

Nuno Crato *Editor*

Improving a Country's Education

PISA 2018 Results in 10 Countries

OPEN ACCESS

 Springer

Improving a Country's Education


Nuno Crato
Editor

Improving a Country's Education

PISA 2018 Results in 10 Countries

 Springer

Editor

Nuno Crato 

Mathematics and Statistics, ISEG

University of Lisbon

Lisbon, Portugal



ISBN 978-3-030-59030-7

ISBN 978-3-030-59031-4 (eBook)

<https://doi.org/10.1007/978-3-030-59031-4>

© The Editor(s) (if applicable) and The Author(s) 2021. This book is an open access publication.

Open Access This book 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 book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book'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.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is probably one of the first to be published, or even the first, about the results of the Programme for International Student Assessment (PISA) 2018. It discusses how PISA results in ten different countries have evolved and what makes countries change. Information on each country's educational system contextualizes the discussion about PISA and other Large-Scale International Assessments' results, such as TIMSS, Trends in International Mathematics and Science Studies.

One reason only made it possible for us to present this work to the reader with such a short delay after PISA results were published in December 2019: we were very fortunate to be able to gather an exceptionally knowledgeable and generous group of international experts.

The ten countries discussed in this volume represent a wide variety of educational systems, from Australia and Taiwan, in the East, to England, Estonia, Finland, Poland, Portugal and Spain, in Europe, and to Chile and the USA, in the Americas. We have high-performing countries, countries that are around the OECD average, and countries that are struggling to attain the OECD average. Each country has its history that reflects efforts to improve educational achievement.

After the introduction, each chapter of this book concentrates on one country. Countries are presented by alphabetic order. Each one is discussed by one of its foremost national experts, some of them with experience in government or in advising governments, many of them with experience in international organizations and quite a few served as national representatives for international assessments. If the reader peruses the biographic notes of each contributor, I'm sure he or she will be as pleased as I was honored when all of them accepted my invitation to contribute.

The idea for this book came about when I had the privilege of convening a roundtable on TIMSS and PISA results at LESE, the Lisbon Economics and Statistics of Education meeting in January 2019. It took place at the Lisbon Economics and Business School of the University of Lisbon, ISEG, where I work. It was the fifth meeting of this biennial conference, and five authors of this book were present. We immediately felt that the diversity of experiences and the independence of spirit of the participants enriched tremendously the analyses presented

for individual countries. We had the idea of preparing a contribution that could help interpret PISA 2018 results and started preparing our work even before the results were released. The outcome is this collective work.

The book is organized as follows. Each chapter is a data-based essay about the evolution of a specific country, discussed and supported by PISA results and other data, and represents the personal stance of the authors. Thus, each author represents his or her own views and not those from his or her institution or government. Each author draws on published data, as well as on a vast set of information and supports his or her view with data and reliable information.

The introductory chapter gathers my reading of the ten chapters. It follows the same principles: I express my views freely, but support them with the best information available. I do not claim to voice the opinion of the authors, and I am the sole responsible for what I wrote.

A final chapter introduced following a Springer referee suggestion provides the necessary background in order to understand what PISA measures and how. It shows examples of PISA and TIMSS questions that convey a better idea on what the results of these surveys mean about students' knowledge and skills.

I am honored to edit this book, and I am sure it will be useful to all those interested in understanding what it takes to improve a country's education system.

Lisbon, Portugal
April 2020

Nuno Crato

Acknowledgements

I feel very grateful to the LESE group and to my research centre Cemapre/REM, at ISEG, for their continuous help and support in the publication of this volume.

We are all grateful to the reviewers who helped improve this collective work. Needless to say, they are not accountable for any insufficiencies or views expressed in this book.

Two anonymous reviewers invited by Springer publishers were very meticulous and particularly helpful for the coherence of the full book and for improving the chapters. We are also very grateful to the following invited reviewers for their expertise and care.

Luisa Araújo, Instituto Superior de Educação e Ciências, Lisbon, Portugal
Jennifer Buckingham, Director of Strategy and Senior Research Fellow, MultiLit, Australia

Patrícia Costa, European Commission Joint Research Centre, Ispra, Italy
Montserrat Gomendio, Spanish Research Council, Spain

Ralph Hippe, European Commission Joint Research Center, Seville, Spain
Isabel Hormigo, Sociedade Portuguesa de Matemática, Lisbon, Portugal

Maciej Jakubowski, Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

João Maroco, Instituto Superior de Psicologia Aplicada, Lisbon, Portugal

Gabriel H. Sahlgren, Research Institute of Industrial Economics and London School of Economics

Mónica Vieira, Iniciativa Educação, Lisbon

Contents

Setting up the Scene: Lessons Learned from PISA 2018 Statistics and Other International Student Assessments	1
Nuno Crato	
Australia: PISA Australia—Excellence and Equity?	25
Sue Thomson	
Chile: The Challenge of Providing Relevant Information from ILSA Studies for the Improvement of Educational Quality	49
Ema Lagos	
England: England and PISA—The Long View	83
Tim Oates	
Estonia: A Positive PISA Experience	101
Gunda Tire	
Finland: Success Through Equity—The Trajectories in PISA Performance	121
Arto K. Ahonen	
Poland: Polish Education Reforms and Evidence from International Assessments	137
Maciej Jakubowski	
Portugal: The PISA Effects on Education	159
João Marôco	
Spain: The Evidence Provided by International Large-Scale Assessments About the Spanish Education System: Why Nobody Listens Despite All the Noise	175
Montse Gomendio	

Taiwan: Performance in the Programme for International Student Assessment 203
Su-Wei Lin, Huey-Ing Tzou, I-Chung Lu, and Pi-Hsia Hung

United States: The Uphill Schools’ Struggle 227
Eric A. Hanushek

Assessment Background: What PISA Measures and How 249
Luisa Araújo, Patrícia Costa, and Nuno Crato

Contributors

Arto K. Ahonen Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland

Luisa Araújo Instituto Superior de Educação e Ciências, ISEC, Lisbon, Portugal

Patrícia Costa European Commission Joint Research Centre, Ispra, Italy

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

Montse Gomendio Spanish Research Council, Madrid, Spain

Eric A. Hanushek Hoover Institution, Stanford University, Stanford, USA

Pi-Hsia Hung Department of Education, National University of Tainan, Tainan City, Taiwan

Maciej Jakubowski Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

Emilia Lagos PISA National Coordinator, National Agency for Educational Quality, Santiago, Chile

Su-Wei Lin Department of Education, National University of Tainan, Tainan City, Taiwan

I-Chung Lu Department of Education, National Pingtung University, Pingtung City, Taiwan

João Marôco ISPA—Instituto Universitário, Lisboa, Portugal

Tim Oates Director Assessment, Research and Development, CBE, Cambridge, England

Sue Thomson Australian Council for Educational Research, ACER, Camberwell, VIC, Australia

Gunda Tire Education and Youth Authority, Tallinn, Estonia

Huey-Ing Tzou Department of Education, National University of Tainan, Tainan City, Taiwan

Acronyms

ACARA	Australian Curriculum, Assessment, and Reporting Authority
AITSL	Australian Institute for Teaching and School Leadership
BCN	Library of the National Congress of Chile
CIVED	Civic Education Study (Chile)
COAG	Council of Australian Governments
ERCE	Regional Comparative and Explanatory Study (Chile)
ESCS	PISA Economic Social and Cultural Status Index
ESSA	Every Child Succeeds Act, the federal program for school accountability from 2015 to present (USA)
EU	European Union
FONIDE	Fund for Research and Development in Education (Chile)
GIS	Geographic Information System
Head Start	Federal program to provide early childhood education for 3- and 4-year-old disadvantaged children (USA)
IALS	International Adult Literacy Survey
ICCS	International Civic and Citizenship Education Study
ICILS	International Computer and Information Literacy Study
IEA	International Association for the Evaluation of Educational Achievement
IELS	International Early Learning and Child Well-being Study
ILSA	International Large-scale Students Assessment
INEE	Instituto Nacional de Evaluación Educativa (Spain)
IRT	Item Response Theory
LESE	Lisbon Economics and Statistics of Education conferences
LGE	General Education Law (Chile)
LOGSE	Ley Orgánica de Ordenación General del Sistema Educativo (Spain)
LOMCE	Ley Orgánica para la Mejora de la Calidad Educativa (Spain)
MCEETYA	Ministerial Council on Education, Employment, Training and Youth Affairs (Australia)

NAEP	National Assessment of Educational Progress, a regular set of tests for representative samples of U.S. students conducted by the U.S. Department of Education (USA)
NAPLAN	National Assessment Program Literacy and Numeracy
NCLB	No Child Left Behind, the federal program for school accountability from 2002 to 2015 (USA)
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
QAS	Quality Assurance System (Chile)
SES	Socioeconomic background
SIMCE	National Learning Outcomes Evaluation System (Chile)
SRS	Schooling Resource Standard
TALIS	Teaching and Learning International Survey
TIMSS	Trends in International Mathematics and Science Study
Title 1	Federal program to find extra funding for disadvantaged students (USA)
UEG	Gender Unit of Ministry of Education
UNESCO	United Nations Educational, Scientific and Cultural Organization

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 composition (1984). However, the two most successful waves of international studies this Association organized were the TIMSS—with its acronym which could stand

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

N. Crato (✉)
Cemapre/REM, ISEG, University of Lisbon, Lisbon, Portugal
e-mail: ncrato@iseg.ulisboa.pt; ncrato@gmail.com

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

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

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 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.

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).

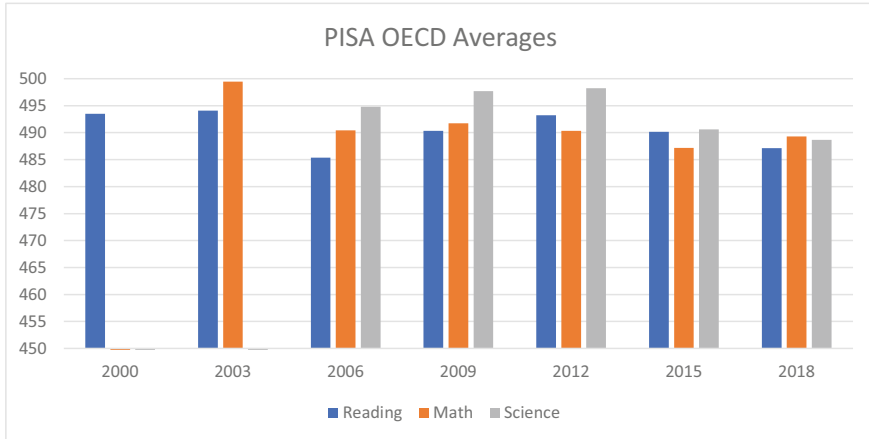


Fig. 1 Evolution of PISA Results for OECD Countries. PISA OECD countries averages include countries that have participate in all PISA waves. *Source* OECD IDE reports with recomputed updated data <https://nces.ed.gov/surveys/pisa/idepisa/report.aspx>

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.

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

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

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, María Victoria Martínez, 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 and María Victoria Martínez explain 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.

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

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).

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.

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.

All this means that a nuanced approach should be adopted as we discuss education spending. As Ema Lagos and María Victoria Martínez explain in their chapter on

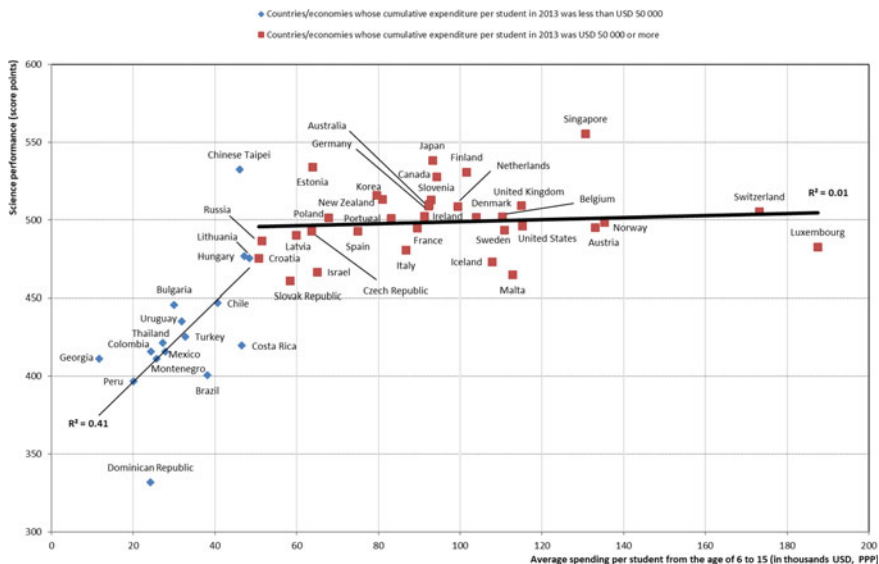


Fig. 2 Student scores in Sciences and spending in education. Only countries and economies with available data are shown. A significant relationship ($p < 0.10$) is shown by the thin line. A non-significant relationship ($p > 0.10$) is shown by the thick line. Source OECD (2016b), Figure II.6.2, p. 186;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

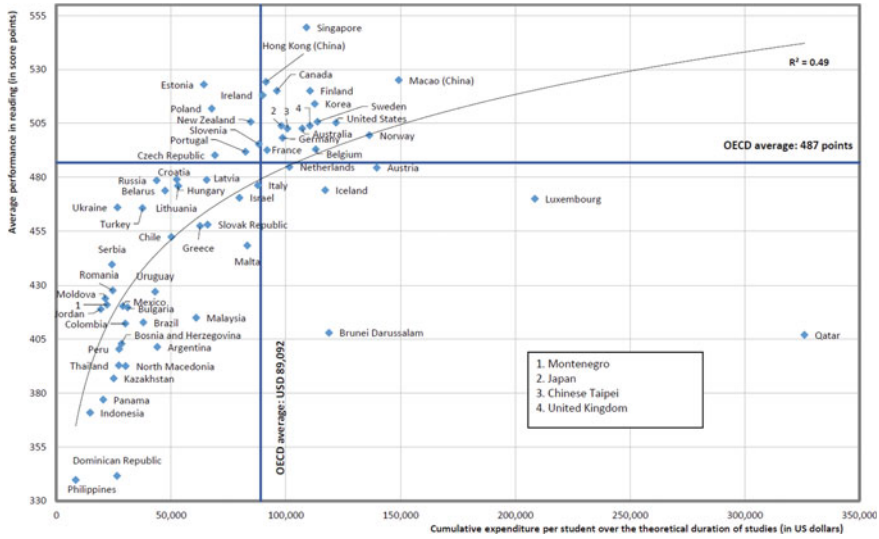


Fig. 3 Student scores in Reading and spending on education *Source* OECD (2019a), Figure I.4.3, p. 65; PISA 2018 Database, Tables I.B1.4 and B3.1.1. <https://doi.org/10.1787/888934028406>

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 Thompson 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

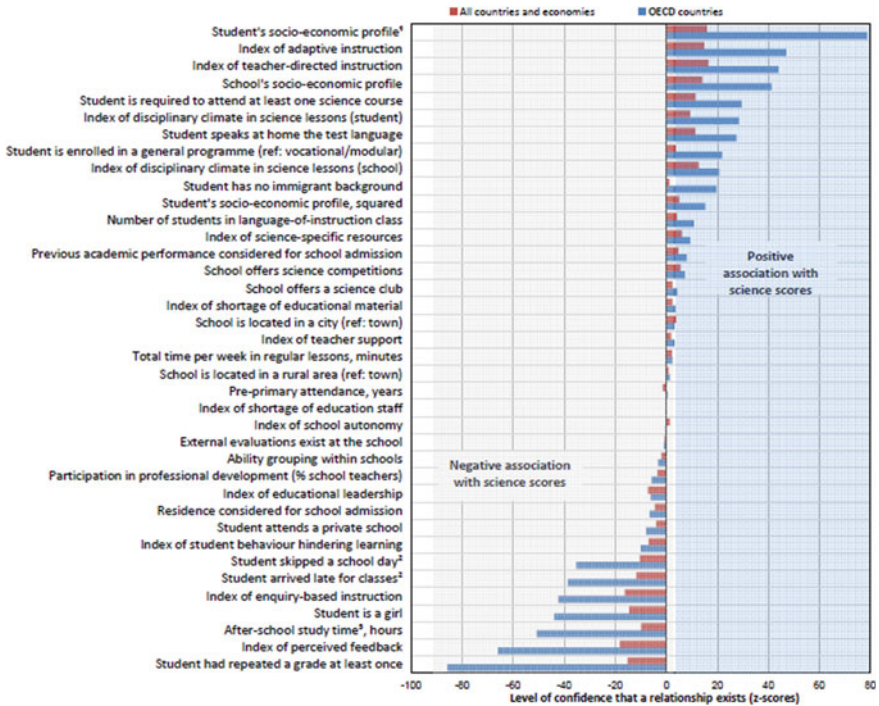


Fig. 4 Factors associated with science performance. *Notes* 1. The socio-economic profile is measured by the PISA index of economic, social and cultural status (ESCS). 2. In the two weeks prior to the PISA test. 3. Includes homework, additional instruction and private study. Factors are ranked in descending 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.

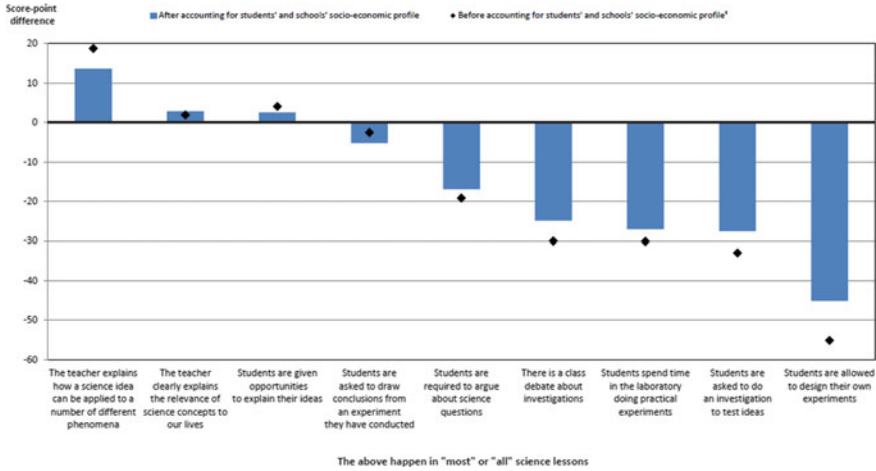


Fig. 5 Enquiry-based teaching practices and science performance. The socio-economic profile is measured by the PISA index of economic, social and cultural status. All differences are statistically significant *Source* OECD, PISA 2015 Database, Table II.2.28. Figure II.2.20 from OECD (2016b). <http://dx.doi.org/10.1787/888933435628>

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.

²¹See 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.jpube.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.



Australia: PISA Australia—Excellence and Equity?



Sue Thomson

Abstract Australia’s education system reflects its history of federalism. State and territory governments are responsible for administering education within their jurisdiction and across the sector comprising government (public), Catholic systemic and other independent schooling systems. They collaborate on education policy with the federal government. Over the past two decades the federal government has taken a greater role in funding across the education sector, and as a result of this involvement and the priorities of federal governments of the day, Australia now has one of the highest rates of non-government schooling in the OECD. Funding equity across the sectors has become a prominent issue. Concerns have been compounded by evidence of declining student performance since Australia’s initial participation in PISA in 2000, and the increasing gap between our high achievers and low achievers. This chapter explores Australia’s PISA 2018 results and what they reveal about the impact of socioeconomic level on student achievement. It also considers the role of school funding and the need to direct support to those schools that are attempting to educate the greater proportion of an increasingly diverse student population including students facing multiple layers of disadvantage.

1 The Australian Education System and Goals for Education

Australia does not have a single national education system; its individual states and territories are responsible for their own education administration, although overall the structures are similar throughout the country. Policy collaboration between state and federal governments takes place in joint councils that include federal, state, and territorial government representatives. While most children attend government (or

S. Thomson (✉)

Australian Council for Educational Research, ACER, 19 Prospect Hill Rd, Camberwell, VIC 3124, Australia

e-mail: sue.thomson@acer.org

© The Author(s) 2021

N. Crato (ed.), *Improving a Country’s Education*,
https://doi.org/10.1007/978-3-030-59031-4_2

public) schools,¹ approximately one-third attend non-government schools, in a sector comprising Catholic systemic schools and other independent schools.

State education departments recruit and appoint teachers to government schools, supply buildings, equipment, and materials, and provide limited discretionary funding for use by schools. In most jurisdictions, regional offices and schools have responsibility for administration and staffing, although the extent of responsibility varies across jurisdictions. Central authorities specify the curriculum and standards framework, but schools have autonomy in deciding curriculum details, textbooks, and teaching methodology, particularly at the primary and lower secondary levels. State authorities specify curriculum for Grades 11 and 12 and are responsible for examining and certifying final year student achievement for both government and non-government schools.

In the last two decades, in particular, the degree of involvement of the federal government and the degree of collaboration between state and territorial governments has increased. In 1989, the first declaration by joint federal and state education ministers arguing for nationally agreed goals of schooling national was released (the Hobart Declaration) (Australian Education Council 1989a, b). This was revised in 1999 and released as the Adelaide Declaration on National Goals for Schooling in the Twenty-First Century (Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA) 1999). For the first time, one of the goals placed a value on equity: “Schooling should be socially just, so that: students’ outcomes from schooling are free from the effects of negative forms of discrimination based on sex, language, culture and ethnicity, religion or disability; and of differences arising from students’ socio-economic background or geographic location.”

In 2008, ministers of education agreed to the Melbourne Declaration on the Educational Goals for Young Australians (MCEETYA 2008), which outlined revised directions and aspirations for Australian schooling. The Melbourne Declaration elevated equity and excellence to the primary goal: “*Australian schooling promotes equity and excellence*”. In addition, it spelt out that “... all Australian governments and all school sectors must ... ensure that the learning outcomes of Indigenous students improve to match those of other students ... [and] ensure that socioeconomic disadvantage ceases to be a significant determinant of educational outcomes” (p. 7).

Since then, Australia’s national reform agenda has included the development of a national curriculum, and introduction of national standards for teachers and school leaders. Two national agencies—the Australian Curriculum, Assessment, and Reporting Authority (ACARA) and the Australian Institute for Teaching and School Leadership (AITSL)—were established to support these initiatives. The Australian

¹Government schools are owned and operated by state or territory governments. They are almost entirely funded by taxes and nominally free for students to attend, though schools frequently charge for other expenses. Catholic schools are owned by the Catholic Church in Australia and the state Catholic education offices distribute funding and provide support to the Catholic dioceses in their state, which own and operate the schools. They receive funding from federal and state governments and charge fees. Independent schools are non-government schools that are run by a variety of private non-profit organisations, although the vast majority are governed by religious bodies. They receive funding from federal and state governments and charge fees.

Government’s National Assessment Program was established and includes PISA as one of several international assessments used as key performance measures for collecting data on the progress of Australian students toward the goals of the Melbourne Declaration. In 2013, the Australian Education Act was passed, which contained a broad range of national targets to ensure that Australia “provides a high quality and highly equitable system for all students”, and “for Australia to be placed, by 2025, in the top 5 highest performing countries based on the performance of school students in reading, mathematics and science” (Australian Government 2013, p. 3).

In the week following the release of the PISA 2018 results, serendipitously, the federal and state education ministers met in Alice Springs, in the Northern Territory, to discuss and agree on a revised statement of national goals. This new statement, the Alice Springs (Mparntwe) Education Declaration has, again, as its primary goal, “The Australian education system promotes excellence and equity”, and commits that “... the education community works to ‘close the gap’ for young Aboriginal and Torres Strait Islander students” (p. 16) and “governments and the education community must improve outcomes for educationally disadvantaged young Australians ... such as those from low socioeconomic backgrounds, those from regional, rural and remote areas ...” (Council of Australian Governments Education Council 2019, p. 17).

2 Funding

To fully explain the methods and history of funding education in Australian schools would require a chapter on its own. In most OECD countries, non-government schools get little or no money from government funding—they are, after all, privately owned and operated. In Australia, the story is convoluted and complicated, and goes back to our origins as a British penal colony, with a population of largely Protestant English and Catholic Irish.² As early as the 1830s, Governor Bourke tried to introduce schools modelled on the Irish National System, with students from all denominations educated in the one school. However, given the sectarianism of the time, these failed. Decades of division between church schools and government-managed schools ensued, and between the 1870s and 1890s each of the Australian colonies passed Education Acts that mandated that education be ‘free, compulsory, and secular’. This essentially stopped most financial assistance to church schools and made education a state responsibility. In addition to cutting them off from state funding, these Acts also cut Catholic and Protestant private schools loose from any state-imposed restrictions. The Protestant schools that remained separate at this time were largely the more elite high-fee schools.

²For far more detailed accounts of this history, see Bonnor & Caro (2007), and Taylor (2018).

The next episode relevant to the growth of the three systems in Australia occurred in the 1960s, when governments began giving money to church schools, with very few conditions. This is summed up perfectly by Bonnor and Caro (2007):

It is a fascinating study of good intentions, short-term solutions, political ambition and expediency, and the final death throes of the old Protestant versus Catholic prejudices that so bedeviled Australian society until the 1960's (p. 35).

The post-war baby boom put huge strain on both government and Catholic schools, the latter of which had traditionally educated children from working class families and were the poor relations of the education system at the time. Fewer people were choosing a life in the church, and, for the first time, Catholic schools were having to employ (and pay) large numbers of lay teachers. In contrast, Protestant schools, also having to employ teachers, took a different path and resorted to charging higher fees, thereby limiting the access to wealthy families. State governments put pressure on the federal government for help in funding education, and eventually this started to occur in various forms. However, Bonnor and Caro (2007) point out that “Among the politics of the day one thing was entirely ignored: that along with public funding should go an established set of public obligations” (p. 37).

This approach to funding, put in place in the 1970s, has had ongoing repercussions that have never been reconciled in terms of funding for the three school sectors, with the funding agreement for Catholic schools flowing on to the rest of the non-Government sector. These repercussions include a change in perceptions of the role of the government schooling system. Connors and McMorrow (2010) noted that “at the beginning of significant Commonwealth funding of schools, the primary obligation of governments was to maintain government school systems at the highest standards, open to all, without fees or religious tests. In 1974, those obligations were enshrined in relevant Commonwealth legislation, but by 2011 they had been expunged from the legislation” (p. 32). By the last year of his government in 2007, Prime Minister John Howard had downgraded the level of education to be acquired from government schools to “... the safety net and guarantor of a reasonable quality education in this country” (Armitage 2007). While the Catholic system, in particular, had traditionally educated children from poor families, this is no longer the case, with many families choosing to send their children to these schools for aspirational, rather than religious reasons.³ The failure of successive governments to tie funding to obligations has provided subsidised private schools with a substantial advantage over their public counterparts, an advantage which is not mirrored in school systems in other countries. Bonnor and Caro (2007) conclude that:

The irony has been that the subsidies [to non-government schools], which were initially aimed at bringing poorly resourced private schools up to the resource and achievement levels of public schools, have continued unchecked until they have neatly reversed the original

³As Bonnor and Caro note, Cardinal George Pell, the then Catholic Archbishop of Sydney, commented that 43% of Catholics are educated in government schools and this figure included 69% of Catholic students from families from the lowest third of family income. “[A]s a consequence Catholic schools are not educating most of our poor ... predominantly our schools now cater for the huge Australian middle-class, which they helped create” (2007, p. 109).

situation they were set up to rectify... public schools are now the resource-poor relations in the education system (p. 38).

In most other countries, there are only relatively small proportions of students attending non-government schools. While there has been a move away from government schools to the non-government sector over the past 20 years, there have been some returns over the past few years to government schools, however, currently, the Catholic system enrolls 23% of Australian secondary school children, independent schools 18% and government schools 59%. This is one of the highest rates of non-government schooling in the OECD.

Constitutionally, school education is the responsibility of the states, and they provide most of the funding for government schools (about 88% nationally). While it does not operate any schools itself, and is under no obligation to do so, the federal government provides the balance of funding to government schools and the majority of funding to non-government schools. According to the latest figures available on the website for the Department of Education, Skills and Employment (Australian Government 2020), around three-quarters of the funding for Catholic schools and less than one-half of the funding for independent schools is from public purses, compared to 95% of funding for government schools. Federal government funding is allocated based on an estimate of how much government funding each school requires to meet the educational needs of its students. This estimate is calculated by reference to the Schooling Resource Standard (SRS), which provides a base amount for every primary and secondary student, along with six loadings that provide extra funding for disadvantaged students and schools. For most non-government schools, the base amount is discounted or reduced by the anticipated capacity of the school community to financially contribute towards the school's operating costs. This is called the 'capacity to contribute' assessment and it is based on a direct measure of median income of parents and guardians of the students at the school. This money is then provided to the state and territory governments and to organisations such as the Catholic education system—which then distribute the money to individual schools according to their own formulas, and with no requirement for transparency as to how funds are distributed.

With widespread dissatisfaction among educational stakeholders in the equity of the funding system, 2011 saw a major review led by David Gonski. The primary aim of this review was to “develop a funding system for Australian schooling which is transparent, fair, financially sustainable and effective in promoting excellent outcomes for all Australian students” (Gonski 2011, p. xiii). Harking back to the aims of the early education agreements, the review argued that funding should aim to ensure that differences in educational outcomes were not the result of non-school factors such as a student's socioeconomic background. One of the primary recommendations of the review panel was that “a significant increase in funding is required across all schooling sectors, with the largest part of this increase flowing to the government sector due to the significant numbers and greater concentration of disadvantaged students attending government schools. Funding arrangements for government and non-government schools must be better balanced to reflect the joint contribution of

both levels of government in funding all schooling sectors” (Gonski 2011, p. xv). Unfortunately the Labor government of the time failed to implement the changes as directly recommended by the Gonski panel (making the promise that “no school will lose money”), and subsequent governments have made a variety of modifications to the model which, it is argued, have not delivered the funding system nor the benefits envisaged by Gonski (Bonnor & Shepherd 2016; Boston 2016; Goss & Sonnemann 2016; Rorris 2016). Over the past 10 years, funding increases have been misdirected towards private schools rather than to government schools. Data released by ACARA show that between 2009 and 2017, government funding (adjusted for inflation) for government schools was cut by \$17 per student (-0.2%) while funding for Catholic schools increased by \$1,420 per student (18.4%) and for independent schools by \$1,318 (20.9%) (Cobbold 2019). To cap it off, while all schools are theoretically able to charge fees, such fees are not compulsory in government schools and are not able to be levied to the extent they are in non-government schools. While many government schools struggle with outdated and worn out facilities, lack of physical resources such as photocopy paper, broken down or inadequate toilet facilities and a lack of teaching staff, some elite independent schools are spending astonishing amounts of money on capital works, including theatres with orchestra pits, indoor Olympic size swimming pools, wellness centres, and equestrian centres. It is estimated that Australia’s four richest schools spent more on new facilities than the poorest 1,800 schools combined between 2013 and 2017 (Ting, Palmer & Scott 2019). The average funding per student, by school sector from all sources, for 2017, is shown in Fig. 1.

