefit from autonomously setting and addressing open challenges. Student-centred versus teacher-centred is not the best framework for researching what works in education.

However, in Fig. 4, some associations provide strong support for general teacher-led learning and strong arguments against child-led learning. We verify that the index of teacher-directed instruction is positively correlated with students' outcomes in science and the index of inquired-based instruction is negatively correlated with the same index. This upsets many assumptions in contemporary discourse. It is also interesting to notice that shortage of materials and shortage of staff seem to make no difference in students' results.

Figure 5 confirms and complements some of these results. Curiously, it is more important that teachers explain how scientific concepts can be manifest in different phenomena than that teachers explain the relevance of scientific concepts for peoples' lives.

This seems counterintuitive, but is a very powerful argument in favour of knowledge—even in favour of pure knowledge. Research has shown that trying to boost student motivation to raise attainment through demonstrating the usefulness of knowledge does not necessarily favour learning. It is knowledge that leads to knowledge curiosity.¹⁶

This same Figure shows that teachers' explanations support good results while students' design of their own experiments, investigations, and class debates hamper good results.

¹³See e.g. Chall 2002, pp. 187-192.

¹⁴If we go back to the original definition, no one or almost no one nowadays will defend a radical child-centred approach. But if we follow today's eclectic and pragmatic classification, no one or almost no one will defend a fully teacher-centred approach—it doesn't sound virtuous, although in its original formulation it is a coherent philosophical stance. To make matters worse, many times teacher-centred approaches are associated with a conservative point of view and student-centred approaches with a progressive approach. When the discussion takes this non-scientific, non-technical, and ideological tone of ideas, we are bound for disaster.

¹⁵See e.g. Kirschner, Sweller, and Clark (2006), Willingham (2010), Boxer (2019), or Dehaene (2020).

¹⁶See e.g. Kirschner and Hendrick (2020), Chaps. 8 and 29.

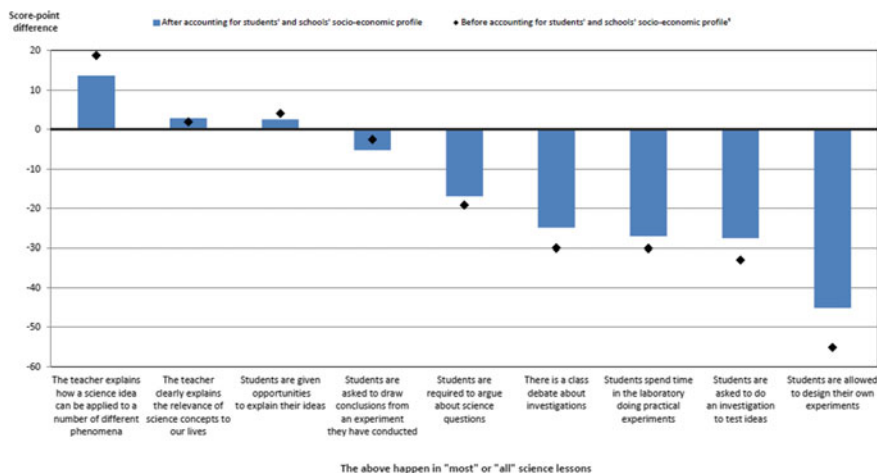


Fig. 5 Enquiry-based teaching practices and science performance. The socio-economic profile is measured by the PISA index of economic, social and cultural status. All differences are statistically significant. Source OECD, PISA 2015 Database, Table II.2.28. Figure II.2.20 from OECD (2016b). <http://dx.doi.org/10.1787/888933435628>

This is surprising on all accounts. Supporters of the so-called enquiry-based teaching cannot accept these statistics (e.g. Sjöberg 2018). By the same token, supporters of psychology research-based methods of direct instruction do not reject the importance of student hands-on experimentation and student active answer-seeking activities. A personal conjecture is as follows: Teacher explanation is associated with confident teaching and with teachers' training and quality. Predominance of students' free investigations is associated with unorganized teaching and teacher's lack of coherent and confident content knowledge. It is not necessarily so. But these are statistical results.