Curiously, however, government schools are funded at 85–90% of the Schooling Resource Standard (SRS), while Catholic and independent schools are currently funded at levels either close to 100% of their SRS or at levels even higher than this.

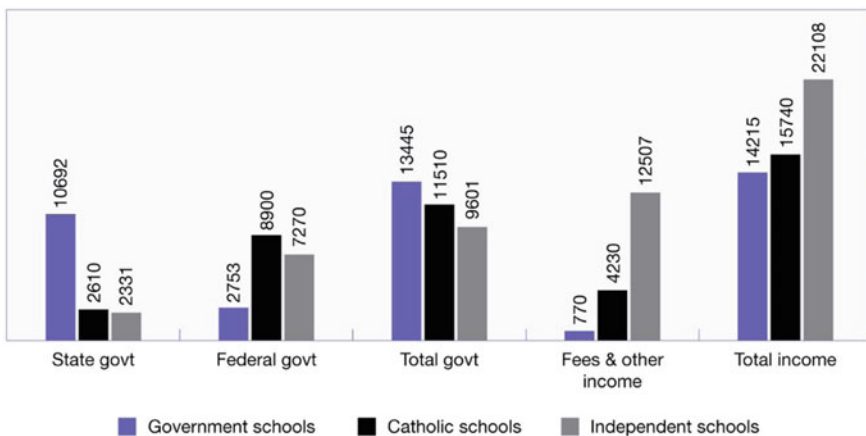


Fig. 1 Australian school income by source per student, by school sector, 2017 (Source ACARA, National Report on Schooling data portal)

3 Is Australia Meeting Its Goals for Schooling?

Given the importance attached to equity and excellence in the Australian National Education Goals since 1999, as well as the attempts to change the funding structures to try and ensure equitable outcomes for all students, it would seem timely to pause and review Australia’s progress towards attaining these goals, using the most recent release of PISA data in 2019.

3.1 Is Australia Attaining Excellence?

Australia’s 2018 PISA results were met on their release in late 2019 with widespread shock and hand-wringing, even though scores have actually been declining since Australia’s initial participation in PISA in 2000. The most recent results saw Australia’s average scores drop to equal the OECD average in mathematical literacy, and those in reading and scientific literacy significantly lower than a decade ago, although still significantly higher than the OECD average.

Figure 2 shows the average scores in achievement for reading, mathematical and scientific literacy for Australian students from 2000 to 2018. In 2000, when reading literacy was first assessed, Australian students achieved a mean score of 528 points, substantially as well as significantly higher than the OECD average of 500 points. In 2009, when it was again a major domain, the score had declined to 515 points, and then in 2018 to 503 points. This decline of 26 points represents a decline of about $\frac{3}{4}$ of a school year in terms of what students can do.⁴

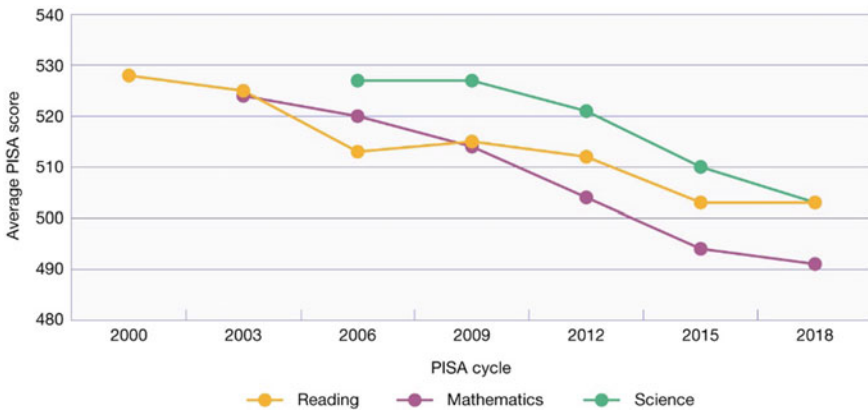


Fig. 2 Australian students’ performance in PISA 2000–2018 (Source OECD 2019)

⁴PISA surveys 15-year-old students nationally. These students are primarily found in Years 9, 10 and 11 in Australian schools. Using regression techniques an approximation can be found in each subject to the number of points that typically represent “one year of schooling” in Australian schools:

In 2003, Australia’s average score in mathematical literacy was 524 points, again substantially as well as significantly higher than the OECD average of 500 points. In 2012 when mathematical literacy was again a major domain, the average score for Australian students was just 504 points, and in 2018 had declined further to 491 points. This score was not significantly different to the OECD average—which had also declined over time to 489 points—and represents a decline from 2003 of almost 1¼ years of schooling in what students can do.

In 2006, when scientific literacy was first assessed, Australia achieved a mean score of 527 points. In 2015, when it was again a major domain, the score had declined to 510 points, and in 2018 to 503 points. This represents almost one full school year decline between 2006 and 2018.

In 2011, the Gonski panel warned that:

Australian schooling needs to lift the performance of students at all levels of achievement, particularly the lowest performers. Australia must also improve its international standing by arresting the decline that has been witnessed over the past decade. For Australian students to take their rightful place in a globalised world, socially, culturally and economically, they will need to have levels of education that equip them for this opportunity and challenge (Gonski 2011, p. 22).

Evidence suggests that this has not been the case. In Fig. 3, PISA proficiency levels in reading, mathematical and scientific literacy have been grouped into *high performers*, those who achieve at proficiency level 5 and above, and *low performers*,

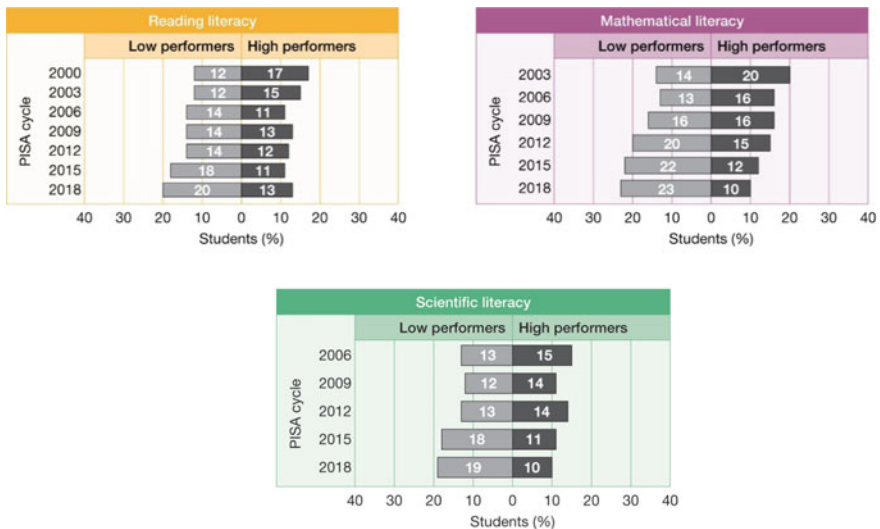


Fig. 3 Percentages of high and low performers in reading, mathematical and scientific literacy, 2000–2018, Australia (Source OECD 2019)

33 points on the PISA reading literacy scale, 28 points on the mathematical literacy scale, and 27 points on the scientific literacy scale.

those achieving below proficiency level 2. In 2000, 17% of Australian students were high performers in reading literacy, and 12% low performers. In 2018, 13% were high performers in reading literacy and 20% low performers. In mathematical literacy the situation has become even more dire. In 2003, 20% of Australian students were high performers, 14% low performers. In 2018, 10% were high performers, and 23% low performers. In scientific literacy the situation has also deteriorated: from 13% low performers and 15% high performers in 2006 when scientific literacy was first assessed, to 19% low performers and 10% high performers in 2018.

Over time, the gap between the high achievers and the low achievers has increased, particularly in reading literacy. This is largely due to a larger decline at the lower percentiles of performance (Fig. 4). Over the PISA cycles, performance in reading literacy at the 10th percentile declined by 38 points (about 1.2 years of schooling), performance at the 90th percentile declined by 15 points (less than half a year of schooling). The difference between the highest and lowest percentiles in 2000 was 261 points (almost 8 years of schooling), which had increased to 284 points (8.6 years of schooling) in 2018. In mathematical literacy scores at the 10th percentile declined by 27 points (about one school year), and at the 90th percentile by 35 points (about 1¼ school years), between 2003 and 2018. The gap between highest and lowest remained roughly the same—246 score points in 2003 and 238 in 2018. Changes in scientific literacy scores have been similar: performance at the 10th percentile declined by 25 points (almost one year of schooling) between 2006 and 2018, and at the 90th percentile by a similar 23 points. In PISA 2006 the difference between high and low performers was 259 points and in 2018 it was 262 points.

3.2 *Is Australia Attaining Equity?*

The Alice Springs (Mparntwe) Education agreement argues that “the education community must improve outcomes for educationally disadvantaged young Australians” (COAG Education Council 2019, p. 17), and identifies educationally disadvantaged as students from low socioeconomic backgrounds, Aboriginal and Torres Strait Islander students, and students from regional, rural, and remote areas—among others—but this chapter will concentrate on these groups.

3.2.1 **Students from Low Socioeconomic Backgrounds**

If a student’s social background is not a determinant of their achievement, then achievement levels would be evenly distributed across socioeconomic groups. To what extent is this the case for Australia?

The primary measure used by the OECD to represent socioeconomic background in PISA is the index of economic, social and cultural status (ESCS), which was created to capture the wider aspects of a student’s family and home background. The ESCS is based on three indices: the highest level of the father’s and mother’s occupations

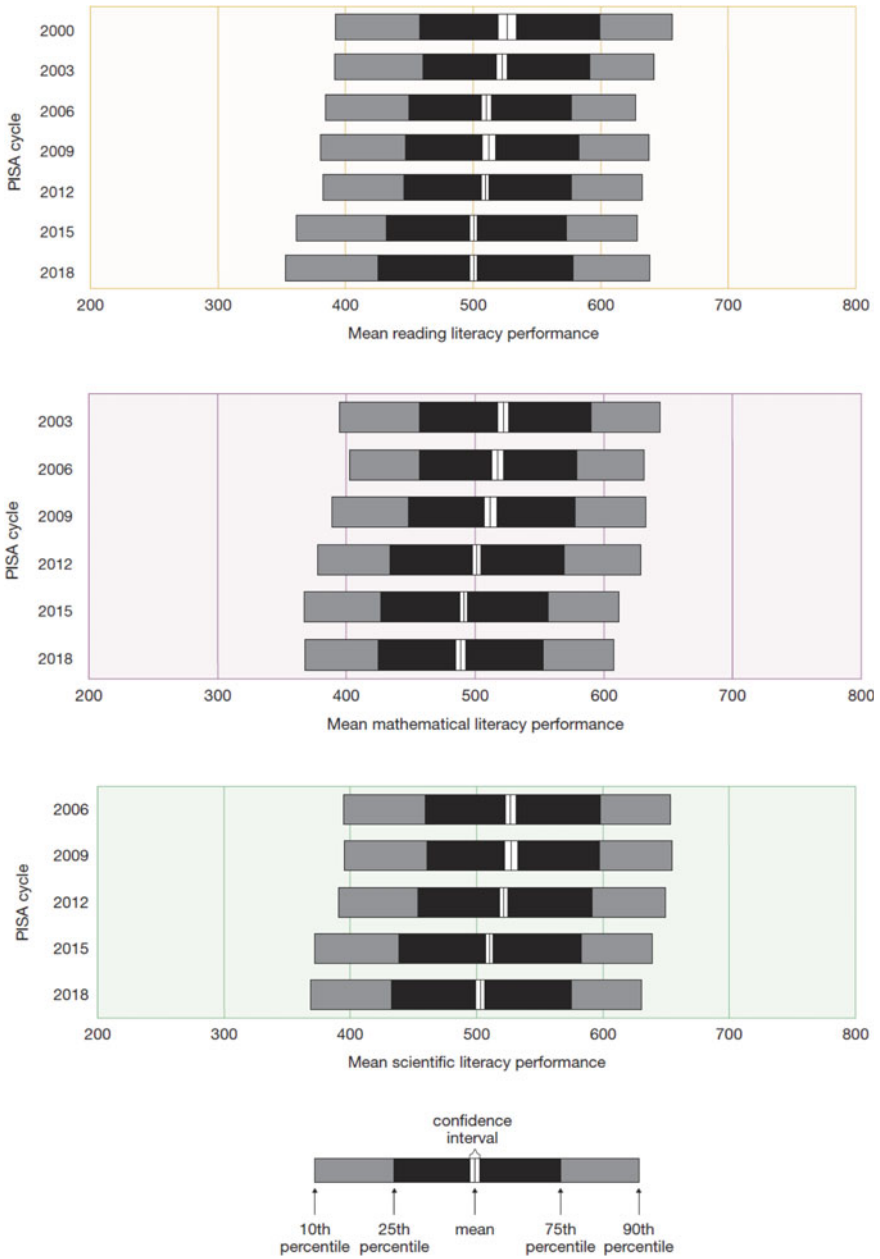


Fig. 4 Distribution of student performance on the reading, mathematics and science literacy scales, PISA 2000–2018, Australia (Source Thomson et al. 2019)

(known as the highest international social and economic index—HISEI), which is coded in accordance with the International Labour Organization’s International Standard Classification of Occupations; the highest educational level of parents in years of education (PARED); and home possessions (HOMEPOS). The index HOMEPOS comprises all items on the indices of family wealth (WEALTH), cultural resources (CULTPOSS), and access to home educational and cultural resources and books in the home (HEDRES). It must be noted that there have been some adjustments to the computation of ESCS over the PISA cycles.

The average score for students who were in the lowest quartile of ESCS in PISA 2018 (disadvantaged students) was 460 points in reading literacy, compared to 549 points which was the average for those in the highest quartile (advantaged students). This difference of 89 points represents about 2.7 years of schooling. In mathematical literacy the average score for disadvantaged students was 451 points and for advantaged students 532 points, a difference of 81 points representing 2.9 years of schooling. In terms of international positions, these scores would place advantaged students at the same achievement level in reading literacy as those in the highest achieving PISA countries, B-S-J-Z China⁵ and Singapore, and disadvantaged students around the same level as the Slovak Republic and Greece. Figure 5 shows the distribution of proficiency levels in reading, mathematical and scientific literacy across socioeconomic background. Clearly, there *are* substantial differences in achievement across socioeconomic level in Australia in these key areas of literacy.

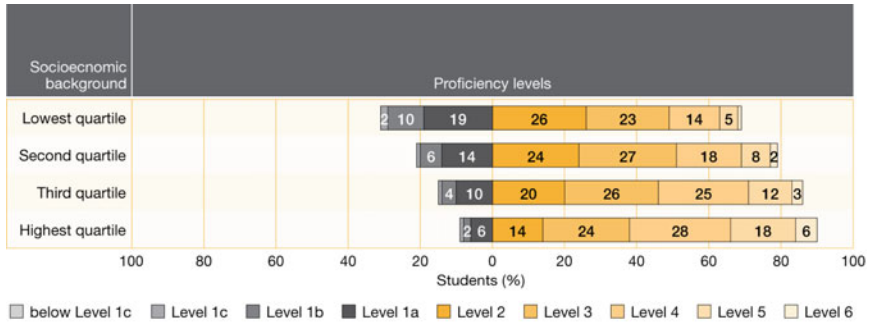
Moreover, this has been the case since the first administrations of PISA. In 2000, as shown in the top left panel of Fig. 6, 21% of disadvantaged students were low achievers in reading literacy. Results from the latest round of PISA in 2018 show that this situation has deteriorated, with 31% of disadvantaged students now classed as low performers. In 2003, 26% of disadvantaged students were low performers in mathematical literacy, and in 2018 this had risen to 37% of this group of students. In 2006, 23% of disadvantaged Australian students were low performers in scientific literacy, in 2018 this had risen to 31% of disadvantaged students.

What should be positive news is that the gap between the average score of advantaged and disadvantaged students has narrowed slightly in all three literacy areas (Fig. 7) from 102 point to 89 points in reading, from 92 points to 81 points in mathematical literacy, and from 91 points to 83 points in scientific literacy. It should be noted though that the gap only narrows from about 3 years of schooling to 2.7 years of schooling in reading literacy, from 3.3 years to 2.9 years in mathematical literacy, and from 3.4 years to 3.1 years in scientific literacy.

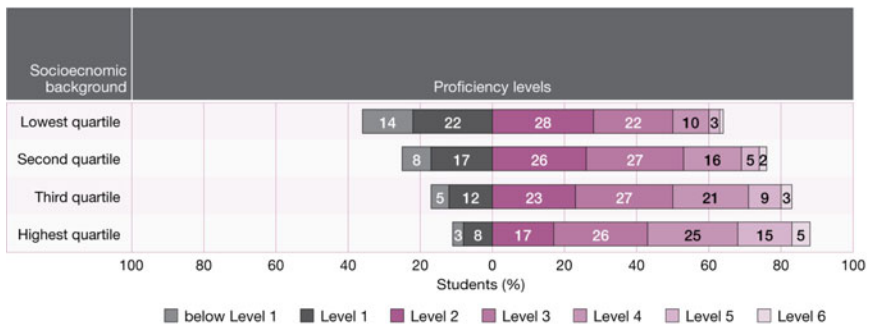
However, in reality, this narrowing is due to the larger decline in the scores of the advantaged students in all areas. In reading literacy, the average scores for disadvantaged students declined by 24 points—where the decline for those in the highest quartile was 37 points. In mathematical literacy the average score for disadvantaged students declined by 28 points, while the decline for advantaged students was 40

⁵The four provinces of China that participated in PISA 2018: Beijing, Shanghai, Jiangsu and Zhejiang.

Reading Literacy



Mathematical literacy



Scientific literacy

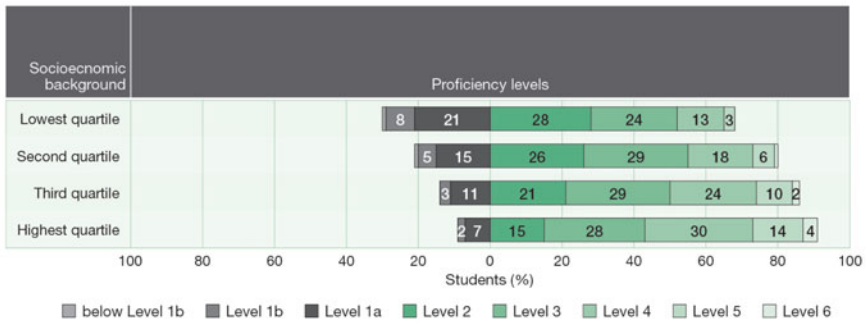


Fig. 5 Percentages of students across the reading, mathematical and scientific literacy proficiency scales by socioeconomic background, PISA 2018, Australia (Source Thomson et al. 2019)

Disadvantaged students

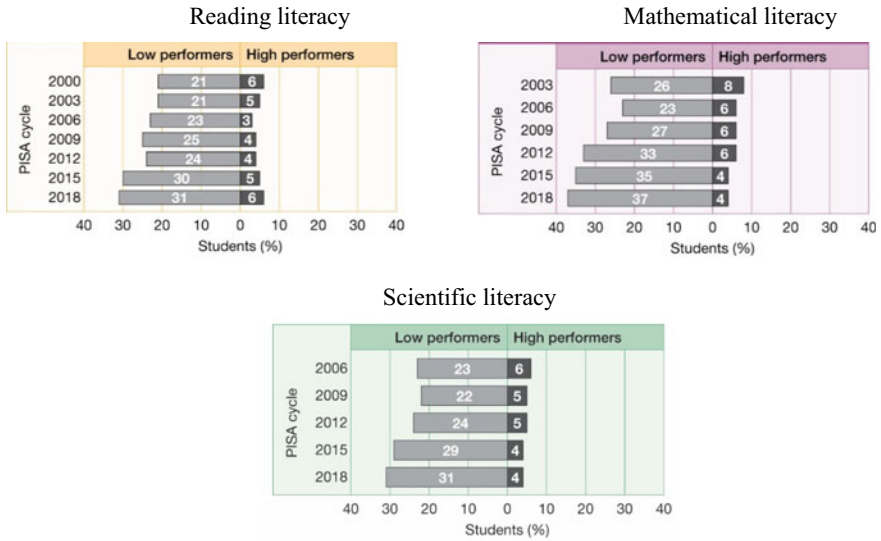


Fig. 6 Proportions of low performers in reading, mathematical and scientific literacy for students from a low socioeconomic background over time, PISA 2000–2018, Australia (Source Thomson et al. 2019)

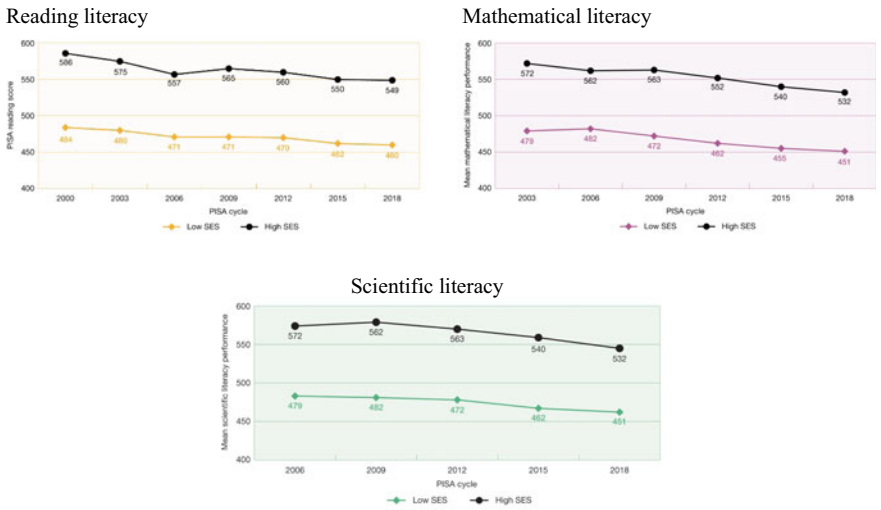


Fig. 7 PISA reading, mathematical and scientific literacy scores over time, advantaged and disadvantaged students, PISA 2000–2018, Australia (Source Thomson et al. 2019)

points. In scientific literacy the average score for disadvantaged students declined by 21 points and the decline for advantaged students 29 points.

3.2.2 Students from an Aboriginal or Torres Strait Islander Background

Traditionally, students from an Aboriginal and Torres Strait Island background have been poorly served by the Australian education system (Gray & Beresford 2008). Reflecting on the first of the reports on *Overcoming Indigenous Disadvantage* in 2003, the Steering Committee Chair commented that “It is distressingly apparent that many years of policy effort have not delivered desired outcomes: indeed in some important respects the circumstances of Indigenous people appear to have deteriorated or regressed” (Steering Committee for the Review of Government Service Provision 2005, p. xix).

PISA 2000 provided a first measure of the gap between Indigenous and non-Indigenous students, with a gap of 83 points in reading literacy (2.5 years of schooling) followed by similar gaps in subsequent rounds of PISA - 86 points in mathematical literacy in 2003 (3.1 years of schooling) and 88 points in scientific literacy in 2006 (3.3 years of schooling). In PISA 2018, a decline in the scores of non-Indigenous students in all three assessment areas brought the gaps to 76 points (2.3 years of schooling) in reading literacy, 69 points (2.5 years of schooling) in mathematical literacy and 75 points (2.8 years of schooling) in scientific literacy. Again, not the envisaged means of closing the gap (Fig. 8).

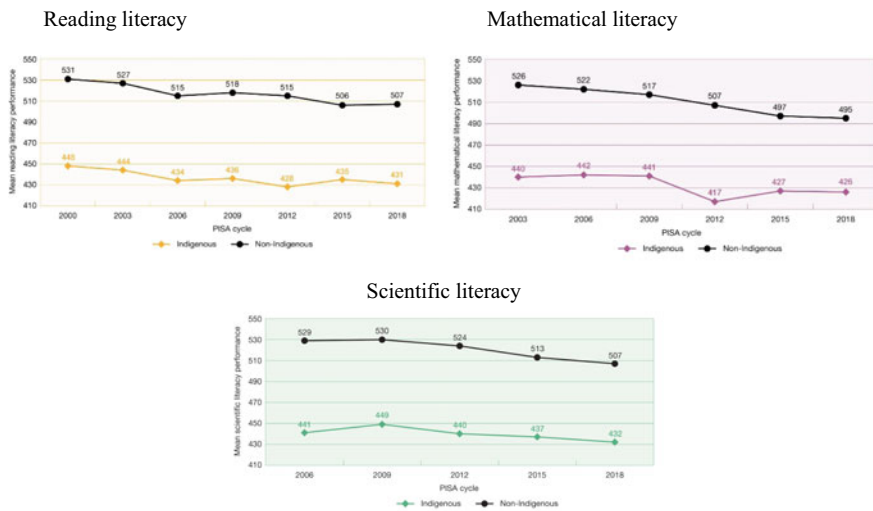


Fig. 8 Mean reading, mathematical and scientific literacy scores over time, for Indigenous and non-Indigenous students, PISA 2000–2018, Australia (Source Thomson et al. 2019)

Indigenous students

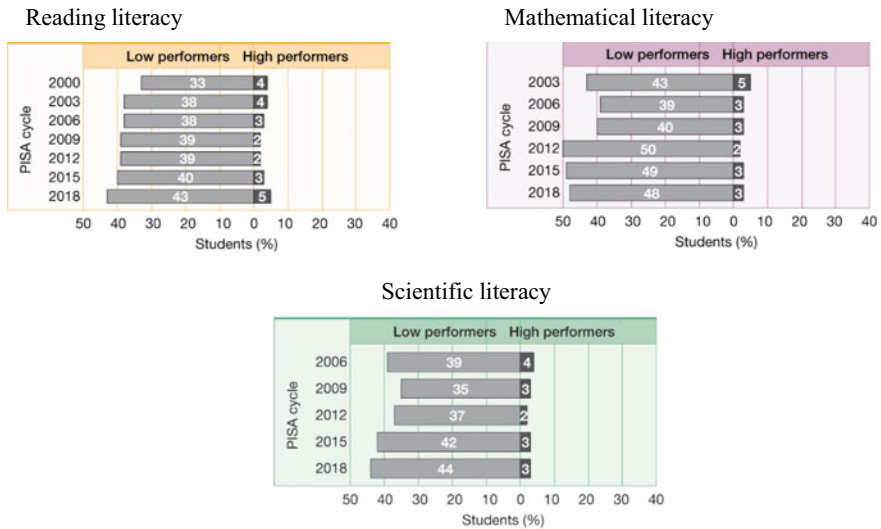


Fig. 9 Proportions of low performers in reading, mathematical and scientific literacy for Indigenous students over time, PISA 2000–2018, Australia (Source Thomson et al. 2019)

Of particular concern is the proportion of low-performing Indigenous students in all three assessment areas, and this has worsened over time (Fig. 9). In 2000, 33% of Indigenous students were low performers in reading literacy, and in 2018 this had increased to 43 per cent. In mathematical literacy in 2003, 43% of Indigenous students were low performers, and this has hovered around the 50% mark in recent years. In scientific literacy, 39% of Indigenous students were low performers, and this increased to around 44% in 2018. These proportions are also most likely an underestimate of the actual proportions as PISA is unable to assess many Indigenous students—those who live in extremely remote areas, those who do not have instruction in English, and those who do not attend on the days of testing.

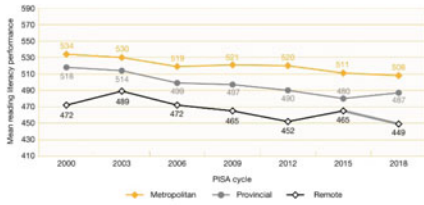
3.2.3 Students from Regional and Remote Areas

In Australia in 2018, participating schools were coded broadly as:

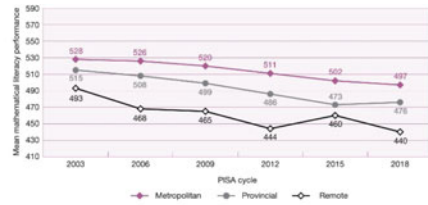
- metropolitan—mainland capital cities or major urban districts with a population of 100,000 or more
- provincial—provincial cities and other non-remote provincial areas
- remote—areas with very restricted or very little accessibility to goods, services and opportunities for social interaction.

The average reading literacy score for students in metropolitan schools in PISA 2018 was 508 points (Fig. 10). This achievement was significantly higher than the

Reading literacy



Mathematical literacy



Scientific literacy

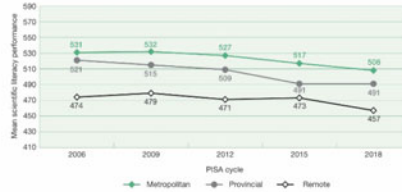


Fig. 10 Mean reading and mathematical literacy scores over time, metropolitan, provincial and remote areas, PISA 2000–2018, Australia (Source Thomson et al. 2019)

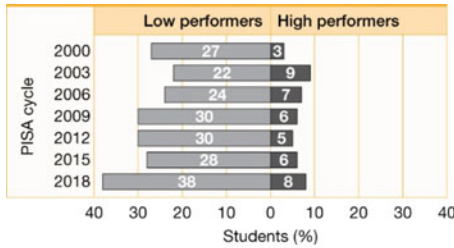
score for those in provincial schools of 487 points, which was in turn, significantly higher than the score for those in remote schools of 447 points. Over time, the average scores for students in both metropolitan and provincial schools have declined significantly (by 26 points and 31 points respectively), while the score for students in remote schools declined from a peak in 2003 of 489 points to the current mean of 449 points. The gap between students in metropolitan schools and those in remote schools is much the same as in 2000, and is a little less than two years of schooling.

In mathematical literacy the differences are more dramatic. The average mathematical literacy score for students in metropolitan schools in 2018 was 497 points. This achievement is significantly higher than the score for students in provincial schools of 476 points, which was in turn, significantly higher than the average score for students in remote schools of 440 points. Over time, scores declined both significantly and substantially for all groups: by 31 points for students in metropolitan schools, 39 points for those in provincial schools and 53 points for those in remote schools. The gap in performance between those in metropolitan schools and those in remote schools has gone from 35 points in 2003 (around 1.25 years of schooling) to 57 points (2 years of schooling) in 2018.

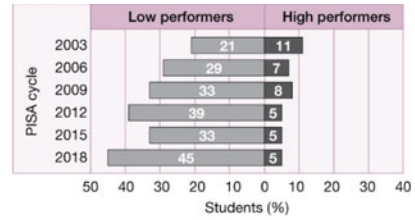
In scientific literacy in 2018 the average score for students in metropolitan schools was 508 points, 18 points higher than those in provincial schools, and 51 points higher than for those attending remote schools. Over time, scores in scientific literacy have declined by 23 points for students in metropolitan schools, 30 points for those in provincial schools and 17 points for those in remote schools. The gap in performance between students in metropolitan and remote schools has remained about the same since 2006—57 points in 2006 and 51 points in 2018. These are both around two years of schooling.

Remote students

Reading literacy



Mathematical literacy



Scientific literacy

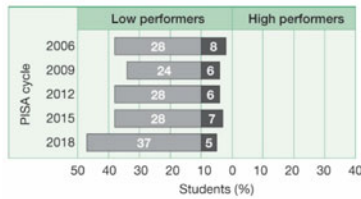


Fig. 11 Proportions of low performers in reading, mathematical and scientific literacy for students in remote areas over time, PISA 2000–2018, Australia (Source Thomson et al. 2019)

In terms of proficiency levels, the proportion of low performers amongst students in rural areas has increased over time across all three assessment domains (Fig. 11). In reading literacy in 2000, 27% of rural students were low performers in 2018 this had increased to 38 per cent. In mathematical literacy the proportion has more than doubled—from 21% of students in 2003 to 45% in 2018, and in science the proportion has increased from 28% of rural students in 2006 to 37% in 2018.

4 Relationship Between School Sector and Disadvantage

It is evident from the results for PISA 2018 that Australia is not meeting its own targets of excellence and equity, and it is far from being on track to meet the goal of being in the “top five by 2025” (however that goal was intended to be measured). Despite apparent increased levels of funding over the last 18 years, the introduction of a national curriculum, the establishment of national agencies to develop national standards for teaching and school leadership and a national testing program of students at a range of age levels, as well as participation in international studies of assessment, average scores have declined year after year.

In investigating the intersection of student performance and funding, it is informative to look at which schools advantaged and disadvantaged students attend, and in particular, where there are multiple layers of disadvantage.

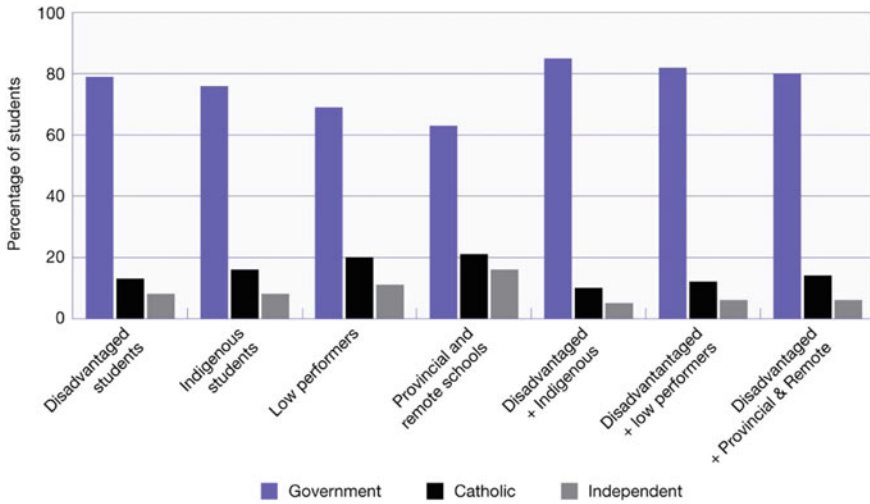


Fig. 12 Type of school attended by disadvantaged groups, PISA 2018, Australia (Source OECD 2019)

Government schools enroll the vast majority of students who fall into the categories of disadvantaged groups as defined by the governments of Australia (Fig. 12). Forty-one percent of government schools can be classed as disadvantaged schools,⁶ compared to three percent of Catholic and less than one percent of independent schools. In contrast, ten percent of government schools, 31% of Catholic schools, and 63% of independent schools, and are classed as advantaged schools. Over 80% of disadvantaged students attend government schools.

Over the past 18 years, analysis of school market share using Geographic Information System (GIS) technology has found that recent enrolment shifts are largely towards government schools in high SES areas, and towards non-government schools in lower SES areas. Further analysis (Bonnor & Shepard 2016) using NAPLAN data shows that, in general, it is the more advantaged students who are moving to the more advantaged schools. As these more disadvantaged students have moved to more advantaged schools, the students remaining in schools lower down the socio-economic scale lose diversity and talent, and their school body contains a higher proportion of disadvantaged students. This creates a cycle where some parents identify schools as low performing or high disadvantage and, if possible (that is, if they are financially able to do so), enroll their children at schools with higher proportions of advantaged students.

Table 1 provides a very brief overview of some of the differences between advantaged and disadvantaged schools from PISA 2018.⁷ These data paint a picture of

⁶Defined as those whose average intake of students falls in the bottom quarter of the PISA index of economic, social and cultural status within the country, compared to advantaged schools, defined as those whose average intake of students falls in the top quarter of that index.

⁷Personal calculations, Australian PISA 2018 data (OECD 2019).

Table 1 Principal's views on hindrances to providing instruction (Australia) (Source OECD 2019)

		Disadvantaged schools (%)	Advantaged schools (%)
Percentage of students in schools whose principal reported that the school's capacity to provide instruction is hindered at least to some extent by	Lack of teaching staff	34	3
	Inadequate or poorly qualified teaching staff	21	0.3
	Teacher absenteeism	28	5
	Teachers not well prepared	18	5
	Lack of educational material	21	1
	Inadequate or poor educational material	21	0.3
	Lack of physical infrastructure	45	6
	Lack of student respect for teachers	16	0.3

less qualified and less well-prepared teachers, issues with teacher absenteeism, lack of materials and lack of physical infrastructure at a substantial proportion of disadvantaged schools, but rarely at advantaged schools.

In addition to a lack of resources, PISA 2018 data show that 21% of students attending disadvantaged schools compared with 0.8% of students attending advantaged schools are enrolled in schools in which the principal reports that the school's capacity for instruction is hindered at least to some extent by students intimidating or bullying other students.

5 Conclusions

The factors described in this chapter have set Australia up to have a large number of young people whose experiences of education are less than they could be, and who are being failed by our current system. Many of these students cope with multiple layers of disadvantage. At present, these students are not being adequately supported by government education policies that fail to provide funding where it is most desperately needed—for basics such as infrastructure and materials, good quality teachers, or enough teachers. Their outcomes reflect this lack of provision of basic educational services. There are a substantial number of studies published in recent years which demonstrate that increased expenditure on schools improves student outcomes, particularly for disadvantaged schools and students (for example Baker 2019; Darling-Hammond 2019; Kirabo Jackson 2018).

Improving Australia's PISA score is not an outcome in itself, it would simply be a reflection of an improvement in the health of the educational system overall, as

that is what PISA was designed to measure. Improving the health of the educational system can be brought about by actively directing adequate funding to schools that are attempting to educate an increasingly diverse student population, many of whom are already experiencing multiple challenges to their engagement with and mastery of the curriculum. To a large extent these are government schools, and increased funding of these schools is essential to provide the human and physical resources needed by these students. As the OECD state:

Achieving equity in education means ensuring that students' socio-economic status has little to do with learning outcomes. Learning should not be hindered by whether a child comes from a poor family, has an immigrant background, is raised by a single parent or has limited resources at home, such as no computer or no quiet room for studying. Successful education systems understand this and have found ways to allocate resources so as to level the playing field for students who lack the material and human resources that students in advantaged families enjoy. When more students learn, the whole system benefits. This is an important message revealed by PISA results: in countries and economies where more resources are allocated to disadvantaged schools, overall student performance in science is somewhat higher. (OECD 2016, p. 233).

Author's Addendum

During the preparation of this chapter world events have not stayed still. In this time, the face of education as we know it has been forever changed, by the COVID-19 pandemic. In Australia it reached our shores in the final weeks of Term 1, 2020. Over the following 6 weeks, education systems moved at frenetic pace to bring online learning to as many students as possible, as quickly as possible. Schools were mostly closed early and required to send students some online work. It is notable that many private schools had the resources at school to provide such curricula/programming much more readily than most public schools. Government education policy had many schools for most of Term 2 and in Victoria at least, all of Term 3, positioned as places to be attended on a face-to-face basis, only by children of parents deemed as essential workers or unable to work at home, or for vulnerable students.

Such a dislocation of schooling and the planned abrupt move to online learning, and its ongoing development, brings with it added pressures in terms of the equity for all students of accessing and achieving equitable schooling outcomes. While about 87% of Australians can access the internet at home (Watt 2019), only 68% of Australian children aged 5–14 living in disadvantaged communities have internet at home, compared with 91% of students living in advantaged communities (Smith Family 2013). Of the Australian PISA students, 84% of those from disadvantaged families have access to a computer at home which they can use for study (meaning 16% do not), compared with virtually 100% of students from an advantaged background. Which would be fine for those cohorts as long as all of these families only have one student in the home. With multiple children comes the imperative for multiple computers, however only 71% of disadvantaged households, compared with 99% of advantaged households, have two or more computers. With added family stresses due to sudden working from home, unemployment of family members, associated anxiety and lack of experience or in understanding how best to facilitate

children's online learning, as well as possibly a lack of appropriate technological skills, this pandemic is almost certain to have a profound and exacerbating impact on the educational outcomes for many disadvantaged students.

Now, more than ever, Australia needs to recognize that our students do not currently benefit equally from their learning, and that online learning, especially in the context of the planned implementation, will almost inevitably worsen the achievement outcomes of the disadvantaged. To assume the nation's students will have an equal capacity to take up and engage in online learning 'inherently privileges the wealthy and further entrenches a multi-tiered educational model' (Graham & Sahlberg 2020).

Australia has an opportunity here and now to make some wholesale changes to the national provision of education provision, and thus necessarily to educational funding. Policy makers and citizens have this unique opportunity, one such as they have never previously had, to insist on equitable educational provision. Future generations, with the benefit of hindsight, will clear-sightedly judge whether this opportunity was grasped with both hands or was squandered.

References

- Armitage, C. (2007, May 16). Vouchers not an option in the new market model. *The Australian*, p. 21.
- Australian Curriculum, Assessment and Reporting Authority (ACARA). (no date). National Report on Schooling: Data Portal. Retrieved from <https://www.acara.edu.au/reporting/national-report-on-schooling-in-australia/national-report-on-schooling-in-australia-data-portal>. Accessed April 13 2020.
- Australian Education Council. (1989a). The Hobart Declaration on Schooling (1989). <http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Hobart-Declaration-on-Schooling-1989.aspx>. Accessed April 13 2020.
- Australian Education Council. (1989b). Common and agreed goals for schooling in Australia. The Hobart Declaration on Schooling. <http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Hobart-Declaration-on-Schooling-1989.aspx>. Accessed April 13 2020.
- Australian Government. (2013). Australian Education Act 2013. <https://www.comlaw.gov.au/Details/C2013A00067>. Accessed April 13 2020.
- Australian Government. (2020). "How are schools funded in Australia". <https://www.education.gov.au/how-are-schools-funded-australia>. Accessed April 13 2020.
- Baker, Bruce D. (2019). *Does money matter in education*. Washington DC: Albert Shanker Institute.
- Bonnor, C., & Caro, J. (2007). *The stupid country: How Australia is dismantling public education*. Sydney: UNSW Press.
- Bonnor, C. & Shepherd, B. (2016). Uneven playing field: The state of Australia's schools. <https://cpd.org.au/wp-content/uploads/2016/05/The-State-of-Australias-Schools.pdf>. Accessed April 13 2020.
- Boston, K. (2016, September 6). What Gonski really meant, and how that's been forgotten almost everywhere. *Inside Story*. <https://insidestory.org.au/what-gonski-really-meant-and-how-thats-been-forgotten-almost-everywhere/>. Accessed April 13 2020.
- Cobbold, T. (2019). The facts about school funding in Australia. <https://saveourschools.com.au/funding/the-facts-about-school-funding-in-australia/>. Accessed April 13 2020.

- Connors, L., & McMorrow, J. (2010). *New directions in schools funding: A proposed model*. Sydney: University of Sydney, Faculty of Education and Social Work.
- Council of Australian Governments Education Council. (2019). Alice Springs (Mparntwe) Education Declaration. <http://www.educationcouncil.edu.au>. Accessed April 13 2020.
- Darling-Hammond, L. (2019). *Investing for student success: Lessons from state school finance reforms*. Palo Alto, CA: Learning Policy Institute.
- Gonski, D. (2011). Review of funding for schooling—Final report. Canberra: Department of Education, Employment and Workplace Relations.
- Goss, P. & Sonneman, J. (2016). Circuit breaker: A new compact on school funding. <https://grattan.edu.au/wp-content/uploads/2016/11/881-Circuit-Breaker-New-Compact-replaced-3.pdf>. Accessed April 13 2020.
- Graham, A. & Sahlberg, P. (2020). Schools are moving online, but not all children start out digitally equal. <https://theconversation.com/schools-are-moving-online-but-not-all-children-start-out-digitally-equal-134650>. Accessed April 13 2020.
- Gray, D. J., & Beresford, D. Q. (2008). A 'formidable challenge': Australia's quest for equity in Indigenous education. *Australian Journal of Education*. <https://doi.org/10.1177/000494410805200207>.
- Kirabo Jackson, C. (2018 Dec). Does school spending matter? The new literature on an old question. NBER Working Paper No. 25368. Cambridge, MA: National Bureau of Economic Research.
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). (1999). Adelaide declaration on national goals for schooling in the twenty-first century. <http://www.educationcouncil.edu.au/EC-Publications/EC-Publications-archive/EC-The-Adelaide-Declaration.aspx>. Accessed April 13 2020.
- Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA). (2008). Melbourne Declaration on Educational Goals for Young Australians. http://www.curriculum.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf. Accessed April 13 2020.
- OECD. (2016). *PISA 2015 results vol II: Policies and practices for successful schools*. Paris: OECD.
- OECD. (2019). *PISA 2018 database*. <https://www.oecd.org/pisa/data/2018database/>. Accessed April 13 2020.
- Rorris, A. (2016). Australian schooling: The price of failure and reward for success. <http://www.aefederal.org.au/application/files/3814/6172/5096/Rorris2016.pdf>. Accessed April 13 2020.
- Smith Family. (2013). *Sport, culture and the internet: Are Australian children participating?*. <https://www.thsmithfamily.com.au/-/media/files/research/reports/participation-research-report-June-2013.pdf?la=en&hash=C9D689EF3DE3C0B8BB99DE6D5EC76A79>. Accessed April 13 2020.
- Steering Committee for the Review of Government Service Provision (SCRGSP). (2005). *Overcoming indigenous disadvantage: Key indicators 2005*. Productivity Commission.
- Taylor, T. (2018). *Class wars: Money, schools and power in modern Australia*. Australia: Monash University Publishing.
- Ting, I., Palmer, A. & Scott, N. (2019). Rich school, poor school: Australia's great education divide. *ABC News*. <https://www.abc.net.au/news/2019-08-13/rich-school-poor-school-australias-great-education-divide/11383384>. Accessed April 13 2020.
- Thomson, S., De Bortoli, L., Underwood, C., Schmid, M. (2019). PISA 2018: Reporting Australia's results. Volume I student performance. <https://research.acer.edu.au/ozpisa/35>. Accessed April 13 2020.
- Watt, E. (21 June 2019). The digital landscape is changing in Australia—Where we are, and where we're going. <https://www.roi.com.au/blog/australian-internet-social-media-statistics-2019>. Accessed April 13 2020.

Dr. Sue Thomson was recruited to ACER in 1998 to work on the *Longitudinal Surveys of Australian Youth*. She has managed Australia’s participation in the IEA’s *Trends in International Mathematics and Science Study* (TIMSS) since 2001, the *OECD Programme for International Student Assessment* (PISA) since 2003 and the *IEA Progress in International Reading Literacy Study* (PIRLS) since Australia joined the study in 2011. This has involved extensive reporting on all studies. Since 2007 she has been the Research Director of the Australian Surveys program at ACER with responsibility for the management of the international studies within Australia, as well as other national assessment programs in schools which assess and report on reading, mathematical and scientific literacy, civics and citizenship education and ICT literacy. Dr Thomson was appointed as ACER’s Deputy CEO (Research) in 2017. She provides senior leadership to staff in several research programs including *Surveys and International Assessments*, *Systemwide Testing*, *Educational Policy and Practice*, *Education and Development* and *Tertiary Education*. In October 2018 Dr Thomson was elected to join the Standing Committee of the International Association for the Evaluation of Educational Achievement (IEA). She is an associate Editor of the *Australian Journal of Education*, and a regular columnist for ACER’s *Teacher* magazine.

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.



Chile: The Challenge of Providing Relevant Information from ILSA Studies for the Improvement of Educational Quality



EMA LAGOS

Abstract Chile has a consolidated culture of evaluation in its educational system because, for more than three decades, first the Ministry of Education and currently the National Agency for Educational Quality have implemented national census tests every year to monitor the established curricula' learning. International Large-scale Students Assessment (ILSA) studies have substantially contributed to this monitoring since the late 1990s. Both, the definition of the disciplines and domains evaluated and the results obtained, have motivated curricular reforms to adapt what is taught to children and young people to prepare them for a globalized world, with a strong presence of information and communication technology. The Chilean students' results have impacted the system, especially by highlighting its weaknesses, related to little improvement over decades, differences in learning achieved by different groups of students, and performance below than expected in the most economically and culturally advantaged sectors. To accomplish these challenges, the system has changed its organization and developed diverse strategies. Data provided by ILSA studies have been used to promote policies and programs for the improvement and strengthening of the most vulnerable groups and a general approach that promotes gender equality in education, politics, and labor. ILSA studies have also been a reference for innovation in educational assessments, allowing the country to evaluate and explore innovative learning areas such as digital and financial competences.

1 Overview of Chilean Education System

The Political Constitution (1980) consecrates and ensures the right to education, which allows the full development of the person at different stages of his/her life; it establishes the compulsory nature of primary and secondary education and the duty of the state to finance a free system to ensure access for the entire population. The Constitution also enshrines freedom of education, and the right to open, organize and

E. Lagos (✉)

PISA National Coordinator, National Agency for Educational Quality, Morandé 360, Piso 10, Santiago, Chile

e-mail: ema.lagos@agenciaeducacion.cl

© The Author(s) 2021

N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_3

maintain schools and the right of parents to choose the school for their children. The Constitution also explains the need for an organic constitutional law that establishes minimum requirements and objective norms for the educational system. Since 2009, the General Education Law (LGE)¹ normalizes Chile's educational system framework (Ley N° 20.370 2009). The LGE defines the goals of school education, regulates the rights and duties of the members of the education community, establishes minimum requirements for completion of each of the education levels, and institutes a process for the recognition of education providers (Biblioteca del Congreso Nacional de Chile 2016).

Organizationally, the Chilean educational system is governed by the Quality Assurance System (QAS),² which is mandated to guarantee good quality education for all students in the country. From the beginning of QAS, in 2012, the educational system comprises four institutions with different duties.

The Ministry of Education is the central institution of the QAS. Its purpose is to implement educational policy by granting official recognition to educational institutions, defining regulations, providing funding, and creating and supporting educational resources, learning standards, and pedagogical training. Other institutions of the QAS are the Superintendence of Education, the National Council of Education, and the National Agency for Educational Quality (Ley N° 20.529 2011).

At a local level, educational institutions differ by administrative dependence and by educational tracks. According to their administrative status: public schools (43.9% of total) are managed by local governments (municipalities) or local education services,³ and funded by the state (Ley N° 21.040 2017); private subsidized schools (48.9%) are managed by private entities and funded by the state⁴ (Ley N° 20.845 2016); and paid private schools (7.2%) are managed by private entities and funded exclusively by families. By 2018, there were 12,021 schools in Chile, serving 3.58 million students.

Chile's current school system consists of eight years of primary education (educación básica), a combination of primary and lower secondary education (Grades 1st to 8th),⁵ and four years of secondary school (educación media), which corresponds to upper secondary education (Grades 9th to 12th). Primary education starts

¹Law N° 20.370 (2009), General education law (LGE). It replaced the Constitutional Organic Law of Education (LOCE), Law N° 18.962 in force since March 10, 1990.

²Ruled by Law N° 20.529 (2011), National system for the quality assurance of early childhood, primary and secondary education and its auditing.

³Law N° 21.040 (2017), creates Public Education System (transfer schools from municipalities administration to local education services administration).

⁴Law N° 20.845 (2016), School Inclusion law. Regulates students' admission, eliminates shared financing, and prohibits profits in educational institutions receiving contributions from the state.

⁵Law N° 20.370 (2009), changes basic and secondary education structure into six years of primary and six years of secondary education. Although it is approved, this measure has not been implemented yet.

when students are six years old. In total, there are 13 years of compulsory education from kindergarten⁶ to 12th grade (Ley N° 20.710 2013).

Schools may offer primary education (mainly small rural schools, offer education for Grades 1st to 4th or Grades 1st to 6th only), secondary school education (Grades 7th to 12th), or both (complete schooling). Schools providers of upper secondary education offer humanistic-scientific education, technical professional (vocational) education, or both (polyvalent). The differentiation between humanistic-scientific and technical professional education occurs in 11th grade, and different curricula for each track accompany it. A small group of schools offers specific artistic education. For students with special needs, temporary or permanent, there are economical, human, and technical resources, and specific knowledge and assistance available. Students with special needs can attend regular schools where facilities and methodologies are adapted for them,⁷ or they attend special schools, organized by type of disability.

1.1 The National Agency for Educational Quality in QAS

In recent years, several efforts have been implemented in Chile to improve quality and equity in the education system. Among the concrete measures to support the improvement of quality, meet the Framework for Good Teaching (published in 2003), the Framework for Good Management (published in 2005), in 2011 was established the Quality Assurance System⁸ of Early childhood education, Basic and Secondary Education and its Inspection (QAS). Within this system, the National Agency for Educational Quality inherited from the former Unit of Curriculum and Evaluation of the Ministry of Education the responsibility to evaluate learning outcomes of students, through national tests (“SIMCE” is a national census assessment conducted every year in particular grades in specific subjects from 1988) and other Indicators of educational quality mostly related with socio-emotional aspects, with also the implementation of ILSA studies. This implementation includes all the procedures and processes for sampling, translation, and adaptation of instruments, test administration, database elaboration, and the publication of a first national results report.

⁶Law N° 20.710 (2015), establishes the obligatory of the second transition level (kindergarten) and creates a free financing system from the middle level.

⁷School Integration Program (PIE). Since 2009, this program seeks to educate children and youth with disabilities for part or all the time in regular schools. It includes providing of additional economic resources to the schools so that they can be used in the hiring of human resources, training of personnel, acquiring material resources, and in the generation of collaborative spaces to meet these students' special educational needs.

⁸Law N° 20.529 (2011). Creates the SAC, Quality Assurance System (QAS).

1.2 Analysis and Dissemination of ILSA Studies Data and Results

The main results produced by ILSA studies are widely disseminated in the country by the National Agency for Educational Quality. Firstly, the data are released in press conferences with media coverage, the same day that the international report is presented to the world by the institution conducting the study, immediately after that the embargo is finished. After that occurs, the publication of national results reports with specific analyses are carried out, and seminars and workshops oriented fundamentally to teachers, principals, education faculties, present and explain the results, and train on the assessment methods developed by these international projects.

The National Agency for Educational Quality maintains a website in which each of the ILSAs is presented, with their specific characteristics: the assessed domain, target population, the general project design, and a series of materials offered to schools, the community-academic, and the general public. Among these materials are released instruments, like assessment frameworks, questionnaires, and test items. The international results reports are also included, as well as the national reports and any thematic reports developed in the Agency.

However, the possibility to conduct in-depth research with data from the ILSA studies is quite limited for the National Agency for Educational Quality. For this reason, it is widely promoted the development of secondary studies by researchers and academics. The ILSA studies' databases with the manuals related to their management are made available to researchers and the community-academic. Agency organizes practical workshops where experts from the team train participants on statistical analysis that can be carried out. The new approaches and discoveries are fundamentals to enable the results to be used for public policy.

This emphasis on technical support and the required training was initiated with great vigor in the second PISA cycle where Chile participated (2006), to develop skills and make this technical knowledge available to research centers in the country. In fact, for the results of the PISA 2006, researchers from different centers were summoned, coordinated, and supported by a technical secretariat established in the Curriculum and Evaluation Unit of the Ministry of Education. With the contribution of an editorial committee composed of national experts, a first volume containing 11 articles was published. It was a selection that presented in-depth analyses and principal findings and lessons for educational policy from PISA 2006 (Ministerio de Educación de Chile 2009).

In 2012, the call was made directly by the Center for Studies of the Ministry of Education, within the frame of the "FONIDE DATOS PISA 2009 Extraordinary call", to encourage the use of the large amount of information provided by the study and development of research aimed at various topics of which PISA delivers data. In this call, the Fund for Research and Development in Education (FONIDE) funded nine projects, each of which focused its questions on different educational process areas using PISA data (Centro de Estudios 2012).

After 2012, several workshops of database analyses have been carried out. The manuals with instructions together with all the materials and information available from ILSA studies is made available for educational communities, researchers, and policymakers for their use in the design of studies, projects, strategies, and initiatives that contribute to the improvement of the quality of the education that Chilean children and youth are receiving.

2 Impact of ILSA on Chilean's Educational Policies

Since the 70s, Chile has built a long history of participation in ILSA studies, covering various subjects and grades. Mostly led by the International Association for the Evaluation of Educational Achievement (IEA), the Organization for Economic Cooperation and Development (OECD), and the United Nations Educational, Scientific and Cultural Organization (UNESCO), these studies have been providing information to the Chilean education system related to mathematics, natural sciences, reading literacy, financial literacy, civic education, writing, computer literacy, among others. Chile's systematic participation's fundamental purpose is to acquire knowledge and international perspective, otherwise not available, to better guide systems, institutions, and practices, deemed of strategic importance to the country's developmental goals (Cox and Meckes 2016).

Participation in ILSA studies has allowed the country to have relevant information to monitor the education system, the current curricula, public policies in education, and the programs that have been implemented, and incorporate international standards into national assessments and study frameworks. This participation also allows the country to compare Chile's results regarding other participating countries that consistently obtain good results (Agencia de Calidad de la Educación 2019a).

The educational national evaluation guidelines are governed by the current National and international assessment Plan that allows projecting medium and long-term efforts to review and design educational policies. The plan reflects national agreements on how and what to evaluate, and the possibility of complement national and international evaluations regarding evaluation frameworks, educational context, subjects of assessment, educational policies, school management, and pedagogical practices. (Decreto N° 183 2016). The current plan determines that Chile is part of the development of Programme for International Student Assessment (PISA),⁹ Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), International Computer and Information Literacy

⁹Developed by the Organisation for Economic Co-operation and Development (OECD).

Study (ICILS), International Civic and Citizenship Education Study (ICCS),¹⁰ and Regional Comparative and Explanatory Study (ERCE).¹¹

For the year 2021, a new National and international assessment Plan will approve national and international evaluation guidelines for years 2021–2025.

Over the past 20 years, ILSA studies have contributed to Chile's education policy by delivering information for decision-making at different levels. Mainly at: (a) Curricular adjustments and reforms designed by experts, (b) Educational system and its regulations, and (c) National assessment system.

2.1 Curricular Adjustments and Reforms Designed by Experts

In 1988, before Chilean participation in modern international studies,¹² the Unit of Curriculum and Evaluation of the Ministry of Education started to collect information about the level of students' knowledge through the national standardized assessment named SIMCE. This assessment measures achievement of fundamental curricular objectives and minimum compulsory contents in Language, Mathematics, Natural Science and Social Sciences. However, through this national assessment, it was not possible to contextualize students' learning regarding students' achievements in other countries or to analyze the national curriculum, teacher training, or pedagogical activities regarding other educational systems.

The participation of Chile in ILSA studies revealed the challenges faced by the national school system regarding the improvement of student learning in different subjects. Despite observing good results in comparison to other Latin American countries, the distance to the average performance of all countries participating in these studies is considerable. These results, along with the content and cognitive domains of different subjects, have been used to support and nourish the curricular revisions and reforms.

The curricular adjustment of 2009 was the first extensive review and update of the curriculum for primary and secondary education since the late '90s.¹³ Among other documentation and inputs (social demands started by secondary students, curriculum analysis, studies of relevance, surveys, revision of other countries' curricula, public consultations), the ILSA studies available up to that time were considered. This adjustment explicitly incorporated the information, both results and

¹⁰All these studies are developed by the International Association for the Evaluation of Educational Achievement (IEA).

¹¹Developed by the United Nations Educational, Scientific and Cultural Organization (Unesco) through its Latin American assessment laboratory (LLECE).

¹²Chile took part in the IEA's Six Subject Survey in the late 60s.

¹³New frameworks for primary and secondary education were promulgated in 1996 and 1998, respectively, and the study programs derived from them were gradually applied between 1997 and 2002. These frameworks were partially updated in 2002 and 2005.

frameworks, obtained in TIMSS for Mathematics and Science, PISA for Language and Communication, and ICCS for Citizen training (Cox and Meckes 2016).

Learning objectives that were not part of the curricula then were identified and integrated. For example, “Earth sciences” was added to the contents of primary and secondary education in Natural Science, and civic contents related to formal political participation and relationships with the political system were added to the History and Social Sciences curricula. (Cox and Meckes 2016).