Regrettably, we do not have similar statistics on PISA 2018. Nevertheless, it is important to know what type of teaching approach is predominant in each country and how our experts assess their influence on each country's results.

Most authors in this book assume a pragmatic approach. It is very rewarding to notice in the chapter on Estonia the importance of its national curriculum and its reform in 1996, which stressed not only a "detailed description of what teachers should teach in their subjects", but a new focus on "what students should know and be able to do". It is a curriculum focused on "learning outcomes". It describes "knowledge, skills, attitudes and values". This cannot be too stressed: a curriculum that is comprehensive but starts with knowledge.

One year after establishing the new curriculum, Estonia established a new external evaluation system. Then, in 2014, it established a new strategy for extending learning skills, taking care of vocational skills, and training teachers.

Discussing teaching styles, Gunda Tire recognizes that Estonian teachers use less frequently student-centred approaches compared with teachers in other OECD

countries. But recognizes “a subtle balance between tradition and innovation”. This balance has been serving Estonia well.

What is then the secret of Estonian success? Gunda Tire stresses the idea that “commitment to education”, plus a “very demanding curriculum” and “high quality examinations built directly on the curriculum” are key ingredients.

Writing about Poland, Maciej Jakubowski stresses the importance of curricular changes for his country. Describing the new curriculum set in 2008, he highlights the curricular “learning outcomes” and the need to have “detailed requirements describing the specific knowledge and skills to be mastered by students”. Next, he highlights the essential role of “central assessments”.

Jakubowski also points out that some so-called “innovative teaching methods are disputable”, namely some recommendations for “twenty-first century skills”. And concludes praising a “good balance between innovations and traditional teaching”.

Tim Oates goes one step further and claims that the strong emphases away from rote learning has harmed students. He argues that some memorization is necessary, not as an end in itself but in enabling knowledge to be retained in long term memory and therefore immediately available for higher level and complex problem solving.

Most interesting is Oates reference to the curriculum as a crucial point of reference. He argues for “curricular coherence”¹⁷, where instruction, assessment, standards and materials are carefully and deliberately aligned. This provides a starting point for standards, schools and teachers’ accountability, professional, practice, institutional development and all subsequent aspects of the educational system.

All this fails if teachers are not able to deliver a good quality education to their students. Teacher initial training, selection, professional development, and promotion are essential aspects of school systems. Although this topic is not systematically discussed in this volume, it is worthwhile to mention that the quality of teachers’ initial training in Finland referred to by Arto Ahonen in this country’s chapter is usually singled out as one of the crucial explanations for Finnish successes.

Teacher quality and teacher experience are also discussed in the chapter on Chile, where Ema Lagos and Vitoria Martínez explain that experienced teachers are not uniformly distributed in the country: the proportion of teachers with less than five years of experience is much higher in disadvantaged schools. Sue Tomson has detailed data on teachers and reveals a worrisome situation: in Australia, disadvantaged schools have a much higher proportion of poorly qualified teaching staff, teacher absenteeism, and ill-prepared teachers, than advantaged schools.

10 Knowledge Versus Competencies

No word in education is more ambiguous than the word “competencies”. In PISA reports, it is usually just a convenient word for a mixture of knowledge, skills, attitudes, values, and capacity for solving applied problems. In some education literature,

¹⁷See Schmidt et al. (2001).

though, competencies are considered as *the* main education goal and not a global designation for equally important education goals.

According to this view, what matters is the mobilization of the above-referred four cognitive and social components to solve practical problems and to be productive in life. This mobilization is then called a competency and knowledge disappears as the starting point in the curriculum. Going one step further, some argue that the focus on knowledge may harm the ability to cooperate, develop critical thinking, and be able to be productive in society. The curriculum focus should then be the application of knowledge.

Although many times introduced as a novel twenty-first century approach, this view is essentially a modern development of some nineteenth-century utilitarian views of Herbert Spencer (1820–1903) and others,¹⁸ and an importation into education of the concept of competencies advanced in the business literature during the last quarter of the twentieth century¹⁹.