The 2009 National adjusted Curricula were understood as a curricular framework and from other instruments in which it was possible to address them. These instruments, with different purposes, were oriented to the achievement of the learning defined by the curricular frameworks.¹⁴

The analysis of international evidence related to higher achiever countries, weakness in Chilean students’ training, and topics emphasized in the international frameworks had allowed national curricula developers to establish requirements and sequence of the learning objectives for the subjects covered by these studies. Consequently, the Learning Progress Maps were developed in 2007. They described the sequence in which a given competence, within the different curricular sectors, is typically developed throughout the school career (12 years), based on the learning opportunities prescribed in the curricular frameworks. Their purpose was to support teachers in the process of observing, analyzing, and monitoring the learning of their students (Ministerio de Educación de Chile 2007). Learning Progress Maps were replaced by Progression Of Learning Objectives, which have similar purposes and were also developed for each curricular sector, per grades.¹⁵

Performance levels of achievement were incorporated from the experience gain in international studies participation. These performance levels detail descriptions of what students know and are able to perform related to their performance in the national assessment SIMCE. From this information, qualitative information about students’ performance is delivered to schools to allow them to identify weaknesses of their students’ learning. Learning Progress Maps explained above and performance levels were developed using the reference frameworks from TIMSS, CIVED and PISA studies applied between 1998 and 2009.

From 2012 to 2019, new processes of curricular updates were developed affecting primary, secondary, vocational, and early childhood education. Especially for primary and secondary education, ILSA studies results were explicitly recognized and documented as an important source. An example of this is the mention made in the curriculum modification decree for 7th to 10th grades about learning outcomes and assessment frameworks, mentioning that this data allows matching the requirements of the national curricula with international requirements in different subjects (Decreto N° 614 2014).

¹⁴Study plans, study programs, progress maps, SIMCE performance levels, and school texts.

¹⁵Example for Language and Communication 1st to 6th grade of primary education. Progresión de objetivos de aprendizaje Lenguaje y comunicación. En: https://www.curriculumnacional.cl/614/articulos-71255_archivo_01.pdf.

In the case of Curricular Bases for grades 1–6, the revision of International assessments of learning applied in Chile (TIMSS, PISA, PIRLS, ICCS) and their assessment frameworks have allowed having comparative information to make decisions about the topics to be covered in each course, and the sequences of content and skills. (Ministerio de Educación de Chile 2018).

In the case of Curricular Bases for grades 7–10, the ILSA studies are widely mentioned, but not related to specific subject topics. Instead, it is indicated in a general way whether the framework or the study's results (report) of a particular cycle were used to develop and revise the subject. For instance, in Language and Literature, they mentioned the PISA 2009 assessment framework and a document with Reading task samples published in Chile (Ministerio de Educación de Chile 2011). In the case of Mathematics, PISA 2003 and 2012 assessment framework, together with PISA 2009 and TIMSS 2011 national report are mentioned as sources. In the case of Natural Science, assessment framework 2009 and TIMSS 2011, together with PISA 2006 international report, were mentioned as sources (Ministerio de Educación de Chile 2015a).

2.2 *The Educational System and Its Regulations*

Information from Chile's participation in international studies has been broadly used as input for evidence-based decisions to revise, propose, and adjust educational policies and practices to improve the school system. Data collected through these studies have been used as a reference in law discussions and adjustments regarding the school system's organization and financing.

Besides, international studies have been extensively used during discussion and design of laws dealing specifically with subjects assessed by some of these studies. For example, the Citizen Education Law¹⁶ was inspired by the poor results about Chilean students' civic knowledge, as shown by the ICCS 2009 study. With this base, on May 15, 2015, the President argued for the need for legislation to mandate every school in the country to define and implement a plan for citizenship formation (Cox and Meckes 2016), and during that year, the chamber of deputies submitted a bill of law to establish the obligation of every school to have a citizenship education plan. ICCS 2009 conclusions were presented as background and proof of this subject's lack of presence during school education (Biblioteca del Congreso Nacional de Chile (2018)). Although citizenship education was already part of the primary and secondary education curricula as a transversal learning objective, the need to strengthen this area was highlighted. Several professionals with knowledge in the subject used ICCS data through the process of generation of the law that was approved in 2016 (Ley N° 20.911 2016). This law established the obligation for schools to

¹⁶Law 20.911 (2016), creates a citizen education plan for educational institutions recognized by the state. It establishes that preschool, primary and secondary education must have a Citizen Training Plan.

define and implement the plan for citizenship. It stated the obligation for the Ministry of Education to promote the incorporation of a compulsory subject of Citizenship Education for the 3rd and 4th grades of secondary education. This new subject of Citizenship education for these grades was approved by the National Council of Education in February 2018 and begun to run in March 2019.

The decision to incorporate Financial Education as an explicit subject into secondary education's national curricula was based on the results of financial literacy from PISA 2015. During law discussions, members of the National Congress, Ministers of Education, and other experts use international studies data to reinforce their arguments. The most common use of these references is to account for the school system challenges in light of the results obtained compared to other school systems. For example, based on international and national PISA 2015 reports, the Commission of Education of the Chamber of Deputies prepared a detailed document with PISA 2015 Financial Literacy results of Chilean 15-year-old students, for the legislative deliberation, and its particular requirements and deadlines (Biblioteca del Congreso Nacional de Chile 2017). The law was promulgated in 2018¹⁷ (Ley N° 21.092 2018). Among others, the relevance that the OECD gives to financial education in PISA was one of the main arguments for this modification. The PISA assessment framework was also considered to define the main learning objectives to be highlighted in the national curricula.

Beyond Chile's need to improve learning outcomes, ILSA studies highlight school system inequities, allowing for discussions regarding the need for system reforms focused on vulnerable groups.

On the one hand, the international comparison has shown that performance differences by socioeconomic and cultural levels exist in all participating countries, but the degree of association between socioeconomic origin and school results varies considerably in different systems (Sandoval-Hernandez and Castejon 2014). From this, it follows that it is possible to develop a more equitable school system, an idea that was presented as a reference for the discussion of the educational inclusion law¹⁸ (Cox and Meckes 2016).

On the other hand, gender gaps have been brought to attention. Traditionally, female students in Chile obtained worse results in mathematics and science assessments, limiting their participation in STEM careers that are the ones with a better salary in the work market. Results from TIMSS and PISA showed that such a gender gap is not common to all countries, and of course, the difference cannot be attributed to innate ease for men in learning those subjects. To move towards quality and inclusive public education, in 2014, the Ministry of Education created the Gender Unit (UEG), a structure that is responsible for promoting the incorporation of a gender perspective in the Ministry's plans. The main goal is building a non-sexist education where everyone's capacities are recognized regardless of sex, identity, and gender. In

¹⁷Law N° 21.092 (2018). Modifies the LGE to include financial literacy contents in secondary education.

¹⁸Law N° 20.845 (2016). School inclusion law. It regulates students' admission, eliminates shared financing and prohibits profit in educational institutions that receive contributions from the state.

that framework, the Ministry established in 2015 a plan for Gender Equality 2015–2018 that made a diagnosis using a series of educational and labor data, including international assessment studies. It proposed a series of measures to incorporate the issue of gender and the need to work in an integrated and synergistic way with different ministries (Ministerio de Educación de Chile 2015b). In February 2019, the Ministry of Education and the Ministry of Women and Gender Equity signed an agreement that includes concrete measures to continue and deepen the initiatives that seek to eliminate gender biases and stereotypes in classrooms and grant equal educational opportunities to women and men.

The opportunity given by these studies to obtain data from the international comparison allowed the country to understand that the inequities which could be considered structural or immovable until that time are not natural. Moreover, information about successful countries related to the management of these challenges encourages the observation of practices replicable in our school system. An example of this is the targeting of resources to vulnerable groups. Preferential grants are delivered to the schools where most vulnerable students attend to mitigate social inequalities and improve school experience. The comparative experience served as an inspiration for these educational policies, for example, the Preferential School Subsidy Law,¹⁹ and its extension and update (Villarroel 2019).

The results of reading in PISA for Chilean students in 2000 and the following years (2009 and 2012), national reading tests, among other inputs, have been used to show the urgency of improving Chilean's reading skills and the need to face the problem with national policy. To date, two National Reading and Book policies have been defined and implemented in the country,²⁰ together with plans to promote reading, composed by a series of initiatives for developing these habits since childhood. Currently, the National Reading Plan 2015–2020 is in place, supported by a large number of government and private entities, which seek to “promote the formation of a society of readers, in which reading is valued as an instrument that allows people to improve their educational level, develop their creativity, sensitivity and critical thinking” (Gobierno de Chile 2015).

2.3 National Assessment System

Regarding the relationship between ILSA studies and Chile's national assessment system, one of the most important goals has been validating through international comparison the national assessment system itself, its methods, approaches, and results, establishing coincidences between similar results, seeking explanations when trends have been different.

¹⁹Law 20.248. Enacted in 2008, updated in 2011, 2016 and 2019.

²⁰Política Nacional del Libro y la Lectura 2006–2010 y Política Nacional de la Lectura y el Libro 2015–2020, both published by the National Council of Culture and Arts.

To improve the national assessment system, not only international evaluation frameworks have been used as a reference to learning objectives but also specific tools for updating and refine other aspects related to item development, sampling, and technical requirements for trend analysis, statistical methodologies, test score estimation, test formats, innovative subjects, among others.

There are countless examples where international studies have been reviewed as a reference to check, revise, or innovate different aspects of the national assessment system. To mention some of them:

- (I) Psychometric technical aspects. From the Classical Test Theory to Item Response Theory (IRT). Since its origin in 1988, SIMCE, the national assessment system, used classical test theory as the main measurement model for data analysis. However, since 1998 a transition process into IRT started, having as principal reference the international studies test scoring experience. (Agencia de Calidad de la Educación 2012). All cognitive tests are currently calibrated and scaled using IRT methodology and progress is also being made in using these models for scoring the context questionnaires' items.
- (II) Inspired by innovative assessments carried out by the international studies, some domains have been raised as initiatives of national interest:
 - PISA Financial literacy led to the generation of a national interdisciplinary workgroup composed of financial, education, and public policy institutions to think of a national financial education plan. Likewise, within the process initiated with integrating into the national curricula of financial education, some materials, courses, and training were developed by governmental institutions. PISA financial literacy study is used in most initiatives as reference.
 - International Computer and Information Literacy Study (ICILS) and International Civic and Citizenship Education Study (ICCS) led to national assessment initiatives²¹ developed to collect national information based on the international studies' frameworks and procedures.
- (III) The development of the Quality and Context of Education Questionnaires related to the national assessment tests has been influenced by the student, parent, teacher, and principal questionnaires from different international studies. Inspiring has been the used question formats, contents addressed, and the analysis methodologies that are constantly quoted as a reference.
- (IV) ILSA studies also have produced in the national assessment system a valuable knowledge in the technical teams regarding items construction, the inclusion of open-ended response items, and the development of coding guides and coding procedures (Cariola et al. 2011).

²¹Link to Citizen national assessment <https://www.agenciaeducacion.cl/evaluaciones/estudios-nacionales/>.

Link to Computational national assessment <https://www.enlaces.cl/evaluacion-de-habilidades-tic/simce-tic/presentacion/>.

3 Students' Movement, a Fundamental Actor in the Generation of Changes in the Chilean Educational System

This section is a general synthesis of the social movements in which Chilean secondary school students have been involved since 2006. These movements have been autonomous, that is, they were not summoned by political parties, and we dare to say that they have given rise and have accompanied the entire reforms process and attempts to improve the Chilean educational system. Contextual elements are offered to the readers to show how a large majority of students have developed their school careers in the last 14 years in Chile. On the one hand, authorities in the country discussed, redefined, and implemented policies related to various aspects of the educational reality. On the other hand, students actively participated or were spectators on the front line in massive movements in which they left classes for long periods to march, occupy the schools, and hold discussion forums, workshops, and other training and development instances. The process of making up the missing classes was usually very demanding. Students had to face extensions of the class calendar and condensation of tests and exams to achieve certification and promotion. Some schools, besides, participated in international assessments administered in the period. Although it is possible to say that there have been students' activities and demonstrations every year, two of them are the most relevant because of their high call and broad participation or in the concrete effects they produced, they are the first, in 2006 and the longest, in 2011. We also include 2012, 2015 and 2018 to complete the picture, and because PISA study or other ILSAs were administered then.

These movements bring together students who attend mostly municipal and some subsidized schools, starting in Santiago, the capital city, but extending later to the regions in the North and South. The students who attend private schools and represent almost 8% of enrollment in the educational system have little or no participation. Among the most active participants, there is a group of schools called "emblematic". They are public schools, free of charge, oriented to academic excellence, with long tradition and prestige (Rivera and Guevara 2017). Traditionally they have been highly selective, but selection practices must be eradicated after the Inclusion Law is implemented. Emblematic schools were targeted at men or women separately, but this feature is slowly changing. They are among the best municipal lyceums in the country, and although in the last years their achievements have decreased, their students get very good results in the SIMCE tests and selection tests for the universities. They promote civic and republican values, and their students are usually active participants in the movements.

Even if not all the students or schools have actively participated, social movements have undoubtedly affected the way in which the daily activities in the schools were developed in the country and has implied modifications in relation to the teaching-learning processes, the school climate, the relationship between students and teachers, and school authorities. This situation has also generated changes in

the attitudes and public behaviors of students as well as the perception that general citizens have of them. Interesting studies on these topics have been developed, but they are not presented here.

At the end of *April 2006*, the first uprising of secondary students (*Revolución Pingüina*) took place. It was the first national and massive social mobilization since the recovery of democracy (García-Huidobro 2009). A center-left wing coalition led the national government. The students started to require local and specific demands (school pass for public transport and the elimination of the payment for the university selection test), but soon more in depth and more cross-cutting themes appeared, such as the defense of the right to education, the improvement of public education quality, the end of municipalization, a rejection of the privatization of education. The students' objective was to repeal the Constitutional Law of Education (LOCE), the legal foundation of the educational system enacted by the Pinochet regime in 1990 (Bellei and Cabalin 2013), March 10, one day before the new democratically elected government assumed power. The movement included the paralyzation of activities in schools, massive street demonstrations, national school strikes, school occupations, and a strong presence in the press, getting important support from diverse actors and sectors of the society. The movement, which was later joined by university students, was active between April and June and then resumed in September and October, around five months of the school year's ten months. The PISA 2006 test was administered between August 21 and September 7, and SERCE was administered between October 16 and 20.

As a consequence of this movement, the government convened a Presidential Advisory Commission for Education. With more than 80 members, this commission met throughout the second half of 2006 to propose to President in December a report that included a series of proposals to improve the education quality. Among several others, the first proposal was to replace the current LOCE with a new law that could give legitimacy to the educational institution and guarantees the right to good quality education (Torres 2003). The General Education Law N°20,370 was promulgated on August 17, 2009.

After minor episodes of demonstrations and other events in the country in the intermediate years,²² the students' mobilizations emerged strongly again in *April 2011* and continued throughout the year. A coalition of right-wing parties led the national government. Started this time by university students, the movement questioned the root of the education market's general model that has produced enormous inequalities among the population and significant indebtedness because the persons who access university training have to apply for a loan to finance it. Secondary school students also joined the movement with their specific demands. However, the essential general demands were: "No profit", the obligation of the state to guarantee a free and good quality public education in secondary and university level, and the end of

²²On February 27, an earthquake and subsequent tsunami that struck central and southern Chile, with an intensity of 8.8 degrees on the Richter scale. The fatal victims reached a total of 525, nearly 500 thousand homes suffered serious damage and an estimated 2 million people were affected. It was the worst natural tragedy experienced in Chile since 1960.

public education administered by the municipalities (Muñoz and Durán 2019). The first demonstrations began on May 12, and in December, many students were still mobilized, some of whom had to repeat the school year. As in 2006, the movement consisted of massive marches, strikes, occupation of school buildings, very massive street demonstrations, a strong presence in the media, and it also included secondary students' hunger strikes (Arrue 2012). The movement managed to install the most important ideas raised by the students in the whole society, especially the critique towards a profit-based education system. It garnered much support from citizens, starting with parents, the general public, even among university authorities. National paralyzations called by unions with national representation, including teachers, to support the students occurred, and other groups with specific demands appeared in public life.²³ It is necessary to point out that there was much repression towards the students. To reject the harsh repression of the Carabineros to dissolve a student march in Santiago downtown, the Students' Federation of the University of Chile called on the public to show solidarity with the protesters by hitting the saucepans, which had been very popular more than 25 years before. The noise made by hitting pans and pots to support the students happened on August 4 night. This demonstration spread through various sectors, from the most popular and the middle classes, transforming itself into a new form of protest that accompanied the movement until its last massive activities and marches (Núñez 2012).

During the seven months of conflict, there were many attempts at dialogue between leader students and the government, with several government proposals, which were not accepted by the students. After three changes of ministers of Education, the year finished without clear solutions²⁴ (Taller de Análisis de Contexto—Varios autores 2012), but the students had become more than protesters in the streets: they became political actors and relevant players in the educational policy debate (Bellei and Cabalin 2013).

In 2012, the students' movement continued with marches in April and June and several strikes and occupations of school buildings in the middle of the year. PISA 2012 test was administered between August 20 and September 12.

In 2015, with a center-left coalition in the national government, a new educational reform was proposed, and its implementation started. The students' movement reorganized, and there were protests between April and June. The teachers' union called to an indefinite strike, which was extended from June 1 to July 27, in protest of the Teaching Career project proposed by the government. It implied that a big amount of students this year missed two months of classes. PISA 2015 test administration had to be delayed in Chile for that reason and was finally administrated between September

²³Mobilizations of poor people without access to housing and environmentalists, as well as the struggle of the Mapuche people and sexual minorities, among others.

²⁴On January 24, 2018, the National Congress approved one of the most important educational reform of the late years, the free education in higher education, understood as a benefit to which persons apply individually and accessed by students who belong to the lowest income quintiles of the country and are registered in higher education institutions attached to this system.

21 and October 10. In 2015, the National Agency for Educational Quality administered two other international assessment projects in Chile: PIRLS 2016, and ICCS 2016, both between October and November.

In 2018, between May and July, mostly university but also secondary students held a feminist student mobilization. The movement sought to denounce and punish cases of sexual harassment and abuse by teachers and demanded a process of social change to eradicate the prevailing machismo and the structural patriarchal system. The movement's demands include taking action against academics accused of sexual abuse, eliminating sexism from education, making changes to curricula, and training gender equality. The movement implied strikes, cultural activities, and occupation of schools and university buildings. The longest occupation lasted 74 days in the Law Faculty of Universidad de Chile, until July 9. PISA 2018 test was administered between August 20 and September 7. In 2018, the National Agency for Educational Quality administered two other international assessment projects in Chile: ICILS 2018 in September, and TIMSS 2019, between October and November.

Students were once again the protagonists of a movement developed between *October 18, 2019, and February 2020* in Chile. It was started by secondary school students who protested the rise in the ticket price of Santiago public transportation system and began to evade payment. The general public quickly and widely supported this action, transforming the movement from one focused on students, to one that concerned a large part of the country's population. The police force harshly repressed it. Trying to keep the country under control, the government decreed a state of emergency and curfew in some cities between October 19 and 28.

However, the movement spread in the form of pacific protests on the one hand, and violent demonstrations on the other, to most of the country's large cities. The main demands were related to the high cost of living, low pensions for retirees, social and economic inequity of Chilean society, criticism of the privatization of health services and housing, a general rejection of politicians and the armed forces. The need to have a new constitution was raised as a fundamental issue.

This extended movement was called "Chile woke up" and consisted of a series of activities of massive street demonstrations in strategic places in the cities, students and workers strikes and paralyzations, barricades, and traffic cuts on the streets, protests from the houses making noise by hitting pans and pots at certain times. Some groups' actions were also directed to destroy Santiago metro stations, public and private buildings, and looting of commercial establishments. National economic and social issues were the subject of conversation at all levels. Many people who had had a very passive attitude concerning their working life, and social, cultural, and political aspects began much more active participation.

There was a high degree of repression against the protesters that motivated that several international humanitarian entities sent observers and published reports about the high amount of human rights violations by agents of the Chilean state. The students' vacations and summer season slowed down the movement, but an increase was expected in March with the return to schools. However, the COVID-19 pandemic paralyzed the actions.

4 Main Marks of Educational Evolution Related to PISA

Chile has been participating in PISA since 2000, being part of every cycle so far, except for PISA 2003. Since PISA 2012, the country has also participated in innovative domains such as Financial Literacy, Global Competence, Problem Solving, Collaborative Problem Solving, and the Well Being questionnaire.

In Chile, PISA has raised questions covering both the performance on the main domains over time, and recognizing specific—not always positive—characteristics of the national school system. The country has been able to comprehend that although the Chilean students’ achievement is at the top within Latin America, it remains below the OECD mean performance in every assessed PISA domain.

15-year-old students of Chile have the best relative performance in reading, science, and mathematics among the countries of the Latin American region that participate in the project. Additionally, trend analyses of Chilean PISA results have shown some statistically significant improvement in reading literacy, which tends to flatten out in the last cycles, while performance in mathematics and science has remained stable (see Fig. 1).

For Chile, and for most participating countries, there were no statistically significant changes in students’ performance in 2018. With an average performance of 487 in reading, most of the OECD countries outperform Chile, which obtained an average of 452 in reading. As far as OECD members are concerned, Chile obtained the third-lowest performance in PISA 2018 in reading, mathematics, and science, only surpassing Mexico and Colombia (OECD 2019).

Some notable differences may be observed comparing Chile and other countries with a similar accumulated expenditure per student between 6 and 15 years of age. On the one hand, Chile obtained reading results similar to Greece, Malta, and the Slovak Republic, although all of them have a higher accumulated expenditure per

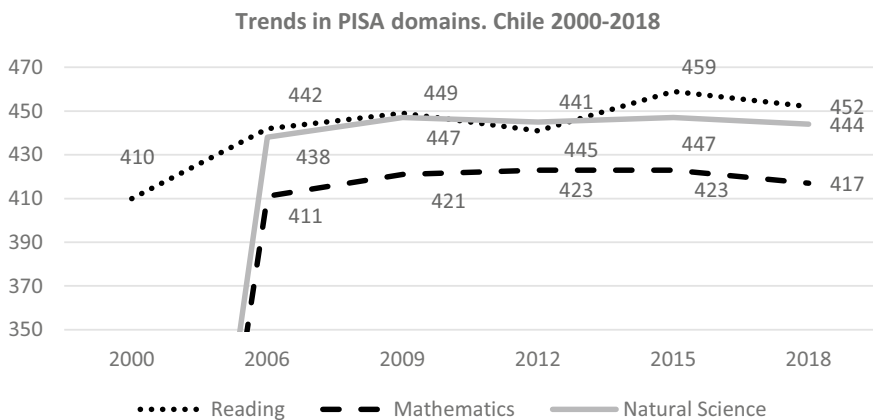


Fig. 1 Trends in PISA domains. Chile 2000–2018. *Source* PISA 2018 Results Volume I. Table I.B1.10 [3/4]; Table I.B1.11 [3/4] Table I.B1.12 [3/4]

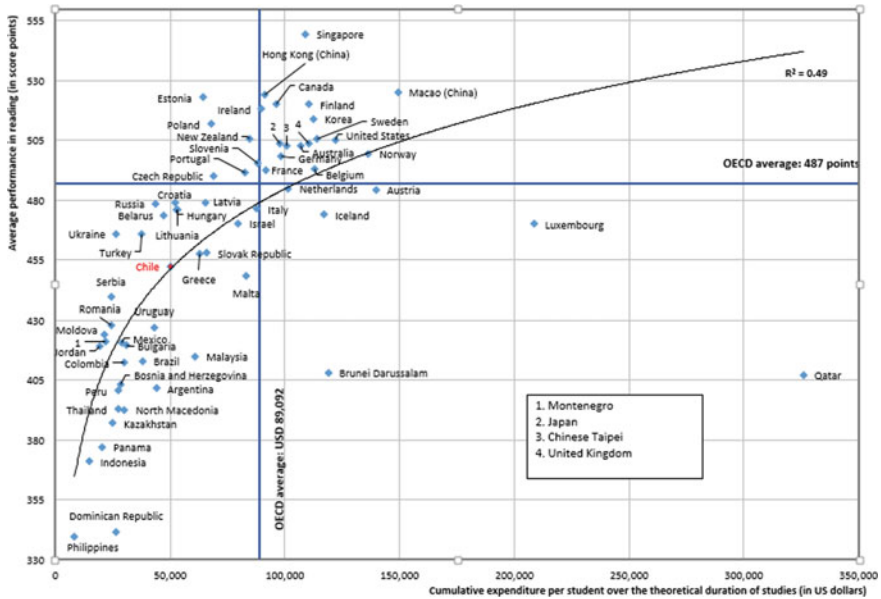


Fig. 2 Cumulative expenditure per student during studies (in US dollars). *Source* OECD, PISA 2018 Database, Tables I.B1.4 and B3.1.4

student. On the other hand, Ukraine, Turkey, and the Russian Federation have similar or lower accumulated expenditures per student but achieved better results than Chile (see Fig. 2).

The reading performance of students in Chile reaches the statistically expected value according to the education expenditure. On the contrary, for the other participating Latin American countries, students’ performance is below the expectation according to their spending.

Regarding the proficiency levels described by PISA, around one-third of students in Chile (31.7%) performed below Level 2 in reading at PISA 2018. PISA designates Level 2 as the base level of proficiency required to address reading-related issues demonstrating the capacity to use their reading skills to acquire knowledge and solve a wide range of problems. The proportion of students in Chile who obtained results below Level 2 was significantly higher than the OECD average of 22.7%. Chile also obtained a higher percentage than the OECD average of poor performance in science (35.3%). Most striking, over half of the 15-year-old Chilean youth obtained results below Level 2 in mathematics (51.9%).

PISA results show that Chile has difficulties in strengthening high-performance students who could help transform the country into a complex and knowledge-based economy later with their professional work. Only 2.6% of Chilean students obtained upper performance (levels 5 and 6) in reading, compared to the OECD average of 9.2%. Only 1.2% of students reached high mathematics performances compared to

the OECD average of 9.4%, and only 1% achieved these levels compared to 7% OECD average in science (OECD 2019).

PISA results show that the Chilean education system reflects substantial inequities of Chilean society. 15-year-old students in Chile show a bigger variation than what is observed in OECD countries average in the Economic Social and Cultural Status Index (ESCS) distribution. Chile's value is 1.03 and for the OECD average is 0.93, with 53 countries with less variation (more homogenous societies) and 23 with a greater variation in the index (more heterogeneous societies).

Although the effect of the socioeconomic and cultural status is very strong, Chile is not the country with the most significant effect of all. The strength of the relationship between ESCS and reading proficiency is expressed by the socioeconomic gradient, which refers to how well ESCS predicts the performance. In this indicator, Chile is quite close to the OECD countries average, with a value of 12.7 (OECD average is 12). There are 47 countries where the effect of ESCS is weaker than what it is observed in Chile, but also, there are 30 countries where the effect of ESCS is stronger than what it is observed in Chile. The countries with the strongest relationship between ESCS and reading performance show values between 18 and 21 and are Peru, Belarus, Hungary, Romania, and Philippines. On the contrary, the countries with the weakest relationship between ESCS and reading performance show values between 5 and 1.7 and are Montenegro, Hong Kong (China), Kosovo Republic, Baku (Azerbaijan), Kazakhstan, and Macao (China) (OECD 2019).

Specific results in Chile for main PISA domains are described below. Two important characteristics are identified within them, *Gender* gap and differences according to *Socioeconomic and cultural status*. Both will be addressed next.

4.1 Reading

Chilean students' performance in reading has significantly improved since the first cycle, with 7.1 points of average change, per a 3-years period, between 2000 and 2018 (OECD 2019). However, the trend is less positive nowadays. Comparing results from most recent PISA cycles to PISA 2000, Chile stands out as one of the best countries in the Latin American region even though, it remains below the OECD average (see Fig. 3).

The following graph summarizes the trajectory of 15-year-old students in Chile for almost two decades and shows that there have been improvements that, although they have not continued, have not been reversed either.

The percentage of students below Level 2 decreased significantly between 2000 and 2009 (from 48 to 31%), and after that this decline stopped. It is crucial to reactivate this decline because that implies advancing justice and integration into society since although these citizens can decode and read a text, their reading competence does not allow them to receive all the information they need to carry out entirely various tasks, to inform themselves, learn new things, or entertain themselves.

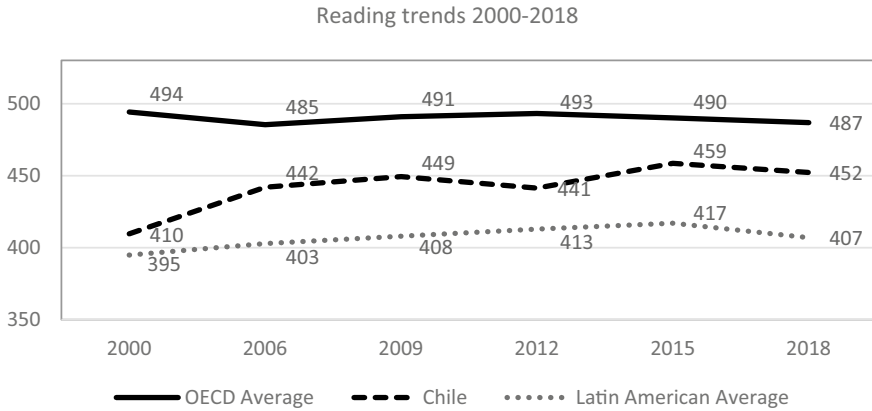


Fig. 3 Mean reading performance, 2000 through 2018. *Source* Developed by Agencia de Calidad de la Educación with PISA 2000–2018 International Databases

The percentage of students who barely reach the minimum skills to enter society successfully (Level 2) has remained constant in the period (with percentages ranging between 28 and 35%). However, there have been small changes in the higher levels, with an increase in the percentages of students reaching levels 3 and 4 and those who have developed more advanced reading skills.

Despite the stable overall performance, the proportion of Chilean students performing at Level 5 or above (top performers) in reading is significantly higher in 2018 regarding 2009 and 2012, with 1.3% and 2.0% respectively (OECD 2019). This trend must also be deepened; it will imply advantages for individuals and the whole society (see Fig. 4).

It is possible to explain, at least partially, the improvement observed in 2006 due to the curricula updating implementation. Most students who took the PISA test in 2000²⁵ were part of a population who had studied the primary education with the curricula defined and implemented by the military government, before the curricular reforms of 1996 and 1998. The reformed study programs were implemented gradually, starting in primary and then in secondary education. In fact, the students in 10th grade taking the PISA test in 2001 were the second generation that had studied that grade with the curricula established in 1998. On the contrary, students who took the test in 2006 were trained during all their educational careers with the curricula established in 1996 and 1998.

In the 2009 cycle, Chilean students showed an increase in the reading mean. Two reading tests (paper and digital) were administered in a group of countries at that cycle. Both tests were reported on the same scale and could be compared. In Chile’s case, students’ performances were significantly different, with a lower average in digital reading. Since PISA is a computer-based test since 2015, it could be expected

²⁵Chile participated in a group of 11 countries that replicated the study in 2001. Their results were expressed on the same scale and included in the report with all the countries, in 2000 and 2001.

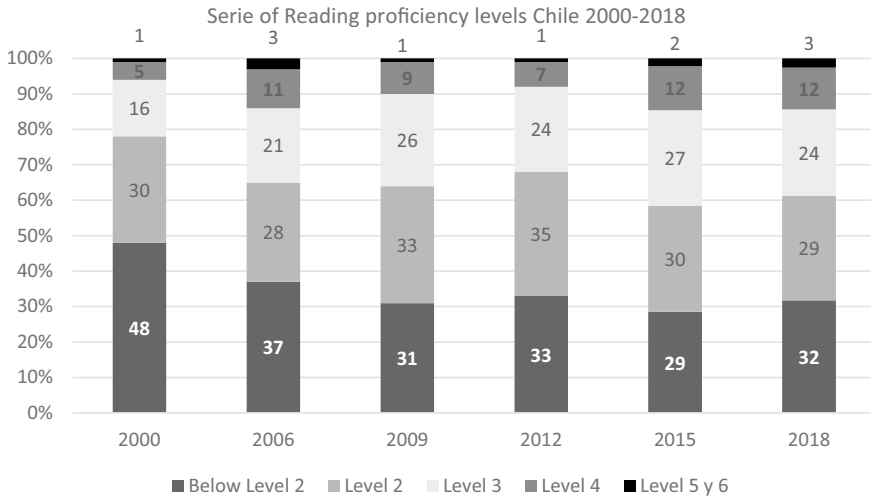


Fig. 4 Reading proficiency levels , Chile 2000–2018. *Source* Developed by Agencia de Calidad de la Educación with PISA 2000–2018 International Databases

that the reading mean in 2018 with a more robust scale because it is the main domain was lower or similar than the paper and pencil test and shows a smaller increase over time.²⁶

4.1.1 Gender Gap

As in all participating countries, in Chile, girls show higher reading competences than boys do. This trend is consistently maintained over time (see Fig. 3). But it is interesting to notice that Chile is among the eight countries with the narrowest gender gap in reading (less than 20 score points): Argentina, Colombia, Costa Rica, Mexico, Panama, and Peru; all of them are Latin American countries with low average and B-S-J-Z (China) with the highest average in the cycle (OECD 2019) (see Fig. 5).

4.1.2 Socioeconomic and Cultural Status Differences

In PISA 2018, there were no notable changes in reading competence in any socioeconomic and cultural quintile (PISA ESCS Index, divided into five groups) for Chilean 15-year-old students regarding 2015. Historically the country has shown relevant inequities in the educational achievements between different social, economic, and cultural origins. The gap between the most disadvantaged students’ performance

²⁶Cycle 2012, where the main domain was mathematics, showed a slightly strange behavior for the minor domains reading and science in Chile. For both areas, there was a decrease, even if it was not statistically significant.

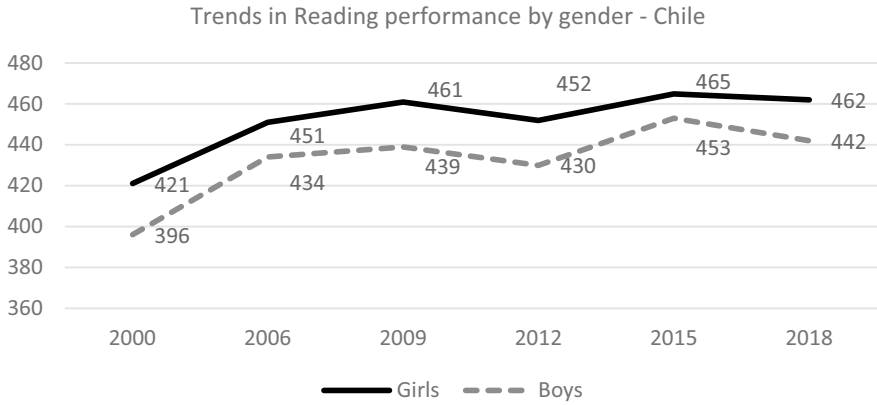


Fig. 5 Mean reading performance, 2000 through 2018 by gender. *Source* Developed by Agencia de Calidad de la Educación with PISA 2000–2018 International Database

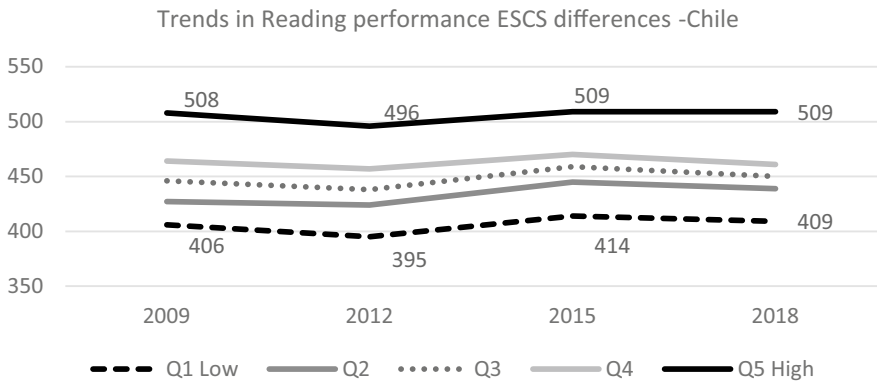


Fig. 6 Mean reading performance, 2009 through 2018 by socio-economic and cultural quintile. *Source* Developed by Agencia de Calidad de la Educación with PISA 2009–2018 International Database

(quintile 1) and the most advantaged students’ performance (quintile 5) has been constant.

However, these advantaged students’ performance is not exceptionally high. It exceeds the current OECD average in the last PISA cycle, but is far from the countries’ average with the best achievements.²⁷ On the contrary, it is possible to notice that quintile 2, the group with serious difficulties but maybe in the limit of extreme poverty, have improved consistently through these years (see Fig. 6).

²⁷Countries with the highest achievements in PISA 2018 Reading: BSJZ-China (555), Singapore (549), Macao-China (525), Hong Kong-China (524), Estonia (523), Finland (520), Canada (520).

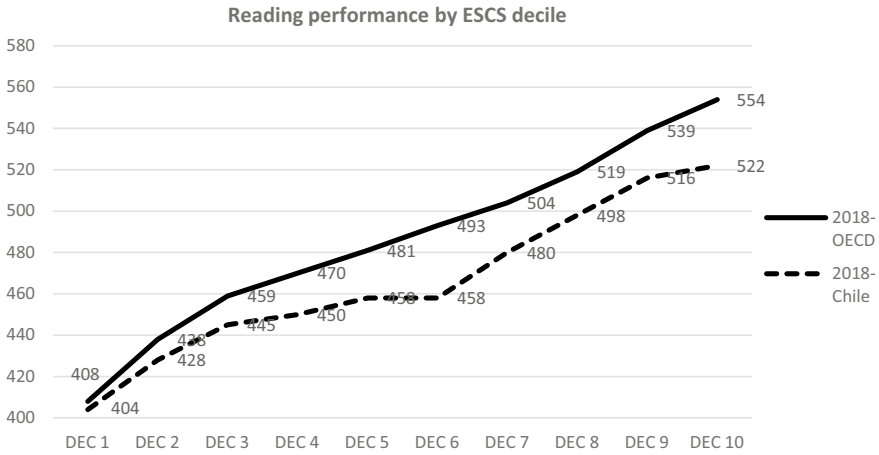


Fig. 7 Mean reading performance by Decile of socioeconomic and cultural status. *Source* Developed by Agencia de Calidad de la Educación with PISA 2018 International Database

On the national SIMCE reading test, it is possible to observe that the performance of the groups with the higher socioeconomic and cultural status in 10th grade²⁸ has gotten worse without stopping since 2012 (Agencia de Calidad de la Educación 2019b, page 30).

Similarly, it is remarkable that students currently belonging to the most socioeconomic and culturally disadvantaged group in Chile (ESCS decile 1) achieved similar performance to students of the OECD average who have similar conditions. On the contrary, students from Chile with higher socioeconomic and cultural resources (ESCS decile 10) perform significantly below their peers in the OECD (see Fig. 7).

This finding shows that although the Chilean educational system manages to produce small improvements in the most vulnerable sectors, has failed to improve the education quality, even in students with higher resources and more significant possibilities to develop their skills and achieve excellence levels.

This weakness was already identified at the beginning of the Chilean participation in PISA when it was clear that compared with students in the similar socioeconomic and cultural conditions in Latin America, Chile’s elite did not stand out. “This means that, even though these young people are the ones who probably get the best results in national assessments, their schools and families should not be satisfied” (Ministerio de Educación de Chile 2004).

4.1.3 Reading Performance Explanatory Model

The following exercise presents a multilevel analysis that is seeking to establish the relationship between a series of explanatory variables at the individual and school

²⁸10th grade is the national modal grade for 15-year-old students.

level and the average of Reading in 2018. The coefficients indicate how much the Reading average changes when the explanatory variables' value changes (see Fig. 8).

The most evident aspect that this graph presents is the importance of students go on their school career without being left behind: the most positive effect is produced by attending a secondary school (ISCED 3) when one is 15 years old, while the most negative effect comes from having repeated a grade.

At the school level, the average of the students' socioeconomic and cultural status has the most substantial effect on their performance. The families' socioeconomic and cultural characteristics are not modifiable by the school. However, it is possible to generate measures that reverse the current social segregation in the Chilean school system to tend towards greater integration of students from different backgrounds within the schools.

At the individual level, enjoying reading and having a positive self-image as a reader are shown to have positive effects on achievement. Instead, feeling that reading is difficult has a negative effect.

Good teaching practices, reported by students, about adapting the instruction to different students, stimulating their engagement with reading, and showing interest in them have a positive effect. On the contrary, teachers who are too directive in their teaching have a negative effect on their students' performance.

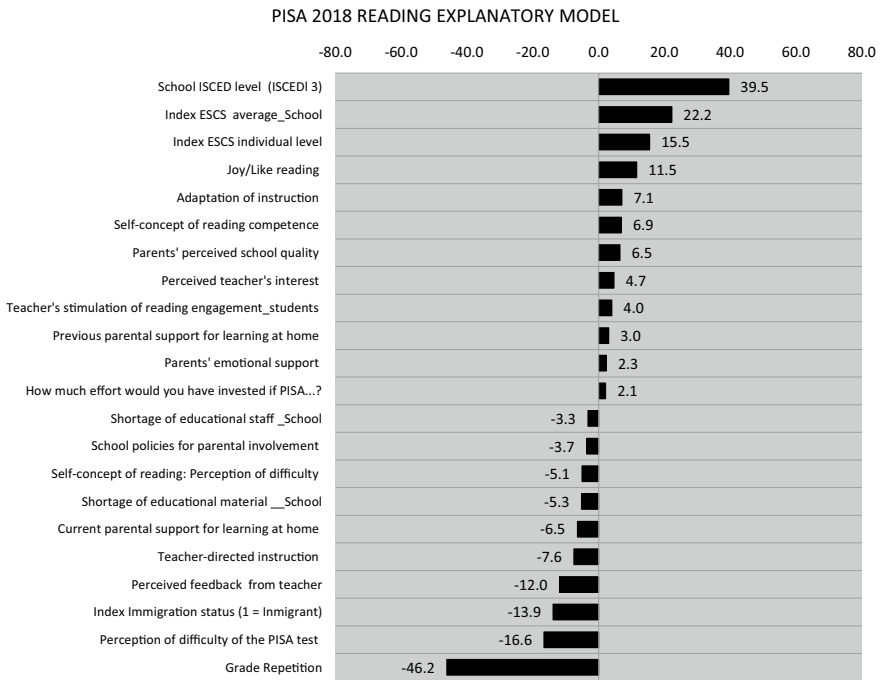


Fig. 8 Reading Explanatory model. *Source* Developed by Agencia de Calidad de la Educación with PISA 2018 International Database. Note: Intercept Value: 446

In turn, the parents’ emotional support and their perception that their children’s school provides quality education have positive effects on performance.

On the contrary, the lack of material and human resources in schools, reported by the principals, negatively affects students’ reading learning.

It is also observed that more feedback from teachers and strong support from parents at the age of 15 years have negative effects. This information is consistent because students with low achievements generally receive more attention from their parents and teachers.

Finally, the graph shows that being an immigrant in Chile at age 15 has a negative effect on reading performance. The educational system and school communities must integrate immigrant students to make them have the same learning opportunities as other students. They arrived to stay; they need to be prepared. It will mean gain for the country, its equity, integration, and population capacities development.

4.2 Mathematics

15-year-old students in Chile’s mathematics performance is significantly lower than the OECD average, although higher than the Latin American average. Through time, students in Chile obtain stable results and have not shown significant variations (see Fig. 9).

In all the PISA cycles, mathematics has proven to be the area in which Chile presents the most significant difficulty, with the highest percentage of students who don’t reach level 2, 51,9% (OECD 2019).

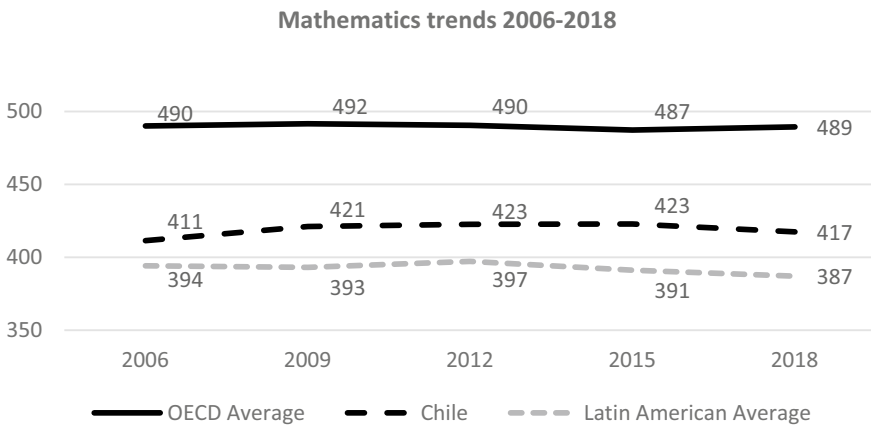


Fig. 9 Mean mathematics performance, 2006 through 2018. Source Developed by Agencia de Calidad de la Educación with PISA 2006–2018 International Database

4.2.1 Gender Gap

Despite the stereotype that boys are better than girls in mathematics, boys significantly outperformed girls in mathematics in only 32 of the 79 countries and economies that participated in PISA 2018, and Chile is one of them (OECD 2019).

Systematically, girls in Chile show lower mathematics performance than boys. However, girls have shown a trend towards stability. On the contrary, boys' scores behaved differently, and their performance fell significantly 11 points in 2018 compared to PISA 2015 (see Fig. 8). It is not clear, based on these PISA data, which can be the reason for this different behaviour. In any case, this information is consistent with the national test SIMCE where after some years of improvement for girls and boys of 10th grade, in 2014, boys showed a worsening in their performance meanwhile the girls showed stability. The situation has not changed lately (Agencia de Calidad de la Educación 2019b, page 35) (see Fig. 10).

It is not good news. Achieving gender equality in education is the first step to achieve a balanced society, but it is not a triumph that the difference is reduced because boys decline. It is essential that girls improve, but that boys also do.

4.3 Socioeconomic and Cultural Status Differences

None of the quintiles of different socio-economic and cultural status show significant changes in their math score between 2015 and 2018. However, the graph shows that quintile 2 has improved, and its trajectory of increasing is permanent (see Fig. 11). National SIMCE mathematic test results are consistent with this finding because,

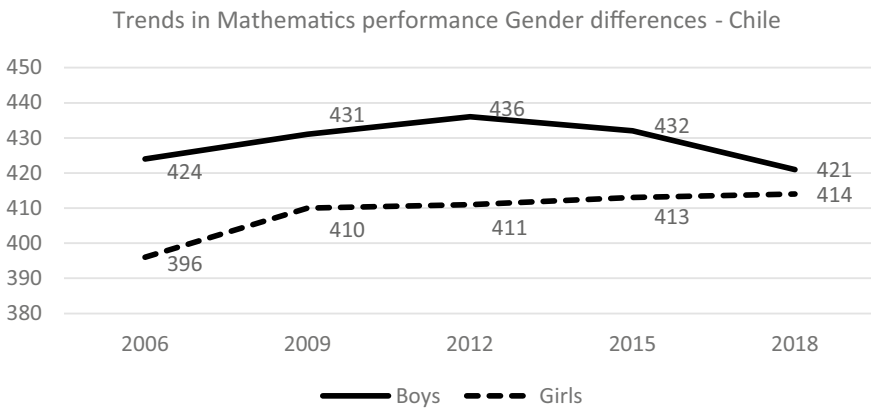


Fig. 10 Mean mathematics performance, 2006 through 2018 by gender. *Source* Developed by Agencia de Calidad de la Educación with PISA 2006–2018 International Database

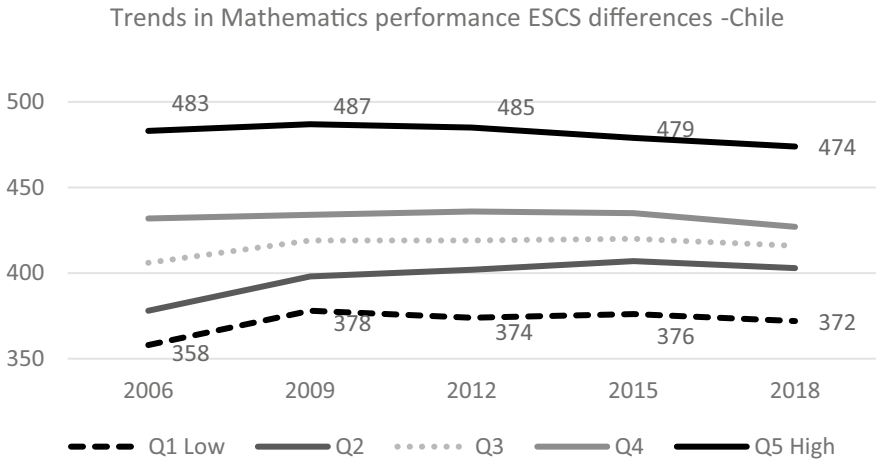


Fig. 11 Mean mathematics performance, 2006 through 2018 by socioeconomic and cultural quintile. *Source* Developed by Agencia de Calidad de la Educación with PISA 2006–2018 International Database

during the last decade, the gap between upper and lower groups has narrowed due to latter's progress (Agencia de Calidad de la Educación 2019b, page 37).

4.4 Natural Science

15-year-old students' performance in natural science has not shown significant variations in the long term. Despite the stable overall performance, the proportion of Chilean students performing at Level 5 or above (top performers) in science shrank in 0.9% between 2006 and 2018 (OECD 2019, page 284). It is significant, even if it is a small percentage.

Natural science results of students in Chilea are lower than the OECD average but higher than the Latin American average, and above the average of the participating countries in the region (see Fig. 12).

4.4.1 Gender Gap

In PISA 2009, girls managed to increase their scientific skills, but there have been no considerable changes since then. Boys have not shown changes through cycles.

PISA 2018 showed no significant differences in natural science by gender. Compared to previous cycles, more equity is observed, but this was due to a significant drop in boys' results and non to a significant increase in girls' performance (see Fig. 13), as also happened in mathematics. On the national SIMCE science test, it is

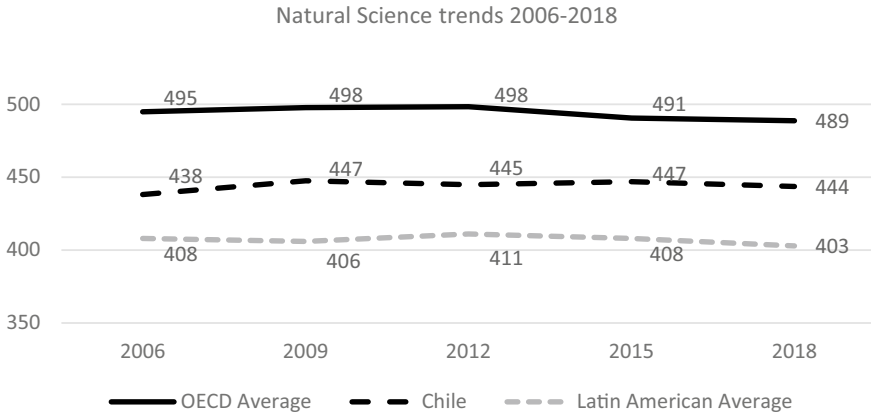


Fig. 12 Mean natural science performance, 2006 through 2018. *Source* Developed by Agencia de Calidad de la Educación with PISA 2006–2018 International Database

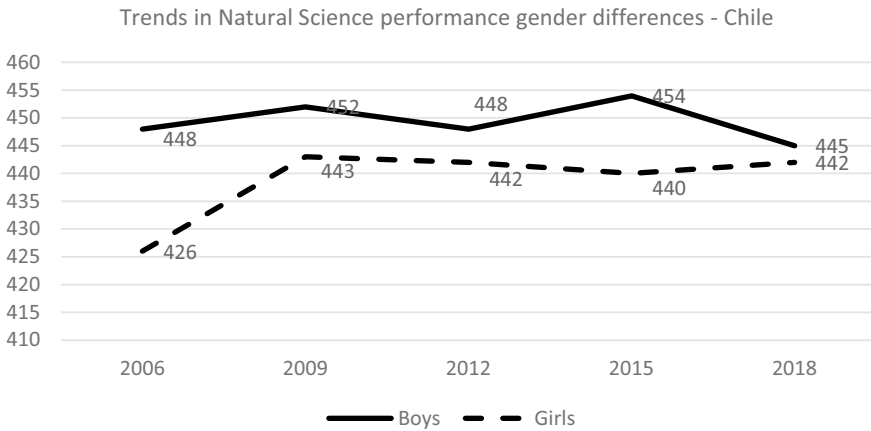


Fig. 13 Mean natural science performance, 2006 through 2018 by gender. *Source* Developed by Agencia de Calidad de la Educación with PISA 2006–2018 International Database

possible to observe the same trend; boys of Grade 10 have reduced their score since 2014, and girls remain stable (Agencia de Calidad de la Educación 2019b, page 40).

4.4.2 Socioeconomic and Cultural Status Differences

There are no significant scientific competence changes for any quintile of socioeconomic and cultural status in PISA 2018 regarding 2015. However, in the long term trend, it is possible to observe that while, the highest quintile is slowly reducing, the two lowest quintiles tend to increase their scores, especially the second, which shows

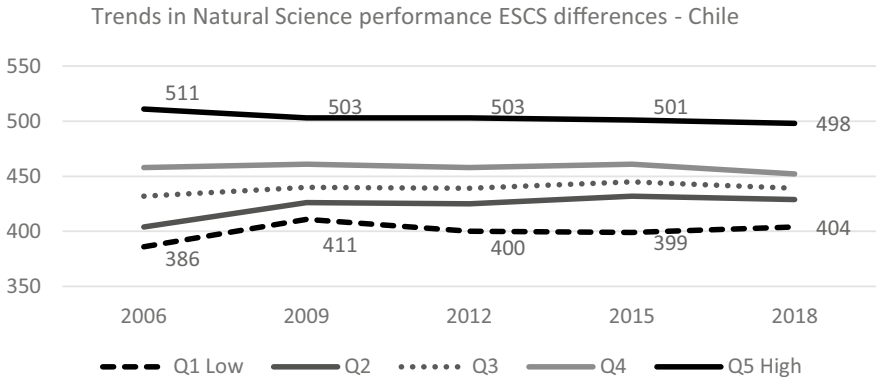


Fig. 14 Mean natural science performance, 2006 through 2018 by socioeconomic and cultural quintile. *Source* Developed by Agencia de Calidad de la Educación with PISA 2006–2018 International Database

a sustained improvement. This finding is consistent with what the national SIMCE science test shows: the most advantaged group has reduced their score since 2012; meanwhile, the two lowest quintiles have remained stable (Agencia de Calidad de la Educación 2019b, page 41) (see Fig. 14).

5 Conclusions and Recommendations

Since the beginning of the ‘90s, Chile is carrying out various educational reforms. From the end of that decade, data collected from international studies are fundamental pieces of the information available within the educational system to monitor its development and the achievement of its objectives.

For 20 years PISA has shown that the students’ performance in Chile has remained without much variation, especially in mathematics and natural sciences. In reading, improvements were observed between 2006 and 2009, but the situation has remained stable after that period. The few observed improvements correspond to the most socioeconomically and culturally disadvantaged students, which is very positive. It shows that the measures aimed at strengthening these groups and the schools that serve these students have made some progress.

Continuous curricular reforms have been made to adjust teaching to the changing demands of the times and the world globalization, but without a doubt, they are not enough to ensure that the majority learn and students who reach levels of excellence emerge.

The national context of recent years, related to social movements and educational adjustments, highlighted the need to make structural changes in the Chilean education system, both concerning the laws and regulations that govern its operation and its practices.

Objective data from the national evaluations periodically carried out by the SIMCE, related with no improvements in the students' achievement and significant differences between groups within the country, made evident the lack of quality of the Chilean educational system. ILSA studies have made a substantive contribution evidencing the system's weaknesses in the international comparison, highlighting high levels of gender and socioeconomic inequity, and the low quality of education received by large masses. Specifically, one of Chile's main concerns is the significant proportion of students of secondary education underperforming in all PISA domains. It is vital to focus efforts to mobilize students at least to Level 2 in all domains, the minimum threshold to be able to join the society.

PISA and other international studies have also shown the Chilean education system's inability to enhance the number of high-performing people who could help improve the country in different innovation areas. The country's efforts to improve low student performance include policies seeking to raise outcomes for those coming from less advantaged backgrounds, strengthening early childhood education, and early intervention mechanisms in case of difficulties. Policies should also include measures to promote excellence for all students and strengthen the students' performance at higher proficiency levels (OECD 2018).

Data show that efforts to deliver more resources to schools to serve and retain in the system, especially the most socio-economically and culturally disadvantaged students, have the effect of making them learn more and be more competent, which is a goal of justice and integration into society. More efforts and efficiency are lacking, but it is going in the right direction.

International studies have been used extensively to improve national curricula by incorporating the knowledge and requirements that are internationally recognized as necessary to face present and future challenges, both concerning preparation for working life and citizen participation. ILSA studies will continue to be a reference in the national Education System. This is how, for example, the financial literacy evaluation framework will undoubtedly be used for the implementation of the law that integrates this subject to the curriculum of III and IV secondary grades. It is important to point out that based on international data analysis, some suggestions can be drawn. For example, the exercise presented with the reading scores (see Fig. 8) shows some aspects that, in each particular situation, can be considered and modified with concrete actions from the national policy, educational institutions, and families. 1. Although the practice of repetition has decreased, it is still used in the country. Evidence showing its ineffectiveness and even worse results provide arguments for seeking alternatives. Early detection and remedial interventions should be the solution to support students facing learning disabilities. Legislation, together with teaching practices and management in schools, should aim at this objective. 2. The country needs to advance in the equity of the educational system. It is fundamental that all schools, without distinction, may have the necessary, high-quality human and material resources to carry out their tasks. The educational system has to provide the means to make this possible. There are huge expectations that the non-selection, the reinforcement of public education, and the end of co-payment in schools that receive state funds will eventually produce a situation of a similar offer of quality education

that families can access. This chain of facts will promote socio-economic integration that would reduce the segregation in schools, which replicates the society's existing. QAS must develop its mission, and all the public institutions work coordinately to promote and facilitate that the dispositions are fulfilled, and the goals can be accomplished. 3. Due to the high diversity of students in the same classrooms, teachers need to develop the capacity to adapt their instruction to different students, which is related to show interest in all of them. For that reason, it is necessary to train the teachers effectively in methods and approaches that allow them to keep and manifest faith in all their students' capacity. In that way, they will not consider useless to try different methods for students with difficulties. Teachers also need to be trained to stimulate the students, proposing them appealing challenges, and supporting their discoveries instead of being too directive in their teaching. 4. Given the enormous importance of the student's ability to enjoy reading and have a healthy self-image about his/her capacity as a reader, it is clear that the first to develop this feature in children are the parents. They can encourage the children to read since the first years of life, reading for them or accompanying them while they read. Having parents who read recreationally also fosters a love of reading. 5. Depending on their particular reality, the schools may generate free and recreational reading spaces, using the available resources, both printed and digital, that allow students to venture into their motivations and interests and thus develop a taste for reading. 6. Specifically related to reading again, the educational system should generate training instances for practicing teachers in reading didactics and new methodologies to encourage reading in children and young people. Of course, schools need to count on physical and digital materials to promote reading for different purposes, starting from the youngest students.

We must recognize that despite decades of efforts, the Chilean education system, in general, remains deficient. Proof of this is that many of the expected achievements have not been accomplished even after the policies' implementation, changes in the governments, and new reforms. Tensions persist in the system, as well as low students' results in national and international studies, with stagnant indicators for years, and unmet aspirations for vast sectors.

Social movements with students as protagonists could be carrying out a long-term cultural and social transformation. They contribute to the generation of new citizens concerned with transforming society and its model, reflected in the design of its public policies in general. This aspect is very positive for the students themselves and the democratic system, but it is also true that it can difficult them to learn and develop the competencies that schools must provide, which is a high cost.

The current state of affairs, with an active, participatory, and demanding citizenry, partially—at least physically—stilled by the planetary emergency brought about the COVID-19, has meant time for reflection, study and the preparation of strategies to apply after the health emergency when it is necessary to face also the deepening economic and social crisis.

Education in Chile is indebted to the country. It should and can improve, and it is widely expected that citizens and authorities can agree peacefully and in democratic channels the best ways to get it. International studies will continue to monitor

the teaching-learning processes. They will offer a comparison with similar and different educational systems, calling for reflection, searching for possible solutions and strategies for the identified problems. All these together will allow in the future—hopefully not too far away—education becomes a useful tool of personal development, satisfaction and social mobility. Then, the country will have more capacity to prepare all the people to decide their lives, reach their goals, become useful and committed citizens with their community to have a more equal, fair and balanced society.