Nowadays, everybody recognizes that students need to go deeper than rote memorization and simple understanding of curricular subjects. Schools pay increased attention to the application of knowledge, to the ability to apply abstract concepts to solve real-life problems, to develop the capability to relate matters and concepts, to be active in formulating learning questions, and to transfer knowledge to new contexts. So, the question is not whether the application of knowledge is important, but whether the application is the only goal and whether there is no value in knowledge itself.

The paradox is that some countries that have embraced competencies as the unifying concept of the curriculum face challenges in the education of their students. Other countries that used to follow a strict curriculum got worse results after redesigning their curriculum around competencies. And other nations, namely Asian, that have developed and followed a very organized and strict knowledge-based and sequential curriculum are obtaining excellent results in the evaluation of student competencies as measured by PISA questions.

Modern cognitive science comes to our rescue in the interpretation of these apparent paradoxes. Firstly, skills are essentially domain-based. To try to develop general transferable skills with no roots in basic subject training, in memory activation, and in curricular knowledge is a vain goal. Secondly, training in interpretation, generalization, and application is a valuable goal, but basic knowledge and skills are the essential tools for interpretation, for generalization, and for application²⁰.

PISA results in 2015 also come to our rescue. As we have seen in discussing Figs. 4 and 5 data, direct teaching is important to obtain results in science application questions, such as those included in the PISA surveys.

In summary, if we want our students to be proficient in knowledge application, we need to be very careful, not so much with applications as with basic knowledge.

¹⁸Spencer (1860).

¹⁹See Chouhan and Srivastava (2014), for a review.

²⁰See e.g. Willingham (2019) and the references therein.

11 Ten Conclusions from Reflecting Upon Ten Countries' Experiences

In sum, what makes countries improve their PISA scores? We will ask an apparently identical question, but a much more important one: What makes countries improve their students' knowledge and skills?

The analyses in the following chapters are very rich. Countries are diverse in their situations and histories, and authors have different points of view. By the same token, needs are unlike and proposals are varied. Any synthesis is somehow arbitrary and personal. It cannot give justice to the diversity of points of view and the wealth of proposals.

With all these caveats and the disclaimer that what follows does not intend to reproduce any agreement among the contributors to this volume, one can list the following major points.

First, everything starts with the curriculum. This is the education founding document²¹. It can be national, federal, regional, or established at local levels. It can be more detailed or less specific, it can be later translated in standards or contain them, but without clear learning goals no education system can progress.

Second, the curriculum, or curricular structure if it is made from different pieces, ought to be ambitious, demanding, and set clear objectives. These objectives must be sequenced, setting solid foundations for students' progress. Knowledge is a necessary foundation to develop skills and values.

Third, everything needs to be coherent around curricular goals. It does not make sense that assessment instruments evaluate some learning goals, textbooks stress others, and schools are rewarded for attaining still different student goals.

Fourth, we need to simultaneously nurture quality and improve low performing students' achievement. To increase average results but allow a significant fraction of students to remain insufficiently prepared for progressing in school and life cannot be a virtuous goal. Similarly, to reduce disparities and to lower everybody's results cannot be a virtuous goal. In sum: a demanding system is not incompatible with caring for low performing students.

Fifth, pedagogy matters. We need a good balance between innovating with new pedagogical approaches and new technology and paying attention to proven basic methods. It is as detrimental to insist on utopic messages that forget basic steps of learning as to insist in maintaining a conformist version of students progresses and not improve ourselves as educators. Students are not little experts that will discover all this brave world by themselves, but can become experts if guided through the necessary intermediate steps.

Sixth, assessment is crucial. PISA and other ILSA tools are important, but an educational system can only progress if it introduces frequent and reliable formative and summative assessment, if student learning goals are verified, if a good independent testing system is in place.

²¹ See e.g. Crato (2019).

Seventh, teachers are the essential mediators and agents of a school system. If their initial training is weak, this hindrance is not likely to be remedied by on-the-job training. The whole process of teachers' initial training, hiring selection, professional development, and promotion is a very serious matter that few countries have managed to address successfully.

Eighth, inform and involve the public. The countries that report a positive effect from participating in PISA and having external evaluations are those that managed to have informed participation from society, which allowed public pressure and public support for improvement.

Ninth, we need to pay attention to what is essential. And the essential is the progress of students, starting with their cognitive development, but including their skills, attitudes and overall development. As the froth of political discussions, professional interests and daily news may diverge to many topics, when reflecting upon education there is one goal above all others: students' progress.

Tenth, education policies need to be judged by students results, rather than by policies' intentions.

Acknowledgements This work and its publication were supported through Project CEMAPREUID/MULTI/00491/2019 financed by FCT/MCTES through national funds.

References

- Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multilevel analysis. *Educational Evaluation and Policy Analysis*, 31(4), 480–499. <https://doi.org/10.3102/0162373709352239>.
- Araujo, L., Saltelli, A., & Schnepf, S. V. (2017). Do PISA data justify PISA-based education policy?. *International Journal of Comparative Education and Development*, February 2017, <https://doi.org/10.1108/ijced-12-2016-0023>.
- Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2019). NBER Working Paper No. 24836.
- Boxer, Adam (Ed.). (2019). *Explicit & direct instruction*. Woodbridge, UK: John Catt.
- Chall, Jeanne S. (2002). *The academic achievement challenge: What really works in the classroom?*. New York: Guilford.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152. https://doi.org/10.1207/s15516709cog0502_2.
- Chouhan, Vikram Singh, & Srivastava, Sandeep. (2014). Understanding competencies and competency modeling—A literature survey. *IOSR Journal of Business and Management*, 16, 14–22.
- Crato, N. (2019). Everything starts with the curriculum, *ResearchED* 3. Available at <https://researched.org.uk/everything-starts-with-the-curriculum/> [accessed 11 April 2020].
- Crato, N. (2020). Curriculum and educational reforms in Portugal: An analysis on why and how students' knowledge and skills improved. In F. M. Reimers (Ed.), *Audacious education purposes: How governments transform the goals of education systems*. Berlin: Springer.
- Dehaene, S. (2020). *How We Learn*, Viking—Penguin Random House (adapted translation of *Apprendre*, Paris: Odile Jacob, 2018).
- European Council (2009). Council conclusions of 12 May 2009 on a strategic framework for European cooperation in education and training ("ET 2020"), Office for Official Publications

- of the European Communities, http://dx.doi.org/10.3000/17252423.C_2009.119.eng [accessed 20 February 2020].
- Figlio, D., & Özek, U. (2020). An extra year to learn english? Early grade retention and the human capital development of English learners. *Journal of Public Economics* 186, 1–16. <https://doi.org/10.1016/j.jpubeco.2020.104184> [accessed 1 August 2020].
- Goldstein, Harvey (2017). Measurement and evaluation issues with PISA. In *The PISA effect on global educational governance*, ed. Louis Volante, New York and London, Routledge.
- Hall, G. Stanley. (1901). The ideal school as based on child study. *Forum*, 32, 24–39.
- Hopfenbeck, Therese N., Lenkeit, Jenny, El Masri, Yasmine, Cantrell, Kate, Ryan, Jeanne, & Baird, Jo-Anne. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), 333–353. <https://doi.org/10.1080/00313831.2016.1258726>.
- IEA. (2018). 60 Years of IEA 1958–2018, Amsterdam. Available at <https://www.iea.nl/publications/60-years-iea-1958-2018> [accessed 20 February 2020].
- Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies*, 2013(1). http://dx.doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx.
- Kirschner, Paul A., & Hendrick, Carl. (2020). *How learning happens: Seminal works on educational psychology and what they mean in practice*. Oxon and New York: Routledge.
- Kirschner, Paul A., Sweller, John, & Clark, Richard E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist. *Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching, Educational Psychologist*, 14, 75–86. https://doi.org/10.1207/s15326985ep4102_1.
- Nunes, Luís Catela, Reis, Ana Balcão, & Seabra, e Carmo. (2018). Is retention beneficial to low-achieving students? *Evidence from Portugal, Applied Economics*, 50(40), 4306–4317.
- OECD (2016a), PISA 2015 results (volume I): Excellence and equity in education, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264266490-en>.
- OECD (2016b), PISA 2015 Results (Volume II): Policies and practices for successful schools, PISA, OECD Publishing, Paris. <https://doi.org/10.1787/9789264267510-en>.
- OECD (2019a). PISA 2018 results (volume I): What students know and can do, PISA, OECD Publishing, Paris. <https://doi.org/10.1787/5f07c754-en>.
- OECD (2019b). PISA 2018 results (Volume II): Where all students can succeed, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/b5fd1b8f-en>.
- OECD (2019c). PISA 2018 results—Combined executive summaries, Volume I, II & III, https://www.oecd.org/pisa/Combined_Executive_Summaries_PISA_2018.pdf [accessed February 10, 2020].
- OECD (2019d). PISA 2018 assessment and analytical framework. Paris: PISA, OECD Publishing. <https://doi.org/10.1787/b25efab8-en>.
- Parker, P. D. et al. (2018). Inequity and excellence in academic performance: Evidence from 27 Countries. *American Educational Research Journal Month XXXX*, XX(X), 1–23, <https://10.3102/0002831218760213>.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Roediger, H. L., Smith, M. A., & Putnam, A. L. (2011). Ten benefits of testing and their applications to educational practice. In B. H. Ross (Ed.), *Psychology of learning and motivation*. San Diego: Elsevier Academic Press.
- Rugg, H. O., & Shumaker, A. (1928). *The child centered school*. New York: World Book Company.
- Schmidt, William H., et al. (2001). *Why school matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey Bass, Wiley.
- Schwerdt, G., West, M. R. & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida, *Journal of Public Economics*, 152(C), 154–169.