Acknowledgements The author gratefully acknowledges the editor of this publication the opportunity to share with colleagues in other countries the challenging experience of implementing and using International Large-scale Students Assessment studies to improve the quality of the education that current and new generations receive in Chile.

Thanks to the National Agency for Educational Quality to support this work's development and María Victoria Martínez for her valuable contribution and respect.

Particular thanks to the participating schools and students all these decades in our studies, and for who will continue contributing to our effort to find and propose creative and efficient ways to ensure that each girl and boy receives the best education and has the opportunity to develop at her/his maximum capacity.

References

- Agencia de Calidad de la Educación. (2012). Informe técnico Simce 2012. Santiago, Chile. Retrieved from: https://archivos.agenciaeducacion.cl/documentos-web/Informe_Tecnico_Simce_2012.pdf
- Agencia de Calidad de la Educación. (2013). PISA: Evaluación de las competencias lectoras para el siglo XXI. Marco de evaluación y preguntas de la prueba. División de Estudios, Departamento de Estudios Internacionales, Santiago, Chile.
- Agencia de Calidad de la Educación. (2019a). Evaluaciones Nacionales e Internacionales de Aprendizaje período 2004–2016. División de Estudios, Santiago, Chile. Retrieved from: https://archivos.agenciaeducacion.cl/Panorama_Final.pdf
- Agencia de Calidad de la Educación. (2019b). Resultados Educativos 2018. 4° Educación Básica, 6° Educación Básica, II Educación Media, Retrieved from: https://archivos.agenciaeducacion.cl/Conferencia_EERR_2018.pdf
- Arrue, M. (2012). El movimiento estudiantil en Chile (2011–2012). Una lucha contra la discriminación. *Ameriqué Latine Histoire Et Mémoire*, 24, 77–88. Retrieved from: <https://journals.openedition.org/alhim/4388>
- Bellei C. & Cabalín, C. (2013). Chilean Student Movements: Sustained struggle to transform a Market-oriented Educational System. *Current Issues in Comparative Education*, 15(2), 108–123. Retrieved from: <https://files.eric.ed.gov/fulltext/EJ1016193.pdf>.
- Biblioteca del Congreso Nacional de Chile. (2016). Historia de la Ley N° 20.370. Establece la Ley General de Educación, Santiago, Chile. Retrieved from: <https://www.bcn.cl/obtienearchivo?id=recursoslegales/10221.3/3824/7/HL20370.pdf>.
- Biblioteca del Congreso Nacional de Chile. (2017). Resultados de la prueba PISA (OCDE) 2015 en Educación financiera. Retrieved from: https://www.bcn.cl/obtienearchivo?id=repositorio/10221/24925/1/BCN-Resultados_PISA_2015_educacion_financieraFINAL.pdf.
- Biblioteca del Congreso Nacional de Chile. (2018). Historia de la Ley N° 20.911. Crea el Plan de Formación Ciudadana para los Establecimientos Educativos reconocidos por el Estado. Retrieved from: <https://www.bcn.cl/historiadela/nc/historia-de-la-ley/4891/>

- Cariola, L., Covacevich, C., Gubler, J., Lagos, E. & Ortiz, M. (2011). Chilean Participation in IEA Studies. In 50 Years of Experiences and Memories, edited by C. Papanastasiou, T. Plomb, and E. Papanastasiou, 373–388. IEA 1958–2008. Nicosia: Cultural Center of the Kykkos Monastery.
- Centro de Estudios. (2012). *Evidencias para Políticas Públicas en Educación: Selección de Investigaciones Concurso Extraordinario FONIDE-PISA*. Santiago, Chile: Ministerio de Educación.
- Consejo Nacional de la Cultura y las Artes. (2006). Política Nacional del Libro y la Lectura. Santiago, Chile. Retrieved from: <https://www.cultura.gob.cl/wp-content/uploads/2012/03/Pol%C3%ADtica-Nacional-del-Libro-y-la-Lectura-2006.pdf>.
- Consejo Nacional de la Cultura y las Artes. (2015). Política Nacional la Lectura y el Libro. Santiago, Chile. Retrieved from: <https://www.cultura.gob.cl/wp-content/uploads/2014/08/politica-libro-lectura-2015-2020.pdf>.
- Cox, C., & Meckes, L. (2016). International large-scale assessment studies and educational policy-making in Chile: Contexts and dimensions of influence. *Research Papers in Education*, 31(5), 502–515. <https://doi.org/10.1080/02671522.2016.1225349>
- Decreto, N. 614. (2014). Establece bases curriculares de 7° año básico a 2° año medio en asignaturas que indica. Ministerio de Educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1059966&f=2015-01-01&p=>.
- Decreto, N. 183. (2016). Establece plan nacional de evaluaciones nacionales e internacionales para el periodo 2016–2020. Ministerio de Educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1094644&f=2016-09-10&p=>.
- García-Huidobro, J. (2009). Qué nos dejó la movilización de los pingüinos?, *Nomadías* 9, Santiago, Universidad de Chile, 2009, pp. 205–207.
- Gobierno de Chile. (2006). Informe Final de Consejo Asesor Presidencial para la calidad de la educación. Santiago, Chile. Retrieved from: <https://educacion2020.cl/documentos/informe-final-de-consejo-asesor-presidencial-para-la-calidad-de-la-educacion/>.
- Gobierno de Chile. (2015). Plan Nacional de la Lectura 2015–2020. Santiago, Chile. Retrieved from: <https://plandelectura.gob.cl/wp-content/uploads/2016/12/Plan-Nacional-Lectura-web-6-12-2016.pdf>.
- Ley, N° 20.370. (2009). Ley general de educación. Ministerio de educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/Navegar?idNorma=1006043&idVersion=2009-09-12>.
- Ley, N° 20.529. (2011). Sistema nacional de aseguramiento de la calidad de la educación parvularia, básica y media y su fiscalización. Ministerio de educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1028635&f=2012-12-11&p=>.
- Ley, N° 20.710. (2013). Reforma constitucional que establece la obligatoriedad del segundo nivel de transición y crea un sistema de financiamiento gratuito desde el nivel medio menor. Ministerio de Educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1057032&f=2013-12-11&p=>.
- Ley N° 20.845. (2016). De inclusión escolar que regula la admisión de los y las estudiantes. Elimina el financiamiento compartido y prohíbe el lucro en establecimientos educacionales que reciben aportes del estado. Ministerio de educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1078172&f=2016-03-01&p=>.
- Ley, N° 20.911. (2016). Crea el plan de formación ciudadana para los establecimientos educacionales reconocidos por el estado. Ministerio de Educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1088963&f=2016-03-03&p=>.
- Ley, N° 21.040. (2017). Crea el sistema de educación pública. Ministerio de educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1111237&f=2019-04-25&p=>.
- Ley, N° 21.092. (2018). Modifica la ley general de educación para incorporar en el nivel de enseñanza media contenidos de educación financiera básica. Ministerio de educación, Santiago, Chile. Retrieved from: <https://www.leychile.cl/N?i=1118807&f=2018-05-23&p=>.
- Ministerio de Educación de Chile. (2004). Competencias para la vida. Resultados de los estudiantes chilenos en el estudio PISA 2000. Unidad de Currículum y Evaluación, Santiago, LOM Ediciones.
- Ministerio de Educación de Chile. (2007). Orientaciones para el uso de los Mapas de Progreso del Aprendizaje. Unidad de Currículum y Evaluación, Santiago, Chile.

- Retrieved from https://archivos.agenciaeducacion.cl/biblioteca_digital_historica/orientacion/2007/orien_mapas_simce_2007.pdf.
- Ministerio de Educación de Chile. (2009). ¿Qué nos dice PISA sobre la educación de los jóvenes en Chile? Unidad de Currículum y Evaluación, SIMCE, Santiago, Gráfica 7 Ltda.
- Ministerio de Educación de Chile. (2015a). Plan de Evaluaciones nacionales e internacionales 2016–2020. Retrieved from: https://www.curriculumnacional.cl/614/articles-34980_recurso_1.pdf
- Ministerio de Educación de Chile. (2015b). Bases Curriculares 7° básico a 2° medio. Unidad de Currículum y Evaluación, Santiago. Retrieved from: https://www.curriculumnacional.cl/614/articles-37136_bases.pdf
- Ministerio de Educación de Chile. (2015c). Educación para la igualdad de género. Plan 2015–2018. Unidad de Género, Santiago. Retrieved from: <https://media.mineduc.cl/wp-content/uploads/sites/28/2017/07/Educaci%C3%B3n-para-la-Igualdad-de-G%C3%A9nero-2015-2018.pdf>
- Ministerio de Educación de Chile. (2018). Bases Curriculares Primero a Sexto Básico. Unidad de Currículum y Evaluación, Santiago. Retrieved from: https://www.curriculumnacional.cl/614/articles-22394_bases.pdf
- Muñoz V & Durán C (2019). Los jóvenes, la política y los movimientos estudiantiles en el Chile reciente. Ciclos sociopolíticos entre 1967 y 2017. *Izquierdas*, #45, pages 129–159. Retrieved from: <https://www.izquierdas.cl/images/pdf/2019/n45/art6.pdf>
- Núñez, D. (2012). Proyecciones políticas del movimiento social por la educación en Chile” en *OSAL*. Buenos Aires: CLACSO. Año XIII, N° 31, mayo. Retrieved from: <https://www.cetri.be/IMG/pdf/OSAL31.pdf>
- OECD. (2018). Education Policy Outlook 2018: Putting Student Learning at the Centre. OECD Publishing, Paris. Retrieved from: <https://doi.org/10.1787/9789264301528-en>
- OECD. (2019). PISA 2018 Results (Volume I): What Students Know and Can Do. OECD Publishing, Paris. Retrieved from: <https://doi.org/10.1787/5f07c754-en>
- Rivera M., & Guevara J. (2017). Selección en la educación secundaria y acceso a universidades de elite por tipos de establecimiento en Chile. Análisis de la cohorte 2016. Centro de Estudios División de Planificación y Presupuesto Subsecretaría de Educación Ministerio de Educación, Santiago, Chile.
- Sandoval-Hernandez, A., & Castejon, A. (2014). A Comparison of School Effectiveness Factors for Socially Advantaged and Disadvantaged Students in ten European Countries in TIMSS-2011. *Šolsko Polje*, letnik xxv (številka 3–4), 36. Retrieved from: <https://www.researchgate.net/publication/271511384>
- Taller de Análisis de Contexto—Varios autores. (2012). Hitos de un conflicto inconcluso. In: Rodríguez & Montealegre, J, (Eds). (2012). *Ciudadanía en marcha. Educación Superior y Movimiento Estudiantil 2011. Curso y Lecciones de un conflicto*. Santiago, Chile, Editorial Universidad de Santiago de Chile.
- Torres, R. (2013). Desigualdad socioeducativa y movilización estudiantil Emergencia del problema público educativo en el Chile del Bicentenario. *Crítica y Emancipación*, (10): 171–214, segundo semestre de 2013. Retrieved from: <https://movimientoestudiantil.cl/wp-content/uploads/2015/12/77-Desigualdad-socioeducativa-y-movilizacion-estudiantilemergencia-del-problema-publico-educativo-en-el-Chile-del-bicentenario-R.Torres.pdf>
- Villarroel, T. (2019). Proyecto de Ley que extiende y moderniza la Subvención Escolar Preferencial. Presentación para la Comisión Educación de la Cámara de Diputados. AcciónEducar. Retrieved from: <https://accioneducar.cl/wp-content/uploads/2019/12/Presentaci%C3%B3n-Proyecto-SEP-Acci%C3%B3n-Educar.pdf>

Emma Lagos has a degree in Sociology from the University of Chile. Emma worked for ten years developing political and social research projects in nongovernment research institutes and private companies (CERC and MORI), participating in the development of “Barómetro CERC”, surveys

that monitored the return and restoration of democracy in Chile in the early '90s, and "Latino-barómetro".

In the Ministry of Education, she was part of the team that implemented international educational evaluation projects in SIMCE (national evaluation system) in Curriculum and Evaluation Unit: Civic Education, TIMSS, PISA and SERCE, between 1999 and 2011.

Since 2012, Ema works in the Education Quality Agency as part of the International Studies team in the Studies Division (PISA, ICILS, PIRLS, TIMSS, ERCE, ICSS). She is the representative in OECD as PISA National Coordinator and has provided advice to the Ministry of Education of Paraguay in the implementation of PISA-D, convened by OECD and financed by IDB.

Ema has participated in the production of two books on international studies of educational evaluation: "IEA 1958-2008 50 years of experiences and memories," edited by Constantinos Papanastasiou, in 2011 and "Implementation of Large-Scale Education Assessment," edited by Petra Lietz, John Cresswell, Keith Rust, and Ray Adams, in 2017.

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.



England: England and PISA—The Long View



Tim Oates

Abstract This chapter does not drill down in the minutiae of the PISA results for England. For that, readers can go to the NFER’s excellent report (Sizmur in *Achievement of 15-year-olds in England: PISA 2018 Results DFE-RR961 2019*) which comprises the UK Government’s commentary on the PISA outcomes. Rather, it tries to do something unique—it places the PISA results in the context of policy changes which may be associated with PISA outcomes, and seeks to explain the factors which determine the trends present in the PISA data. It looks briefly at the other administrations of the UK (Scotland and Wales in particular), but highlights the vital differences between those administrations. I maintain that ‘The UK’ cannot be treated as a unitary system.

1 Introduction

This chapter does not examine methodological issues associated with the PISA data for England. For this, I refer readers to the work of John Jerrim (particularly his 2011 paper on the limitations of both TIMSS and PISA), Benton and Carroll (2018) and Micklewright and Schnepf (2006).

The time series 2000–2018 is hampered by two issues. Firstly, possible mode effects following the switch to on-screen administration (Jerrim et al 2018) and secondly, the failure of England in 2000 and 2003 to meet the full sample criteria. The 2000 data are regarded as problematic, however the 2003 data are available in Micklewright and Schnepf 2006, are considered by them to be adequate, and are included in Table 1 here.

The curriculum- and instructional-sensitivity of the PISA items has been compared with TIMSS (Schmidt 2018), and I make here no assumption that PISA is

T. Oates (✉)

Director Assessment, Research and Development, CBE, Cambridge Assessment, Cambridge, England

e-mail: oates.t@cambridgeassessment.org.uk

© The Author(s) 2021

N. Crato (ed.), *Improving a Country’s Education*,
https://doi.org/10.1007/978-3-030-59031-4_4

83

Table 1 PISA scores 2000–2018 for England

	2000	2003	2006	2009	2012	2015	2018
Maths	531	507*	495	493	495	493	504
Reading	526	507*	496	495	500	500	505
Science	535	516	516	515	516	512	507

Source OECD *England failed to meet the 85% participation school sample criterion in 2003, but met the 80% student response criterion

an entirely unproblematic or infallible measure of underlying improvement or deterioration of educational standards in England. But what I look at here is a set of correspondences—between changes—or not—in PISA scores, and policy actions. When timelines are aligned with sensitivity, and plausible time lags taken into account, some interesting possible relationships emerge. It is these on which this chapter focuses.

2 The National Scene

It's November 2019 in England, and some of the educational journalists are getting restless. They are beginning to make enquiries about the expected PISA results. In December, immediately after publication of the results, their stories will enjoy a brief flurry of prominence and they know they will swiftly move onto other things. In England, the PISA results subsequently will be cited in scattered articles which talk about educational performance and government policy (for example see *Schools Weeks 2019*; *TES 2019a*). Various politicized organisations will issue immediate forthright comment, seldom agreeing (*Financial Times 2016*). Domestic researchers will of course then begin their wide-ranging scrutiny of the outcomes, but their reflections and insights will take time, and will most likely only be seen in academic journals.

It's not that PISA in England is treated as a trivial matter. John Jerrim continues to provide influential methodological critique on both survey approach and interpretation (Jerrim 2011). The National Foundation for Educational Research provides important comparisons of PISA, PIRLS and TIMSS (e.g. *NFER 2018*). The Department for Education's international unit provides thorough and reflective time series perspectives. And Cambridge Assessment's researchers and policy advisers—Tom Benton in the foreground (e.g. *Carroll and Benton 2018*)—review method and carefully weigh up the OECD PISA reports against their own extensive transnational comparisons of education system performance around the world.

But for English policy makers, while important for international benchmarking, PISA is by no means the most important body of data for domestic discussions of educational performance. There are a number of reasons for their views, which now I will explore.

PISA has the known limitation of being a cross-sectional analysis of 15-year olds, lacking the quasi-longitudinal structure of TIMSS—‘quasi-longitudinal’ since on a four-year survey cycle TIMSS tests a sample from year 4 and year 8, although the pupils in the year 8 sample are not the exact individuals tested in the previous cycle’s year 4 sample. This presents specific limitations on inference from the PISA data. Timeframes of system reform have to be taken carefully into account, a heavily contested issue which I will deal with in more detail later in this chapter. Even seasoned analysts can forget to think about the total experience 0–15 of those 15-year olds—what was their experience in life, and what were the characteristics of the education system as they passed through it? What exactly did the 10 + years of school experience prior to the assessment point in PISA at age 15 contain? Is what is happening at 15 consistent with earlier education and experiences? This provides a different take on the unsound conclusions which can float around PISA. The following provides a telling example.

An April 2018 headline declared ‘*Exclusive: England held back by rote-learning, warns PISA boss—England’s schools system is losing ground to the Far East because of an emphasis on rote-learning and a narrowing of the curriculum, says the official behind the Pisa international education rankings*’ (TES 2018a). The article emphasises the 2015 PISA finding that “...*Britain comes out right on top...*” in terms of the amount of rote-learning in its schools. But this highlights starkly the limitation of inference from a survey of 15-year-olds—quite the wrong conclusions about the system as a whole can be assumed. These pupils are close to their GCSE examinations, taken at age 16. In this, England is typical: most systems have high stakes assessments at 16 (Elliott et al. 2015). England is atypical in using externally set examinations for these. The GCSE examination results are high stakes for schools as well as pupils—schools are measured on the grades obtained, and the ‘value added’ which each school’s programme presents. Particular subjects are counted towards a target—the English Baccalaureate ‘basket’ of GCSE qualifications grades. Rote learning as a feature of education in this exam-focussed phase is widely acknowledged in England (Bradbury undated; Mansell 2007) but its origins are complex. The 2010 Coalition Government introduced new GCSEs in English and Mathematics—first teaching from September 2015. In sharp contrast to the older qualifications—which included large amounts of coursework—the new GCSEs require extensive memorization—for example of poetry and segments of drama. Whilst the 2015 PISA cohort were not on programmes directed to these new qualifications, there was widespread discussion permeating schools about the dramatic rise in memorization required. Staff had seen new sample assessment materials and were busily preparing new learning programmes of high demand. Fifteen-year olds measured in PISA 2018 were highly sensitized to the reformed qualifications about to be taught in the system. Understandably, memorisation was a preoccupation of these students, and their teachers, as they approached their public examinations. But this should not be generalized to all pupils of all ages; it speaks of the reality for 15-year-olds.

There is further fundamental background to the PISA finding. Rote learning is seen as an essential component of learning in some Asian systems (Cheong and Kam 1992; Crehan 2016)—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 (Christodoulou 2014; Au and Entwistle 1999). This approach is endorsed by contemporary cognitive science (Abadzi 2014; Kirschner et al. 2006). However, it was an approach which was strongly discouraged in primary schools in England as a consequence of the recommendations of the 1967 Plowden Report.

Concerned to improve reading and maths attainment, the 2010 Coalition Government re-emphasised the importance of rote learning, particularly in respect of elements of mathematics—and particularly in respect of multiplication tables (BBC 2018). The action subsequently taken in the revised National Curriculum of Sept 2014 to strategically re-introduce rote learning into primary education was irrelevant to the 15-year-olds in the 2015 PISA survey. But this highlights an important *absence* of rote learning from the 2015 PISA cohort's primary education. With a general absence of rote learning of the form which supports higher level functioning (Sammons et al. 2008), it is unsurprising that the PISA cohort's perception is one of a highly pressured year immediately prior to high stakes examinations, characterized by subject content which need to be memorized. An appropriate reading of the PISA findings is the exact reverse of the headline '...a system dominated by memorization...'. That is, the findings should not be read as an indication of the prevalence of rote learning throughout the compulsory school system in England, but as a sign of its general neglect and absence—and an urgent preoccupation with it as pupils approach demanding national assessments.

This illustrates the extent to which extreme care needs to be taken in the interpretation of the outcomes of a cross-sectional survey of 15-year-olds. Historical context, time lags, domestic analysis all need to be taken into account.

Whilst aware of the limitations, and whilst sceptical of some of the top line conclusions from PISA reporting, policy makers and politicians in England certainly maintain a keen interest in the underlying measurements of mathematics, reading and science—and certainly the trend data. But they view PISA data as an important *part* of the picture. A part, not the whole—and whilst important for a short time around publication, in England the PISA data quickly become of secondary importance. Why? Not because of PIRLS or TIMSS—although they too are of course of interest when the reporting from them begins. No, for understanding domestic performance, PISA is of secondary interest because of the quality and comprehensiveness of England's National Pupil Database (NPD).

While exact equivalents of the PISA data on classroom climate and social/familial background contextual data is not collected in the NPD, the NPD includes essential school and pupil characteristics (birthdate, gender, school location, etc.) and attainment data (phonics check, national tests, national qualifications) for every child and every educational setting. It is massive, comprehensive, underpinned by law, of high quality, and well-curated (Jay et al. 2019). The NPD supports vital and wide-ranging analysis of equity and attainment throughout England. It is now linked to educational databases for further and higher education, and to data from the labour market. Scrutiny of these data allow sensitive analyses of the distribution of attainment within pupil groups and across geographical areas, the performance of schools,

right through to analysis of the comparability of qualifications and standards in qualifications over time. It is these massive domestic datasets which are at the forefront of policymakers', politicians' and researchers' enduring interest (Jay et al. op cit; TES 2018b). Sitting in the sidelines there also are the domestic cohort studies: the 1958 National Child Development Survey (NCDS)—all children born in the first week of February 1958- the British Cohort Survey 1970 (BCS), and the Millennium Cohort Survey 2000 (MCS). These are genuine longitudinal surveys, following life outcomes in health, education, employment—allowing extraordinary insight into the role of education in society, and society's impact on education.

3 PISA 2018 in 2019

So...PISA is interesting, but not sole reference point for English commentators and analysts. Let's go back to November 2019 and the growing anticipation in the countdown to the publication of the 2018 PISA results. What was anticipated...and why? Since the election of a coalition government in 2010, there have been genuine structural changes in provision and shifts in aims.

Principal amongst these have been:

- (1) From 2010, massive shifts of schools from local authority (municipality) control to direct contractual relations with central government. This 'academy' policy originated in the 1997–2010 Labour administrations, with the first 'academies'—essentially a change in governance—appearing in 2001. The policy was extended to potentially include all schools from 2010. 'Free schools' also were introduced as a new category of school from 2010; schools suggested and constituted by parents and community groups, approved by the State (Alexiadou, Dovemar & Erixon-Arreman 2016). Again, these are in direct contractual relationship with central government.
- (2) In Sept 2014, a new National Curriculum. A new curriculum for secondary was designed in 2007, but was rejected by the 2010 Government as poorly theorised and lacking international benchmarking. By contrast, the 2014 revisions emphasised clear statement of demanding content—concepts, principles, fundamental operations and core knowledge—with this content strongly benchmarked to high performing jurisdictions. It should be noted that the National Curriculum is not a strict legal requirement in all schools in England. Independent schools (private schools) are not under legal obligation, although their performance tends to be judged by their outcomes in national examinations at 16 and 18. Other classes of schools are required to participate in national testing and national examinations at 16 and 18, but the 72% of secondary schools and the 27% of primary schools which are academies are not required to follow the programmes of study of the National Curriculum (NAO 2018).
- (3) From 2010, a strong emphasis on improving reading in primary schools. While the previous Labour governments 1997–2010 had put in place the Literacy

and Numeracy strategies, only in the closing years of the Strategies had there been a move from diverse approaches to reading towards more evidence-based methods (Chew 2018). This followed Andrew Adonis' commission for the Rose Report on early reading (2006). From 2001 to 2006, outcomes fell in PIRLS—the Progress in International Literacy Study—which tests year 5 pupils. Scores fell from 553 to 539. They then rose between 2006–2011, climbing back to their 2001 level (score 552 in 2006) and continuing to improve for the 2016 survey (score 559)—with the 2016 figures representing a substantial closing of the gender gap—previously, in 2011, England had possessed one of the largest gender gaps in PIRLS.

Although synthetic phonics increasingly had been the focus of the last years of the Literacy Strategy, the lack of professional consensus around methods was clear in the vigorous and adverse reaction to the incoming 2010 Coalition Government's emphasis on phonics. This was considered to be highly controversial and was widely discussed in the media. Many then-prominent educational commentators were critical of this strong emphasis and argued that specific techniques should be decided upon by schools (Guardian 04 03 14; Clark 2018). Government was not deterred, and asserted phonics-based reading schemes through textbook approval procedures and, from 2012, a statutory 'phonics screening check' was introduced for Year 1 pupils (a 40-item test with a threshold score of 32). Notably, highest attainers in the 'phonics check' also were high performers in the 2016 PIRLS survey (McGrane et al. 2017). The Government's commitment to phonics subsequently was justified by both a new community of researchers (Willingham 2017; Machin et al. 2016) as well as by the continued improvement in PIRLS results for year 5 pupils. The phonics check for year 1 children also saw escalating scores—from 31.8% reaching the threshold score in the trial test of 2010, climbing to 81% in 2017. Reading attainment improved; the gender gap was significantly reduced.

- (4) From 2010, alongside the focus on enhancing reading, maths education was enhanced through a series of measures. Targeted funding was allocated to the National Centre for Excellence in the Teaching of Mathematics (NCETM, founded in 2006 by non-government bodies). Under specific contract to Government, NCETM has managed a wide ranging programme of support to schools through designated 'Maths Hub' schools (37 in total, working with over 50% of schools in the country, with over 360,000 registrations for information from the Centre), managing a research-based teacher-exchange with Shanghai, providing teaching resources and professional development, and participating in Government textbook approval processes. This emphasis on professional development and high-quality textbooks is a highly distinctive feature of the policy work on maths.
- (5) In 2015, revised GCSEs (the key examinations at the end of lower secondary, typically taken at age 16) in Maths, English and English Literature were introduced, following national debate about declining assessment standards in key examinations (Cambridge Assessment 2010). Teaching started in September

2015, with first examinations in 2017. These dates are important for interpretation of the 2018 PISA survey. New GCSEs in all remaining subjects (sciences, geography, languages etc.) were introduced in 2016, with first examinations in 2018. Significant changes were made to content, assessment methods, grading, and overall level of demand. Increasing ‘content standards’ was fundamental to the reform, with a recognition that a significant increase in demand was required to ensure alignment with high performing jurisdictions (Guardian 2013a). At the same time, a new grading scale was introduced (Ofqual 2018), moving from the old system of ‘A*-G’ (with A* being highest) to 9–1 (with 9 being highest). Great attention was paid to managing, through statistical means, the relationship between grading standards in the old GCSEs and the revised GCSEs (Ofqual 2017).

- (6) Revised targets: with the introduction in 1988 of the National Curriculum and allied national testing, there emerged policy options regarding the use of the data from those tests and from the existing national examinations at age 16 (GCSE) and 18 (GCE A Level). A national survey-based analysis of standards had been operated in previous decades by the Assessment of Performance Unit (Newton 2008) but with the introduction of national testing for every pupil, government sensed that greater accountability for each school could be introduced into the system, rather than simply the production of policy intelligence on overall national standards. The publication of individual school results, and the strong use of assessment data in school inspection became a feature of the system from 1992. For over a decade, the published data focussed on simple measures such as the percentage of pupils achieving specific ‘levels’ in national assessments in each primary school (with Level 4 being the ‘target’ level in Maths and English) and in each secondary school achieving 5 GCSEs at grades A*-C. More elaborated measures were added in 2002 (‘value-added’) (Leckie and Goldstein 2016). The evolution of school league tables in England occurred during the period 1992–2016: ‘contextual value-added’, ‘expected progress’ and ‘progress 8’, modified in 2006 (‘contextual value-added’) and 2011 (‘expected progress’) and significantly remodelled in 2016 (‘Progress 8’). The early GCSE targets had driven damaging behaviour in schools, with many schools focusing on grade C/D borderline pupils, to the detriment of both higher attaining and lower attaining pupils—and prior to 2010, successive governments had persisted with these crude targets despite clear research regarding adverse impact (Oates 2014). Later refinements, particularly Progress 8, are a committed effort to improve the validity of the representation of the performance of individual schools (principally, accounting for intake), and to drive desirable behaviours in schools. As a response to ‘gaming’ previous cruder GCSE qualification targets and measures, the Government in 2010 introduced the ‘English Baccalaureate’ (EB) performance measure—designed to ‘increase the take up of ‘core’ academic qualifications which best equip a pupil for progression to further study and work’ (House of Commons Library 2019). Typically pupils take 10 GCSE examinations. The EB measure requires pupils to achieve specific grades in English Literature and Language, Mathematics, History or Geography, two Science and a Language.

This allows a degree of choice in the 3 or more GCSEs which pupils take in addition to the EB requirement. The national target was to reach 90% of pupils reaching the EB target by 2020, but in 2017 a new target was set at 75% by 2022. It is not a legal requirement, but it is viewed as high stakes by schools.

- (7) Changes in teacher supply: contrary to popular press stories around PISA 2000, Finnish teachers are not the most respected in the world, with the ranking from NIESR data giving Brazil 1 (lowest) Finland 38, England 47, and China 100 (highest) (Dolton and She 2018). Likewise, while starting salaries are higher in Finland, there is far greater post-qualification growth in salaries in England (OECD 2003). Yet in England teaching generally is portrayed as an undesirable occupation, with commentary focusing on (i) the high level of direct government scrutiny of schools via accountability arrangements and (ii) a high level of non-teaching activities which detract from quality of professional and personal life (NFER 2019).