- Sjøberg, Svein. (2018). The power and paradoxes of PISA: Should inquiry-based science education be sacrificed to climb on the rankings? *NorDiNa*, 14(2), 186–202.
- Spencer, H. (1860). *Education: Intellectual, moral, and physical*. New York: Appleton. Reprint in 1911 as *Four essays on education and kindred subjects*. London: Dent.
- United Nations Statistics Division (2019). Global indicator framework for the Sustainable Development Goals and targets of the 2030. Agenda for Sustainable Development, <https://unstats.un.org/sdgs/indicators/indicators-list/> (accessed on 13 June 2020).
- Willingham, Daniel T. (2010). *Why don't students like school?: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco: Jossey-Bass.
- Willingham, D. T. (2019). How to teach critical thinking, occasional paper series, New South Wales Department of Education.
- Zhao, Y. (2020). Two decades of havoc: A synthesis of criticism against PISA. *Journal of Educational Change*. <https://doi.org/10.1007/s10833-019-09367-x>.

Nuno Crato is Professor of Mathematics and Statistics at the Lisbon School of Economics & Management of the University of Lisbon. Visiting senior research scientist at the European Commission Joint Research Center, his academic research focus on stochastic models, time series applications on financial and social problems, and statistical data-based evaluation of policy measures, namely education. He has published extensively on econometrics and statistics and worked on education policy analysis.

From 2011 to 2015, he was the Education and Science Minister of Portugal. During his tenure, mandatory schooling was raised from nine to 12 grade years, English was introduced as a mandatory subject starting at third grade, the dropout rate was reduced from c. 25% to 13.7%, retention rates improved, and Portuguese students achieved the best results ever in international surveys, namely PISA and TIMSS.

He was President of the Portuguese Mathematical Society (2004–2010) and director of the International Institute of Forecasters (2016–2020).

A prolific science writer, some of his books have been translated and published in various countries, including the U.S., U.K., Italy, and Brazil, namely his *Figuring It Out* (Springer 2010). His recently co-organized *Data-Driven Policy Impact Evaluation* (Springer 2019) is becoming an international reference in the field.

He has been a vocal voice in educational debates, publishing critical articles and books on education, advocating a structured curriculum, external evaluation of students, and a better content-knowledge training of teachers. He is founding organizer of the *Lisbon Economics and Statistics of Education* conference series and directs *Iniciativa Educação*, a program for fostering education.

For his work, he has received prizes from the European Mathematical Society (2003) and the European Union (2007). He was awarded Commander (2008) and Great-Cross (2015) of Prince Henry Order of the Portuguese Republic.

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

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