Government rightly has seen the problem as twofold—a problem of supply (both in the nature of training and in the quantity of supply) and a problem of retention (content of the professional practice). Principally, in respect of supply, from 2010 the focus of initial training was switched from universities to schools and school-university partnerships (Whitty 2014) and a flagship scheme, Teach First, was launched to encourage high performing graduates into teaching (Teach First undated). In respect of professional practice, a teacher workload survey and review was commissioned, to both understand and act on the reported problems of workload and role.

These initiatives and developments—and the problems to which they are a response—are major aspects of the context which needs to be taken into account when interpreting the 2018 PISA results. A lot has changed in the period prior to the 2018 PISA data capture on 15 year olds. Both the scale of change and, crucially, the timing of the impact of the changes are embedded in the pattern of the PISA outcomes.

The scale of the post-2010 changes—so many aspects of the education system simultaneously being re-cast by new policies—were criticized by Parliamentary Select Committee as ‘hyperactivity’ on the part of the then-Secretary of State Michael Gove (Guardian 2013b). But the need to work across all aspects of arrangements was driven by explicit theory, not mere personal tendency of the Secretary of State. The commitment to international benchmarking and to ‘policy learning’ from international comparative research including examination of the characteristics of high performing systems. Bill Schmidt’s work on ‘curriculum coherence’—where instruction, assessment, standards and materials carefully and deliberately are aligned—was extended into a wider examination of coherence across all key aspects of arrangements (Cambridge Assessment 2017). In 2010 this mobilized wide ranging policy review by the Secretary of State. This emphasised the need to ensure coherence across accountability, curriculum standards, professional practice, institutional development and so on—aiming to remove in particular the tensions between accountability (targets, reported data, inspection) and curriculum aims which had been evident for so long (Hall and Ozerk 2008). The use of international evidence on effective pedagogy

in reading and maths was focussed intently on securing ‘curriculum coherence’. All this was not so much ‘hyper-activity’ as a perceived need to align key elements of arrangements quickly and effectively; a demanding act of public policy.

4 PISA as a Reflection on Post-2010 Policy Moves

So, it was in the context of these changes which journalists and researchers waited for the 2019 publication of the 2018 PISA results. It was clear that a simple look at timing of the policy actions—emphasis on reading, new National Curriculum, new qualifications—should give everyone pause for thought. Timing of change, and the time lag involved in genuine system impact is essential in interpreting international survey data such as PISA, TIMSS and PIRLS. This seemed entirely to be forgotten in the noise in 2001 after Finland was announced as ‘first in the world’ (BBC 2015). Time lags and the necessity of relating the timing of actions to effects were ignored by the majority of commentators. Profoundly misleading narratives have been created as a result.

It is a necessary statement of the obvious that the pupils in the PISA survey were 15 years of age. In England, most of the surveyed pupils were in their 11th year of schooling. For those in year 11, they had only experienced the new National Curriculum for their secondary education—from year 7 (age 11). They had not experienced a full education aged 5–15 under the new curriculum. The new curriculum was introduced in September 2014, when the PISA cohort already had entered secondary education. But it is important to note that the revised National Curriculum was intended to have its biggest impact in Primary Education. The PISA cohort was not exposed to this new curriculum. And no new national curriculum, especially one committed to a radical reshaping of learning, is implemented perfectly in its first year of operation. The intention of National Curriculum policy in England was to shape the curriculum in secondary education through the revised, more demanding specifications of national qualifications, along with national targets and accountability instruments—particularly the English Baccalaureate requirement.

The policy around years 7, 8 and 9—the first 3 years of secondary education—was controversial. The new National Curriculum was stated in a very specific and detailed form for Primary education, including a move to a new year-by-year format. Years 7, 8 and 9 were stated as a much more general requirement and, unlike the new Primary specification, was treated as a ‘Key Stage 11–14’ rather than as separate years. The controversial assumption of the policy was that the period from year 7 to year 11 would be seen by schools as a continuum of learning, culminating in GCSE examinations in around ten selected subjects. This attracted criticism, since there was a dominant notion in educational discussion that ‘learning already was far too heavily determined by exams and assessment’ (Mansell *op cit*). But the 2010–13 policy advisers committed to a ‘continuum of learning 11–14’ principle, with the detail of learning targets linked to the detailed and carefully designed objectives of national examinations rather than a set of detailed year-by-year National Curriculum

statements. They felt that this would not lead to narrow, restrictive learning, and by contrast should equip pupils with the wide reading, extended writing, critical thinking, rich discussion etc. which would lead to enhanced outcomes in national exams at 16. Policy advisers noted the evidence, from teachers, of the strong lower secondary ‘washback effect’ from the demands of national examinations and, rather than working against it, intended that it be used to intensify and focus the learning in the first years of secondary education. There was no strong, explicit statement of this principle, since the washback effect was so strongly evident in the system. While national inspection reports noted that years 11–13 frequently were the ‘lost years’ in the system (Ofsted 2015), the policy assumption was that the significantly increased demand of the new national examinations and accountability requirements would intensify these early secondary years.

But for interpretation of the PISA results, it is essential to note that the new examinations were only introduced in 2015 (maths and English) and 2016 (other subjects). The 2018 PISA cohort therefore did not experience either the new National Curriculum during their primary education 5–11 nor the intended intensification of 11–14 education. They also experienced the new qualifications only in the immediate years after first implementation, a period typically associated with sub-optimal performance of the system due to (i) lack of established, stable and refined teaching approaches, (ii) uncertainty regarding exact expectations, (iii) unrefined professional support (Baird et al 2019).

Thus, in November 2019, when expecting the PISA results, some researchers and commentators—including this author—were taking these factors into account and anticipating the possibility of static or even depressed PISA outcomes for England. Yet, on publication, England’s results showed significant improvements in mathematics, an apparently stable position in reading but improved performance relative to other benchmark nations, and high but static performance in science. How should we interpret this? The simplest explanation is also a positive one; that despite the depressive effects of system changes, real gains in attainment were being secured.

The improved maths performance comes after decades of protracted flat performance. Again, carefully considering timelines, it was anticipated that pupils in previous PISA cycles might benefit from the Numeracy strategies of the late 1990s—but no such effect is obvious in the data, unless the time lag is unfeasibly long. The increase corresponds more exactly with the post-2010 emphasis on mathematics—the wide and varied practical policy combined with high profile public discourse, demanding targets, and some ‘washback’ from elevated standards in public examinations.

The reading scores demand careful interpretation. PIRLS data shows an increase 2006–2011 which suggests some elevation of reading prior to 2010. The high level of reaction against phonics in 2010 suggested that varied methods for early reading were established in the system. School inspection reports reinforce this view (Ofsted 2002; Department for Education 2011). The strong emphasis on synthetic phonics, re-enforced by the ‘phonics check’ in Year 1 co-incides with an elevation of performance and closing of the gender gap in PIRLS. In PISA—assessing 15-year-olds—with only small variation in scores since 2006, performance of the education system

appears moribund. But this is deceptive. Background trends need to be taken into account. As we know from the Finland context, reading is significantly influenced by factors outside the school system (Tveit 1991). The 2018 PISA elevated reading scores in the USA are all the more remarkable in the light of the significant gradient of decline in reading speed and related comprehension since the 1960s (Spichtig et al. 2016). Likewise, England's score should not be seen solely in terms of the country's static trend in PISA, but in the relation between England's trends and those of other high-performing nations. In 2015, 12 countries outperformed England. Germany, Japan, New Zealand and Norway all outperformed England in 2015, but had similar scores to England in 2018. Those that had similar scores in 2015—Belgium, France, Netherlands, Portugal, Russian Federation, Slovenia, and Switzerland) were outperformed by England in 2018. These comparisons cast an interesting light on the seemingly static performance in England. In addition, the gender gap is significantly lower than the OECD average. However, equity remains challenging. Notably, the increase particularly has been amongst higher performing pupils. The lowest achievers' scores have remained static, but the difference between high and low achievers in 2018 is similar to the OECD average. The international picture, from both PISA and other sources, suggests a *significant international widespread decline* in reading—and this without measuring the technology-driven switch in reading habits and family environment which is occurring over a dramatically compressed timeframe (Kucirkova and Littleton 2016). With the PIRLS data showing similar relative improvement in performance for England, the results appear to endorse the post-2010 policy emphasis on reading—similar to the macro and micro policy emphasis on maths—and substantial benefit in the practical action which has been put in place with schools.

A sub-domain in the 2018 sweep, science results in England provide a different story to maths and reading, and suggest that government should both sustain its approaches in those two subjects *and* attend to policy action aimed at primary and secondary science performance. With very few specialist science teachers in Primary, and science testing withdrawn from national assessment prior to national qualifications at 16, incentives and drivers have declined in the both the pre- and post-2010 period. While over this period over 400 initiatives from various sources (House of Lords 2006), of various types and various scales have been implemented across arrangements, TIMSS Grade 4 data in 2011 showed a fallback to late 1990s performance levels (Sturman et al. 2012).

Performance in science is not crashing, it just remains static—and shows challenging equity outcomes—a high gap between high and low achieving pupils, and with a higher proportion of pupils achieving at the highest proficiency levels. England bucks the international trend in terms of gender: no significant gap in attainment, unlike the OECD gap in favour of females. But it is essential to recognise that overall the gender gap data in England is not positive: national qualifications data in science subjects at 16 and particularly at 18 remain highly gendered (Cassidy et al. 2018).

The 2018 mean score for England remains higher than the OECD average, and 10 nations had overall scores higher than England. But 12 others in 2018 had a significant

drop in performance: Australia, Canada, Denmark, Finland, Japan, Norway, Spain and Switzerland. Only two secured an increase: Poland and Turkey.

5 England Within the United Kingdom; the Devolved Administrations

All of this gives the post-results overview for England, linking it to policy actions and long term timelines. But this is England—what of the United Kingdom, of which England is a part? Usefully, the survey design and the implementation requirements result in PISA providing valid data for the devolved administrations of the UK—Scotland, Northern Ireland and Wales. I have said nothing of these latter territories. And deliberately so. They are different from England in vital ways, and different from each other (with an emerging important exception in the apparent increasing convergence of Wales with Scotland). All of this provides the most extraordinary natural experiment—the late David Raffe’s ‘Home international comparisons’ (Raffe 2000).

Scotland has for the past two decades worked on the ‘Curriculum for Excellence’; an increasingly controversial approach to defining, specifying and delivering the primary and secondary curriculum which emerged from a 2002 consultation exercise. Implemented in 2010–11, with new qualifications in 2014, it uses models strongly contrasting with those in England (Priestly and Minty 2013). For Scotland, an increase in reading in PISA 2018 accompanies an unarrested decline from 2000 in maths and science.

Wales is undertaking radical reform, in the light of a previous history of low results relative to the rest of the UK and of previously declining scores. It is looking to the Scottish model in the ‘Curriculum for Excellence’ (Priestly *op cit*; TES 2019b) rather than the post-2010 actions and models in England. But timelines remain important. The new curriculum model in Wales has not yet been fully designed and specified, let alone implemented—the target date for enactment being 2020. Yet in 2018, reading scores improved over 2015, science reversed a severe decline, and maths continued the improvement which was first seen in the 2015 PISA outcomes.

Northern Ireland remains distinctive and different—its arrangements heavily shaped by its history and its size and geography. Pupils there performed better than pupils in Wales, but slightly lower than England in all three domains. Performance was below Scotland in reading—Scotland’s improved domain—but above Scotland in science and maths. However, science in Northern Ireland has shown significant decline since 2015, despite an unchanged position 2015–2018.

When considering the relative performance of England, Wales, Northern Ireland and Scotland, I have emphasised just how essential it is to avoid lapsing into any assumptions or view that ‘The UK’ can be regarded as a unitary system. It cannot. Indeed, far from it: the systems in the different administrations are highly distinctive, increasingly driven by very different assumptions, models and policy instruments.

6 Diversity and Difference

And it is on this note of ‘difference’ which this chapter ends. David Reynolds, co-author of the influential and incisive transnational 1986 report ‘Worlds Apart’ (Reynolds and Farrell 1996) has repeatedly emphasised the importance of within-school variation in England, which remains amongst the highest in high-performing PISA nations. Arrangements in England also manifest high between-school variation. This hints at a level of variation in educational forms which has been under-analysed and under-recognised. Diversity of institutional forms, curriculum assumptions, professional practices is extraordinary in England. In a country of 66.4 million (ONS 2019) there are extremely large and very small schools and everything in between. There are free schools, state schools, independent schools, academies and academy chains. There are selective authorities and non-selective areas. The transfer age between primary, lower secondary and upper secondary varies from area to area. Setting and streaming is managed in very different ways in different schools. Micro-markets have emerged in different localities with different patterns of schools (Gerard 1997). There is historical legacy which gives large variations in school funding from area to area. There is a possibility of parental choice of school in some areas and operationally, none in others. Regional variation in growth and economic activity shows similar extreme variation (ESCOE 2019).

The picture of diversity in educational performance in England is rendered complex by the peculiar distribution of the ‘unit of improvement’. In some cases, schools in ‘academy chains’ (allied groups of schools) are clustered in a locality—in other instances, they are widely geographically distributed. In addition, city development strategies of the early 2000s—the most prominent being ‘London Challenge’—lent policy integrity to a specific urban localities. While the underlying causes and extent of improvement associated with London Challenge are contested (Maccougall and Lupton 2018), the ‘City Challenges’ indicate a period of change where key ‘units of development’ were at the level of large urban centres, rather than the nation as a whole.

Research which the author undertook across England during the National Curriculum review showed not only high variation in forms of education and professional practice, but high variation in assumptions and values regarding curriculum, assessment, and pupil ability. Few other nations appear to possess such high structural and internal variation across all dimensions of provision. This poses a massive challenge to policy makers, who must anticipate very different conditions for the impact of national policy measures, and high resilience and resistance (Cambridge Assessment 2017). Sustained improvement therefore suggests particular potency and design integrity in the forms of public policy which legitimately can causally be associated with that improvement. The Reading and Maths results should be seen as signs of genuine policy achievement in a highly diverse and challenging context. When improvement bucks the trends present in society, as in the case of literacy, policy makers can be doubly satisfied with their and teachers’ endeavours.

References

- Abdazi, H. (2014). How to define education goals for the post-2015 period? Guidance from memory functions unpublished working paper.
- Alexiadou, N., Dovemark, M., & Erixon-Arreman, I. (2016). *Managing Inclusion in Competitive School Systems: The Cases of Sweden and England Research in Comparative and International Education*. <https://doi.org/10.1177/1745499916631065> accessed 03/03/20.
- Au, C., Entwistle, N. (1999). Memorisation with understanding. In *approaches to studying: cultural variant or response to assessment demands?* University of Edinburgh Education Line Database.
- Baird, J. A., Caro, D., Elliot, V., El Masri, Y., Ingram, J., Isaacs, T., et al. (2019). *Examination Reform: Impact of linear and modular examinations at GCSE*. Oxford/Coventry: OUCEA/Ofqual.
- BBC. (2015). *Finns aren't what they used to be 15 04 20 BBC News* <https://www.bbc.co.uk/news/business-32302374>. Accessed March 27, 20 18:20.
- BBC. (2018) Times tables check trialled ahead of rollout 14 02 18 BBC News <https://www.bbc.co.uk/news/education-43046142>. Accessed March 27, 20 14:10.
- Benton, T., & Carroll, M. (2018). *The link between subject choices and achievement as GCSE and performance in PISA 2015*. Cambridge: Cambridge Assessment.
- Bradbury A undated Pressure, anxiety and collateral damage—the headteachers' verdict on SATs, UCL Institute of Education, London.
- Cambridge Assessment. (2010). *Exam standards: The big debate*. Cambridge: Cambridge Assessment.
- Cambridge Assessment. (2017). *A Cambridge Approach to improving education*. Cambridge: Cambridge Assessment.
- Cassidy, R., Cattan, S., & Crawford, C. (2018). *Why don't more girls study maths and physics?* London: Institute of Fiscal Studies.
- Cheong, A. Kam, H. (1992). *Singapore students are rote learners; Fact or Myth? ERA conference paper*, September 24–26, 1992, Educational Research Association of Singapore (ERAS).
- Christodoulou, D. (2014). *Seven myths about education*. Abingdon: Routledge.
- Clark, M. (2018). How justified is it to dominate government policy on early literacy with synthetic phonics and the phonics screening check part II a critique of the evidence 14. *Education Journal, Issue, 352*, 14–19.
- Crehan, L. (2016). *Cleverlands: The secrets behind the success of the world's education superpowers*. London: Unbound.
- Department for Education. (2011). *The National Strategies 1997–2011*. London: Department for Education.
- Dolton, P., & She, P. (2018). *Global Teacher Status Index*. London: National Institute of Economic and Social Research.
- Elliott, G., Rushton, N., Darlington, E., & Child, S. (2015). *Are claims that GCSE is a white elephant red herrings?* Cambridge: Cambridge Assessment.
- ESCOE. (2019). *Differences in economic growth across the UK remain substantial in the year to Q1 2019, Economic Statistics Centre of Excellence* <https://www.escoe.ac.uk/differences-in-economic-growth-across-uk-remained-substantial/>. Accessed February 14, 20 12:50.
- Financial Times. (2016). *Ambitious education reforms fail to lift UK Pisa school rankings* December 06, 16 Financial Times <https://www.ft.com/content/1ea10484-bb14-11e6-8b45-b8b81dd5d080>. Accessed March 19 20, 16:15.
- Gorard, S. (1997). *Market forces, choice and diversity in education: the early impact Sociological Research Online* <https://doi.org/10.5153/sro.111>. Accessed February 12, 20 15:50.
- Guardian. (2013a). *Michael Gove unveils GCSE reforms*, June 11, 13, Guardian.
- Guardian. (2013b). *Graham Stuart: 'You can't drive education like a sports car'*, March 25, 13, Guardian.
- Guardian. (2014). *Reading Lessons; why synthetic phonics doesn't work*, March 04, 14, Guardian.
- Hall, K., & Ozerk, K. (2008). *Primary Curriculum and assessment: England and other countries Cambridge Primary Review*. Cambridge: University of Cambridge.

- House of Commons Library. (2019). *English Baccalaureate* <https://researchbriefings.parliament.uk/ResearchBriefing/Summary/SN06045>. Accessed March 12, 20 10:20.
- House of Lords. (2006). *Science teaching in schools: Report with evidence; 10th report of session House of Lords Science and Technology Committee*. London: The Stationary Office.
- Jay, M. McGrath-Lone, L., Gilbert, R. (2019). Data Resource; the National Pupil Database (NPD). *International Journal of Population Data Science*, 4(1) <https://ijpds.org/article/view/1101>. Accessed March 01, 20 16:00.
- Jerrim, J. (2011). *England's "plummeting" PISA test scores between 2000 and 2009: Is the performance of our secondary school pupils really in relative decline?* DoQSS Working Papers 11–09, Department of Quantitative Social Science, UCL Institute of Education, London.
- Jerrim, J., Micklewright, J., Heine, J. H., Salzer, C., & McKeown, C. (2018). PISA 2015: How big is the 'mode effect' and what has been done about it? *Oxford Review of Education*, 44(4), 476–493.
- Kirschner, P., Sweller, J., & Clark, R. (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*, 41(2), 75–86.
- Kucirkova, N., & Littleton, K. (2016). *The digital reading habits of children*. Manchester: Arts Council England.
- Macdougall, A., & Lupton, R. (2018). *The 'London Effect': Literature review*. Manchester: University of Manchester & The Rowntree Foundation.
- Machin, S., McNally, S., Viarengo. (2016). Changing how literacy is taught: evidence on synthetic phonics. *American Economic Journal*, 10 (2), 217–241.
- Mansell, W. (2007). *Education by numbers: the tyranny of testing*, Politico's Publishing, Michigan.
- McGrane, et al. (2017). *Progress in International Reading Literacy Survey (PIRLS): National Report for England December 2017 DFE-RR770*. London: Department for Education.
- Micklewright, J., & Schnepf, S. V. (2006). *Response bias in England in PISA 2000 and 2003 Department for Education Research Report RR771*. London: Department for Education.
- National Audit Office. (2018). *Converting maintained schools to academies PN: 15/18*. London: National Audit Office Press Office.
- Newton, P. (2008). *Monitoring national attainment standards—a collection of working papers*, Office of Qualifications and Examinations Regulation, Coventry.
- NFER. (2018). *NFER educational briefings—Key insights for England from PIRLS, TIMSS and PISA*, National Foundation for Education Research, Slough.
- NFER. (2019). *Teacher labour market in England—Annual Report 2019*, National Foundation for Educational Research, Slough.
- Oates, T. (2014). The 'qualifications sledgehammer': why assessment-led reform has dominated the education landscape. In: Heller-Sahlgren, G. (ed) *Tests worth teaching to*, Centre for Market Reform of Education, London.
- OECD. (2003). *Attracting, developing and retaining effective teachers: Country background report for Finland*. Paris: Organisation for Economic Co-operation and Development.
- Ofsted. (2002). *The National Literacy Strategy: the first four years 1998–2002 Office for Standards in Education*, London.
- Ofqual. (2017). *Modelling GCSE grade 9 to 1 outcomes in 2017: A technical explanation*, Office of Qualifications and Examinations Regulation, Coventry.
- Ofqual. (2018). *GCSE 9 to 1 grades: a brief guide for parents 02 03 18 Blog* <https://ofqual.blog.gov.uk/2018/03/02/gcse-9-to-1-grades-a-brief-guide-for-parents/accessed> February 21, 20.
- Ofsted. (2015). *Key Stage 3: the wasted years?* Office for Standards in Education, London.
- ONS. (2019). *Population estimates for the UK, England, and Wales, Scotland and Northern Ireland: Mid-2018*. Newport: Office for National Statistics.
- Priestly, M., & Minty, S. (2013). Curriculum for excellence; 'A brilliant idea, but....' *Scottish Educational Review*, 45(1), 39–52.
- Raffe, D. (2000). *'Home international' comparisons of post-16 education and training*. CES Briefing: University of Edinburgh, Edinburgh.

- Reynolds, D., & Farrell, S. (1996). *Worlds apart?: A review of international surveys of education achievement involving England*. London: Stationary Office Books.
- Sammons, P., et al. (2008). *The influence of school and teaching quality on children's progress in primary school DCSF-RR028*. Schools and Families, London: Department for Children.
- Schmidt, W. (2018). Why there is no "holy grail" in education. Lecture Cambridge 30 05 18 available at <https://cambridgeassessment.org.uk/news/why-theres-no-holy-grail-in-education/>. Cambridge: Cambridge Assessment.
- Schools Week. (2019). *PISA 2018: Maths up, no improvement in reading or science*, December 03, 19, Schools Week <https://schoolsweek.co.uk/pisa-2018-maths-up-no-improvement-in-reading-or-science/>. Accessed January 23, 20 14:10.
- Sizmur, J. et al. (2019) *Achievement of 15-year-olds in England: PISA 2018 Results DFE-RR961* Department for Education, London.
- Spichtig, A., Hiebert, E., Vorstius, C., Pascoe, J., Pearson, D., & Radach, R. (2016). The decline of comprehension-based silent reading efficiency in the United States: A comparison of current data with performance in 1960. *Reading Research Quarterly*, 51(2), 239–259.
- Sturman, L., Burge, B., Cook, R., & Weaving, H. (2012). *TIMSS 2011: Mathematics and science achievement in England*. Slough: National Foundation for Educational Research.
- Teach First undated Our work and its impact, Teach First, London.
- TES. (2018a). *England held back by rote-learning, warns PISA boss*, March 23, 18, TES Global Ltd.
- TES. (2018b). *English national pupil database one of the most impressive in the world*, January 22, 18, TES Global Ltd.
- TES. (2019a). *Pisa results must be a relief for the government*, December 03, 19, TES Global Ltd <https://www.tes.com/news/pisa-results-must-be-relief-government>. Accessed January 12, 20 13:40.
- TES. (2019b). *Scotland's mistakes will shape Welsh education reform*, February 18, 2019, Times Global Ltd.
- Tveit, K. (1991). The development of popular literacy in the Nordic Countries. *a Comparative Historical Study, Scandanavian Journal of Educational Research*, 35(4), 241–252.
- Willingham, D. (2017). *The Reading Mind: A cognitive approach to understanding how the mind reads*. San Francisco: Jossey-Bass.
- Whitty, G. (2014). *Recent developments in teacher training and their consequences for the University Project*. London: University College London.

Tim Oates is Group Director of Assessment Research and Development at Cambridge Assessment, focusing on national and international research on assessment and measurement. In 2010 he published 'Could do better' which laid down the principles for the review of the National Curriculum in England. He was chair of the Expert Panel for Review of the National Curriculum in England. Emerging from this review, subsequent research on the quality and function of textbooks and other resources has been taken up around the world and discussed at two international summits on learning resources. He chairs various curriculum groups for the Department for Education in England. He has published widely on assessment and curriculum issues, and routinely provides briefings and advice to UK and other governments. He has worked with OECD on curriculum matters and is leading a new UNICET project on a curriculum framework for displaced children. He is Fellow of Churchill College Cambridge and in 2015 received a CBE for services to education.

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.



Estonia: A Positive PISA Experience



Gunda Tire

Abstract According to Programme for International Student Assessment (PISA) run by Organization of Economic Cooperation and Development (OECD), Estonian education system stands out as a high performing system where students from different socio-economic backgrounds achieve high results. In PISA 2018 Estonian students ranked first in reading and science and third in mathematics among the OECD countries. What has Estonia done to be at the top of the PISA league tables? There are many aspects that have contributed to the success of Estonian education. The following chapter will look at the historical background, describe the factors, policies and conditions that have contributed to the current educational landscape that has attracted considerable attention from all over the world.

1 Estonia in the Spotlight

PISA is a household name in Estonia. Everybody knows something about it, schools never refuse to participate, and every new round of PISA data release is expected with certain amount of curiosity. Estonians by nature are very self-critical and without PISA they would never admit that Estonia has one of the best performing education systems in the world.

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.

Teachers, students and school principals received official acknowledgement on job well done by the whole society and the establishment.

G. Tire (✉)
Education and Youth Authority, Tallinn, Estonia
e-mail: gunda.tire@harno.ee

© The Author(s) 2021
N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_5

101

PISA has put Estonia in the spotlight and in 2019 the top three positive topics about Estonia in the foreign media were: success stories about the digital society, genome research in Tartu University and outstanding results of Estonian students in PISA 2018.

2 Estonia and Its Education System

Estonia is a small country with 1.3 million inhabitants. It covers 45,000 square kilometres and is larger than, for example, the Netherlands, Denmark or Switzerland. It is located on the Eastern shores of the Baltic sea in the Northern part of Europe and has borders with Russia on the East and Latvia on the South. It has beautiful, pristine nature, it is rich in forests and has more than 1500 islands along its coastline. The official language is Estonian which belongs to rare Finno-Ugric family of languages. Estonian population comprises 69% of Estonians, 25% of Russians and 6% of other ethnic groups (Statistics Estonia 2017). The language division is reflected also in the education system. There are two types of schools—one with Estonian as the language of instruction, the other where instruction is mainly conducted in the Russian language.

In order to better understand the origins of Estonian education system, we should take a step back and look at the countries' rather turbulent history.

For centuries, the territory of Estonia has been almost always conquered by some great powers. The start of formal education can be dated back to thirteenth century when German rulers opened the first church schools in Estonia but did not really influence literacy or numeracy rates of the population. More systematic approach to education opened to people only after Reformation, which happened simultaneously in Europe and in the Baltic countries, once Tallinn became part of the Hanseatic trade route. In the seventeenth century the northern part of Estonia became under the rule of Swedish Kingdom, consequently, agriculture and education were organised according to the Swedish model. Some influences of “Good old Swedish times” are present even today. On the Swedish initiative, the academic gymnasiums were opened in Tallinn in 1631 and Tartu in 1630, the University of Tartu was established in 1632. The school in Tallinn, dedicated to the Swedish king Gustav Adolf, is still functioning today and is one of the best schools in Estonia.

The church taught a large part of peasantry children to read from the second half of seventeenth century. The idea that all children should be educated, regardless of their social standing was applied (Ruus 2002). In the beginning of eighteenth century during the Great Northern War with Russia, the territory of Estonia was conquered by the Russian empire. Because of the lasting war and devastating plague, population diminished nearly tenfold, the only university and a lot of schools were closed. However, Baltic Germans were the ones who de facto ruled in Estonia as they had a strong stand in Czar's court in St. Petersburg and Russians lacked administrative power to manage their remote territories themselves. Russification appeared for a short time at the end of the nineteenth and beginning of twentieth century.

The rules and education reforms implemented in the tsarist Russia applied formally also to Estonian elementary schools. According to census in 1897, the level of literacy among Estonians was 80% which was the highest in Russian empire (Lees 2016). The presence of two foreign cultures, German and Russian, encouraged the development of Estonian national identity in the nineteenth century.

Estonia became an independent state in 1918 and introduced free, compulsory and public education for all. The new country quickly implemented prevailing European ideas about the democratic nature of schools, mother-tongue instruction, secondary schools, developing talents of every child, supporting children's initiative and developing extracurricular activities. All this was cut short in 1940 when Estonia together with other Baltic countries was occupied by the Soviet Union. Estonia lost about fifth of its population due to the losses in the Second World War, many fled to the West from the Soviet regime and many were deported to Siberia and never returned.

Education during the Soviet area remained in Estonian, although massive instruction of Russian language as "a language of friendship" was added to the curricula. Undisguised ideology was added even to math and science lessons; however, history and social sciences suffered the most. Foreign languages were poorly taught with minimal hours and learning materials, heavily saturated with Soviet ideology. The goal was to keep people in isolation from the rest of the world and foster the growth of *Homo Sovieticus* as a new species. In schools, strong emphases were put on subjects such as maths and science due to the military needs. The Soviet regime tried hard to keep people in isolation from the West, however, Estonia had a privilege to peek through the Iron Curtain due to its proximity to Finland (80 km). People could watch the Finnish TV and since Finnish and Estonian languages are related, many Estonians mastered the Finnish language independently and followed the life in the West with the help of Finnish television.

The breakthrough for Estonian education was Estonian Teacher's Congress that took place in 1987. The Independence of the Estonian state was re-established in 1991 and Estonian teachers were the voice of freedom four years prior to that. The teachers in 1987 demanded a new, independent, Marxism-Leninism ideology free curriculum for secondary schools, they formed committees consisting of teachers, scientists, university professors, etc. A lot of help was received from Estonians living abroad who were well organised and paid a lot of attention to supporting education in Estonia. Many Estonians whose parents had emigrated during the Second World War, returned and helped in re-arranging the system.

Because of the language similarities and exchanges with Finnish universities, Finnish education system and practices had their influence on processes in Estonia. At that time Finland had already for two decades followed the comprehensive school system. Estonians looked at their curricula, teaching materials and practices and learned from their neighbours.

After intensive work, the curriculum for newly independent Estonia's education was ready and introduced to schools in 1989, two years before the country officially regained its independence.

The curriculum created in 1989 was reformed in 1996. If until then teachers were given quite a detailed description of what they should teach in their subjects,

then after the reform more attention was on what students should know and be able to do (output-oriented curriculum). Teachers were provided with contemporary ideas, popular in European countries, such as competences-based curriculum, general and cross-curricular competences, subject strand competences. This created a bit of confusion and resistance among teachers, but more than twenty years later we can say that it was a very innovative and positive change that was implemented. The next curriculum revision was done in 2002. The national curriculum is updated approximately every ten years and it states the learning outcomes that students should master during different stages of their formal education.

The education system is mostly public, the private schools comprise 11%. The description of Estonian education system is reflected in Fig. 1.

Estonia follows the comprehensive school system and compulsory education, called the “basic education”, lasts from grades 1 to 9. The comprehensive school (ühtluskool in Estonian) is aimed to provide all students with the best education, regardless of their background. The first streaming into academic or vocational tracks takes place after grade 9 when students are 15–16 years old. For some historical



Fig. 1 Education system in Estonia

reason, majority of students, partially encouraged by their families, choose academic (upper secondary) path, around 25% of students opt for vocational education.

Estonian primary education system is based on a strong pre-school education. Around 94% of children attend kindergartens and children start school at the age of 7. Preschools similarly to general education schools follow the state curriculum. The idea is to provide also young children with a playful, competence based, guided and structured plan for their activities. Although children start school at a relatively late age of 7, many of the activities that in other countries are done at school, Estonian children do in kindergarten in a more playful and relaxed environment. Most children know how to read and write when they start first grade at school.

Schools are generally owned by the local municipalities and consequently enjoy quite extended autonomy. All schools can decide on their culture, goals and the focus of studies. They can specialize in science, languages or any other subject. They follow the state curriculum, which is the framework for developing school curriculum as the national curriculum leaves space for the school to develop their identity. Principals can hire and fire teachers, decide on how to allocate the budget and evaluate the needs for teacher training. Teachers decide on the textbooks and teaching methods that they consider appropriate and would like to use in their lessons. There are basically no school inspectors, interference from the state in school matters is only case based, for example, if there has been a complaint on some matter.

It was also decided back in the nineties that all teachers must have a master's degree to work at school. Teachers used to have master's degree equivalent diplomas already in the Soviet times and it seemed only natural that this requirement should remain.

3 What is Estonia's Experience in International Assessment Studies?

It had been slightly more than a decade since the breakoff from the Soviet Union and complete reconstruction of its education system when Estonia started to participate in the international large-scale assessments.

The first international assessment Estonia participated was Trends in International Mathematics and Science Study (TIMSS) in 2003. The results were surprising. Estonian students were seventh in the international rankings and many people assumed that this must be a happy accident. Estonia never repeated TIMSS, but instead joined PISA in 2006. Since then, it has participated in all consequent PISA cycles. The other large-scale international survey Estonia has taken part is the OECD Teaching and Learning International Survey (TALIS) for teachers in 2008, 2013 and 2018. Student readiness as future citizens has been assessed by International Civic and Citizenship Education studies in 2009 and 2016 run by International Association for the Evaluation of Educational Achievement (IEA). Estonia has also been part of the OECD Programme for the International Assessment of Adult Competencies

(PIAAC) which is a survey on adult skills in literacy, numeracy, and problem solving in technology rich environment. In 2018 Estonia participated in the OECD International early learning and child well-being study (IELS) with a focus on 5-year-old children.

The high rankings in PISA kept repeating and scores have been increasing in some areas of assessment. Estonians went through a positive “PISA shock”, their critical nature would not believe that they have “a rather decent school system” and every time before new PISA data release the question in the air is “have we started to fall?”.

4 What Does PISA 2018 Say About the Student Performance in Estonia?

When exploring different education systems, a significant and important factor is to note how much each country invests and spends on its education. Undoubtedly education needs resources, but high level of resources does not immediately result in high student performance (OECD 2019a). The Estonian case in PISA 2018 shows that high results can be achieved with less money. Estonia spends on education 30% less than other OECD countries. Nevertheless, it ranked first among the OECD countries in reading literacy and science and was third in mathematics. The mean score in reading of 523 points was statistically not different from the results of Macau (China) (525 points), Hongkong (China) (524 points), Canada (520 points), Finland (520 points) and Ireland (518 points). If the general message from the OECD is rather pessimistic about little or no improvement in student performance since the beginnings of PISA, then Estonia has shown positive improvement in reading and mathematics and has kept stable (and high) results in science. The improvement in reading is mostly due to the decreasing number of low performing students and the increase of the top performers. Altogether 89% of Estonian students have reached baseline level of proficiency in reading (OECD mean is 77%). Between 2009 and 2018 the share of top performing students (levels 5 and 6) has increased by almost 8 percentage points. The performance gap between boys and girls has decreased from 44 points in 2009 to 31 points in PISA 2018 (OECD mean 30 points). The gender gap decreased in PISA 2015 when the test was moved from the paper to computer-based test. Figure 2 shows the trends of Estonian student performance in reading, maths and science over different PISA cycles.

There has been a slight improvement in mathematics. In PISA 2018, for the first time Estonia ranked right after the high-ranking Asian countries with a score of 523 points, which is statistically similar to the results of Japan, Korea and the Netherlands. 89.8% of Estonian students have reached the baseline level in mathematics and 15.5% of students are top performers on levels 5 and 6. Boys perform slightly better than girls with 9 score points, which is a statistically significant difference.

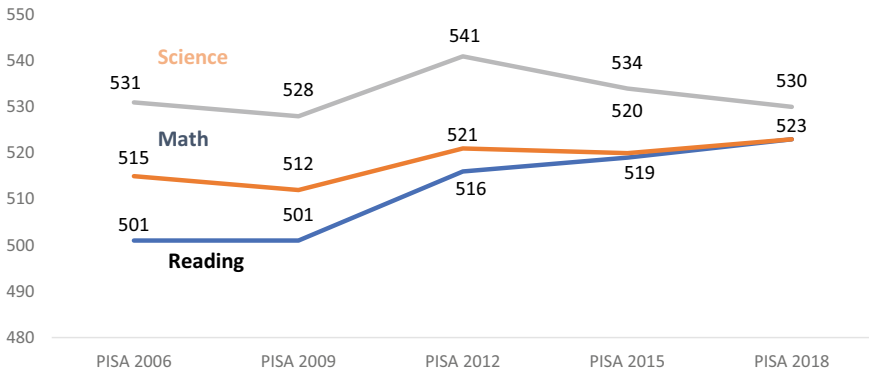


Fig. 2 Performance trends for Estonia

If countries had to pick, which is their favourite domain among the PISA assessed subjects, then Estonians would pick science. For some reason, scores in science have always been higher than in other domains. In PISA 2018 Estonian students scored 530 points in science, which is statistically indistinguishable from results of Japan with 529 points. If in previous PISA cycles, there was no gender gap in student performance, then in PISA 2018, for the first time, there is a statistically significant 5-point difference in favour of girls. Altogether 91.2% of Estonian students have reached the baseline level in science (OECD 78%) and there are 12.2% of students at the two highest levels of performance. This share has slightly decreased and again for the first time there are more girls performing at level 5.

Figure 3 shows the shares of students at different levels of reading proficiency in PISA 2009 and 2018. It can be observed that in PISA 2018 the share of low performing students (below level 2) has decreased, whereas there has been an increase in the numbers of the top performing students (levels 5 and 6).

If the overall picture of Estonian student performance is very positive, then looking more closely in how different groups of students perform, there is some room for improvement. As already mentioned, the Estonian population comprises of 69% of Estonians and 25% of Russians. Both groups of schools are treated equally. They receive funding, based on the same principles, follow the same national curriculum, etc. Through PISA we have learned that there is a considerable gap in achievement between the two groups. Russian schools have a good command of the basic skills and knowledge; however, they are less successful in the application, and PISA is all about the application of knowledge in real life situations. The gap between Estonian and Russian speaking students is 42 points in reading and science, 29 points in mathematics. The gap points out that there are more low performers and less high performers among the Russian students. At the same time the results of the Russian speaking students are above the OECD mean and it is a very good performance. However, in comparison with their peers in Estonian schools the gap persists. In PISA it is often referred that 39 points equals to one year of schooling. Why is there such a gap between the groups? We have tried to research both groups and can

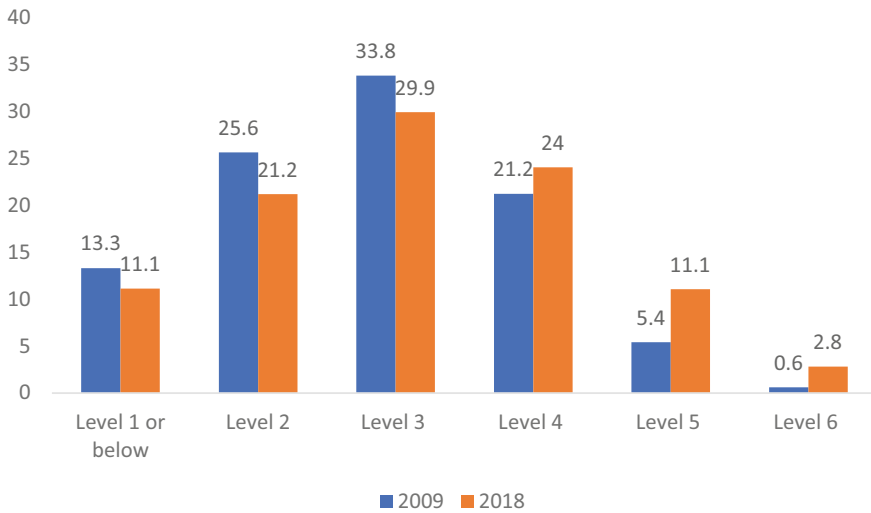


Fig. 3 Percentage of students at different levels of reading proficiency in PISA 2009 and 2018

mark the differences, but it is difficult to name the reason and fix it. There has been a strong suggestion from the political parties that Estonia should not maintain the education system with two languages and merge the schools which would lead to better integration of the society. Currently it has not been done yet.

4.1 Educational Equity

Another important aspect that describes the quality of education is equity. Estonia follows the comprehensive school model, where all students follow similar education path until the end of compulsory education (grade 9). The first streaming to academic and vocational tracks takes place at the end of basic school, when students are at the age of 16. Moreover, grade repetition is not commonly practiced. It is believed that struggling students should be noticed early enough, and they should be helped while they are with the same age group peers.

PISA has consistently shown that Estonian education system manages to provide education for all students regardless of their socio-economic background. In fact, it is uncommon to classify schools according to the composition of the students and divide them as socially advantaged or disadvantaged. Only 6.2% of variation in reading scores could be explained by the student's socio-economic background (12% in OECD). As in all countries, there is a significant difference (61 points) in average reading performance between students with high and low socio-economic background (99 points in OECD countries) (OECD 2019b). Students who come from poor families get financial help, which is granted by the social systems of the state.

All students receive free hot lunch, some schools provide also a breakfast. For some of the disadvantaged students the school lunch might be the only hot meal of the day they get.

In Estonia, education by law is for free, unless the parents decide otherwise and choose private schools for their children. Apart from free services, such as lunch, textbooks, school transport, students get supporting services if needed. Many schools have their own psychologist, speech therapist, social pedagogue as a part of their staff. Smaller schools get those services from a state-run network that extends all over the country. Many students stay at school after the lessons are over. They use classrooms to do homework under teacher supervision or participate in extracurricular activities such as sports, art or computer clubs, which are often free and provided by the school.

Estonian school manages to compensate for what has not been provided to children at home and students from disadvantaged families often achieve high results. In PISA we call them “resilient students”. Altogether 7.4% of students with disadvantaged background have reached the top levels of performance (2.9% in OECD countries). Moreover, 15.6% of Estonian students with disadvantaged background belong to the best performing 25% of students. In fact, the mean score of students from the bottom quarter of socio-economic status is 497 points. This score is above the OECD average and shows that the poorest students in Estonia manage to perform better than the top quarter with the most affluent background in many countries. This proves that if a student is born poor, it necessarily does not have to stay that way and the school system is able to contribute to social mobility, care and develop potentially everybody to high levels. At the same time, 16% of Estonian students from disadvantaged socio-economic background have not reached the baseline level of proficiency, whereas in the OECD countries this share of students is 36%.

4.2 Student Well-Being

PISA is following the general trend of other international and national assessments in paying more and more attention to aspects of school climate, student well-being and learning habits.

In recent years, student well-being has been high in the listing of national policy priorities in Estonia. Therefore the “soft outcomes” from PISA 2018 are analysed with care. Already in two consecutive PISA cycles students were asked the following question: “How satisfied with your life in general are you these days?” Students are given a scale from “one” to “ten”, “one” being the lowest and “ten” the highest level of life satisfaction. Although the general life satisfaction has fallen by 5% in the OECD countries since PISA 2015, Estonian students scored on average 7.19 (7.04 for the OECD countries). Estonia together with Finland, Germany and France show high levels of performance and relatively high levels of life satisfaction. Although 70% of Estonian students are satisfied with their life (67% in OECD countries), 89% feel often happy and 9% always feel sad (OECD 2019c).

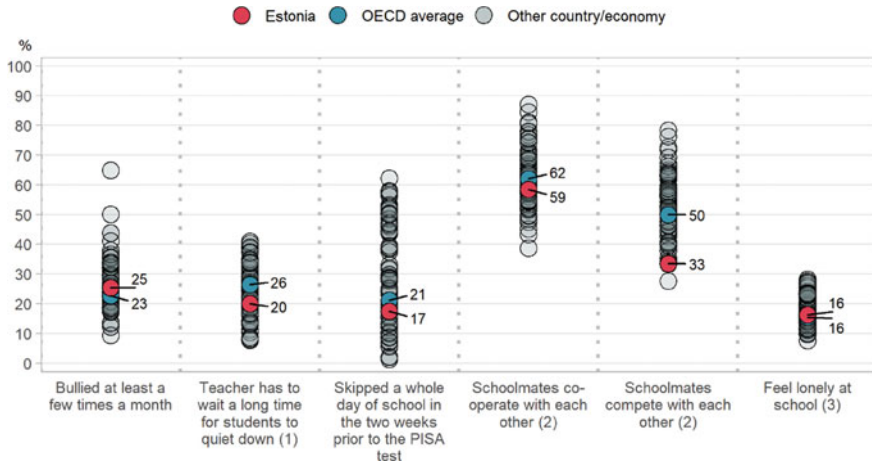


Fig. 4 Estonian School climate (Source OECD, PISA 2018 database)

How is life satisfaction related to reading performance? In general, Estonian data does not show any correlation between life satisfaction and reading performance. However, the data represents an interesting phenomenon. The lowest results in reading are for those students, who reported the highest levels of life satisfaction.

The biggest influence on student’s well-being at school is exposure to bullying. 25% of Estonian students experience some sort of bullying and that slightly exceeds the corresponding levels in other OECD countries (23%). There is better discipline and less skipping school when compared to other countries as can be seen in Fig. 4. Students value cooperation more than competition, but in comparison with other countries, the reported cooperation and competition levels are rather low (see Fig. 4).

5 What is a Teacher in Estonia Like?

In order to understand the key player—the teacher in the education system, we should have a glimpse at TALIS, the OECD Teaching and Learning International Survey. TALIS gives voice to teachers and school principals. The survey studies around 4000 teachers and principals from 200 schools per country. It explores issues about initial teacher training and continuous professional development, provides an overview of practiced teaching methods and different manifestations of classroom climate, etc. It also asks the teachers about their satisfaction levels with their job and how they feel about their profession. Estonia has participated in TALIS survey three times (TALIS 2008, 2013, 2018).

According to TALIS 2018 data, 86% of Estonian teachers are female and with an average age of 49. The average age of teachers across OECD countries and economies is 44. Altogether 54% of teachers (OECD average is 34%) are aged 50 and above

and this is the prime issue of sustainability of Estonian education system (OECD 2019d). Within the next decade there will be an urgent need to renew the workforce. Teachers in Estonia are very experienced and 81% of teachers have received all required formal qualifications. They report positive classroom environment, which is also in good concordance with the student reporting from PISA 2018.

Estonian teacher is inclined to follow more traditional, already well-established practices in their everyday work. Altogether 86% of lesson time is spent on teaching and learning. This exemplifies more effective time management than 78% of average in OECD countries. If over the past five to ten years, in most of the countries the actual teaching and learning time has decreased in about a half of the countries, then in Estonia it has increased with one percentage point. Teachers assess students' progress regularly, but they do not use that much the more modern approaches where students evaluate their own progress. In Estonia student self-evaluation is used only by 28% of teachers, in OECD countries 41%.

98% of teachers and 100% of principals have attended some sort of professional development activity during the year of survey. The areas where Estonian teachers expect to get more additional training are related to ICT skills, teaching in multicultural or multilingual settings and teaching students with special educational needs.

6 What Policy Measures Have Supported Us Along the Way?

Current legal framework of education, as a successor to pre-war existing Estonian Republic legal framework, was established in 1990s after the end of Soviet occupation. The main goal of initiated reforms was to liberate education from the burden of Soviet ideology and set the foundations for modern education system. As already described, the first steps were to find a common understanding of a new curriculum, write corresponding textbooks and retrain teachers. Many of the educational institutions went through restructuring and name restoration/change.

In the mid 1990s the Open Estonia Foundation (allied with Soros Foundation) played an important role in promoting ideas of free and independent schools. They trained school principals and launched the project on quality management system in Estonian schools.

In the 1990s the following main laws on education were adopted:

- The Law on Education of the Estonian Republic (1992) which outlined the general principles in education and its availability.
- The Law on Basic and Upper Secondary Schools (1993) described the grounds of operating and governing municipal basic and upper secondary schools.
- The National Curriculum (1996) provided a framework for all educational establishments regardless the language of instruction. Since then schools are supposed to develop their own curricula, based on the national curriculum. The national

curriculum lists the compulsory subjects with a syllabus and states the number of lessons for each subject. The new curriculum was outcome oriented: it described knowledge, skills, attitudes and values—all together as competences, which were to be mastered in the learning process. The national curriculum was updated in 2011, splitting the curricula for basic schools and upper secondary schools into two different curricula. More detailed attention was designated to development of subject strand competences, general competences (adopted from corresponding EU framework) and cross-curricular competencies that all teachers should include in their subject lessons. The curriculum is updated approximately every ten years.

- In 1997 the external evaluation system was established.

Since education in Estonia is mostly free and paid by the taxpayer, the assumption is that the state needs to know how well the money is spent. The goal of the external evaluation is to see how well students have mastered the learning goals from the national curriculum at different levels of study. External assessments are conducted at the end of grades 3 and 6. The tests cover Estonian language (or Estonian as second language for Russian medium schools) and mathematics. In addition, one other subject is tested in grade 6 on rotating bases in different years. It can be foreign languages, social sciences or science. Externally developed tests with corresponding marking schemes are administered to these grades. Tests are sample based and compulsory for 10% of schools, but since schools find them a valuable tool of quality measurement, they volunteer and administer them to their students outside the compulsory sample. Already for several years the grade 6 tests are computer based. Paper based tests are gradually disappearing. The goal is to have all national tests and examinations transferred to computer-based assessment by 2021.

At the end of compulsory education, in grade 9, students take three exams. Test forms and marking schemes are centrally provided, but the marking takes place at school by the subject teacher. The requirements for finishing the basic school consist of centralized examinations in mathematics, Estonian language, one freely chosen subject by the student (from a list of 10 subjects) and a completed research project organized by the school. The national modal grade for PISA is grade 9 and the test is administered to students approximately one month before they take these school leaving exams.

At the end of upper secondary school, grade 12 students take three centrally set and centrally marked national examinations that are also valid for entering universities or other higher educational establishments. Students should pass national examinations in Estonian or Estonian as a second language, mathematics (two different curricula are offered with different number of learning hours and corresponding exams), and in a foreign language, prepared by foreign examination companies. In addition to centralized examinations, students are required to pass a school exam, and conduct an independent research project in the topic of their interest.

Another school quality control instrument for school strategic development is school internal self-evaluation. It is compulsory and has been in use for more than a decade. During the self-evaluation process schools must analyse their past achievements from many different aspects and set goals for the future.

An interesting project that had a huge impact on Estonian education system and society in general was initiated by the former president of Estonia Toomas Hendrik Ilves in 1996. The project was called “Tiger leap” and its goal was to provide all Estonian libraries and schools with internet and computer classrooms. The prime activity of the project was to provide know how and competence for incorporation of the technology in everyday life (using ID card, internet banking, reading online news and sending e-letters). This overarching project was very helpful in the digitalization of the society and Estonia has managed to achieve remarkable results. Now Estonia is among the most digitalized societies in the world with vast number of online services provided to its citizens. Almost everything can be done online—casting a vote in elections, filing a tax return or registering a childbirth, etc.

The demand for digitally educated citizens has put a significant pressure also on the education system to teach digital competence and this has been a high priority in state policies. Schools have integrated a variety of digital solutions and employed education technologists to support teachers in orienting in the jungle of education technology (educational programs, smartboards, robotic kits) and use the digital solutions in their lessons.

On the state level there are different databases that give access to information about schools and their quality of teaching. Schools have access to free digital textbooks and assessment banks, school administration software. All schools use some version of E-school system for communicating information about student’s achievement, absences, homework, notices and exchanging information between schools and home.

Estonian education system was “upgraded” in 2014 when the government adopted the Estonian Lifelong Learning Strategy 2020 (Estonian Ministry of Education and Research 2014). The strategy document set out five priority areas for development.

1. **Change in the approach to learning** or focus on student centred learning. Each learner at all stages and types of learning should be provided with education that supports their individual and social development. The goal stresses the need to acquire appropriate learning skills, foster creativity and entrepreneurship.
2. **Competent and motivated teachers and school leadership.** This priority focuses on extensive teacher training and evaluation of teachers and headmasters.
3. **Concordance of lifelong learning opportunities with the needs of labour market.** Lifelong learning opportunities and career services that are diverse, flexible and of good quality should result in an increasing number of different age people with professional or vocational qualifications and in increasing overall participation in lifelong learning across Estonia.
4. **A digital focus in lifelong learning.** Modern digital technology is used for effective learning and teaching. The focus is on improvement in the IT skills of general population and grant access to the new generation of digital infrastructure.
5. **Equal opportunities and increased participation in lifelong learning.** Equal opportunities for lifelong learning should be created for every individual.

All goals were described in more detail, and indicators of measurement were attached. Data from national and international assessments intend to estimate if the

goals have been reached. One of the key indicators about the top achievers is used from PISA and if the goal set in 2014 was to have 10% of students to be top performers in reading by 2020 then PISA 2018 shows that the goal has been achieved. There are 13.9% of students at the top levels of proficiency. The goals for mathematics and science were set slightly higher than the PISA 2018 results.

Substantial funding jointly with co-financing of structural funds of the European Union was added to support the implementation of the strategy.

The first goal gears the education away from the more traditional way of teacher centred teaching towards more progressive student-centred educational approach. As we know from TALIS 2018, Estonian teachers use less frequently those approaches if compared with teachers in other OECD countries despite the government priority. This again shows the consequences of autonomy of school system where schools and teachers are free to choose the teaching methods and apply what they feel most appropriate in their teaching.

The strategy document also marks the change from traditional summative assessment towards more child centred formative assessment that should support individual learning and development. According to legislation, the goal of external assessment is to give students, parents, schools, school administrators and the state an objective and comparative feedback to the learning objectives, stated in the national curriculum, as well as provide an input for formation of education policy (Basic and Upper Secondary Schools Act § 34).

To support every student more effectively, the proposed policies suggest using digital technologies more effectively. This complies with the digital focus of the strategy document. The Ministry of Education and Research has launched the development of innovative digital assessment tools. There are mixed opinions about the effectiveness of digital technologies in the educational process. PISA has repeatedly shown that those educational systems that use technology extensively in the study process, show poorer achievement results (OECD 2015). The opinion in Estonia is that ignoring technology and not involving it in the study process is disempowering students from participation in digital society. Digital devices are here to stay. They will be more user friendly in the future and we must learn to use them smartly in favour of student learning. The use of technology could be made more effective with additional teacher training. We see from PISA 2018 data that Estonian students feel that they are digitally advanced, and technology leaves a positive impact on the quality of their life.

Lots of effort and funding is put into the development of computer based “diagnostic tests” that would enable teachers to detect what students already know and what are their gaps in a specific topic or skill. Diagnostic tests are developed in most subjects along with collections of tasks and sets of digital learning materials. The assessment system is moving towards more precise measuring and reporting on subscales and measuring of value added. In addition, non-cognitive tests such as tests in social-emotional and digital competencies have been developed and already administered to several student cohorts. New tests have been developed with the help of experts from universities.

The lifelong learning strategy document recommends the state to collect evidence about student development and wellbeing. As a result, in 2015 the decision was made to create an instrument (questionnaire) to measure student well-being. The same year the theoretical framework was developed, and the survey instruments were piloted in 2016 and 2017. The first full scale student well-being survey was administered in 2018 for grades 4, 8 and 11. In order to widen the picture, teachers and parents of the surveyed students were included in the study. Separate surveys were administered also to pre-schools, vocational schools and academic branch of educational establishments. The goal of the study is to get the big picture at the system level and provide each individual school with comparative reports about the general well-being and school climate in the context of national average. Each school is given a detailed report of indicators, concerning general wellbeing of their students, teachers, and parents. It also points out the problematic areas of the school that need attention (Ministry of Education and Research 2019). Schools use this data as an input for evidence-based self-development and quality improvement. The creation of centralised wellbeing measurement tool has spared schools from inventing their own wellbeing questionnaires and added the quality and comparability dimension.

7 How Are Policy Measures Reflected by International Comparison Studies?

PISA has affected Estonian education in a positive way. It has captured the picture of the education landscape since 2006 for five times. So far Estonia has not changed its education policy measures with a goal to excel in PISA outcomes. All the changes are a strive to improve and advance in line with the demands of the constantly challenging and changing world.

Table 1 is an attempt to summarise different policy measures and explain, how they are mirrored in assessments on national level.

8 In a Nutshell: What is the Secret of Estonian Success?

There have been tremendous changes in Estonian education system since 1990. Several post-soviet countries have asked the question—we all had the same starting platform, what did you do differently? There is no clear-cut answer to this question. Estonia has pursued the system of equity by treating every student equally, regardless of their background trying to provide the best learning conditions for all. Schools have enjoyed a lot of autonomy for decades; they have been very little disturbed by school inspectors. At the same time, there has been a strong strive to improve from within, to provide the best education for each child. This started with the introduction of mandatory school self-assessment as schools have been obliged to evaluate their

Table 1 Estonian education policy measures and their reflection in PISA or TALIS

Policy measure	Reflected in PISA or TALIS
National curriculum in year 1996 determined core and cross curricular competencies. Amendments in 2002 introduced problem-solving skills, social and emotional skills. Updates from year 2011 focused on teaching core competences in all subjects	PISA assesses application of competences in real life settings. The national curricula emphasise cross curricular skills and competences which make students think broader and reflect on what they can do with their knowledge
Nationwide external assessment system was introduced in 1997. Students are assessed in grades 3 and 6 (low stakes). Tests assess math and reading literacy, open-ended responses are used. Since 2017, tests in grade 6 are computer based. High stakes exams in grade 9 to certify the end of basic school	Students are generally familiar with the format of items presented in PISA as they have taken national external evaluation tests during their earlier years of schooling. Modal grade for PISA is grade 9. PISA is administered about one month before the final basic school leaving exams
Comprehensive school since 1990s implements social equity. All students are provided with free school meals, free textbooks, etc. Schools must ensure the best learning environment for all students Grade repetition is not commonly practiced	Small influence of socio-economic background on test results
In 2014, state established a nationwide specialist network for students with learning and behavioural difficulties, where schools can get additional support services of psychologists, social pedagogues, speech therapists	The supporting systems for struggling students might have contributed to the low number of students below the level 2, general share of low performers has also decreased
Amendment to the basic school and upper secondary school acts: <ul style="list-style-type: none"> • New management and funding principles for schools • Limitations to state supervision-only case-based school inspection is practiced, usually after receiving a complaint 	As seen from PISA data schools are very autonomous. They choose textbooks and teaching methodologies. Principals hire and fire school staff, allocate resources for teacher training. Relatively small variation in average achievement between schools, larger variation within schools

(continued)

Table 1 (continued)

Policy measure	Reflected in PISA or TALIS
<p>Lifelong learning strategy 2020, adopted in 2014. Main goals were:</p> <ul style="list-style-type: none"> ● Change in the approach to learning: more progressive, student-centred way of learning should be applied ● Competent and motivated teachers and school leaders ● Digital focus in lifelong learning; move towards digital format in teaching, learning, assessment and publishing of learning materials ● Key indicator (based on PISA) with a goal to increase the share of top performing students in reading, maths and science 	<ul style="list-style-type: none"> ● TALIS 2018 reveals that teachers use their teaching time effectively, mostly using teacher-lead structured instruction. Student-centred methods are also applied, but comparatively less than in other countries ● TALIS reveals that almost all teachers have participated in some sort of teacher training in the year of survey ● Students seem to enjoy the computer-based PISA tests. Tests on computers are not a novelty for them ● The share of top performing students has increased, and the goals for 2020 have been achieved in reading and almost achieved in the other domains. (Goal in reading 10%, in PISA 2018 it is 13.9%; goal in maths 16%, in PISA 2018 it is 15.5%, goal in science 14.4%, in PISA 2018 it is 12.2%)

activities, set plans and visions for the future. Schools have received substantial funding to renovate or build new infrastructure and have created pleasant learning atmosphere.

In Estonia, like in many other countries, the aging of the population is a reality and it has reflected in the decline of the student numbers. In response to that school optimization has taken place all over the country. Some schools have been closed, others merged or reformed. This has created many emotions and put pressure on schools to improve. They must have a vision for future and come up with new solutions in order to survive. Schools have been very active in participating in different projects, in gaining international experience, for example, in getting exchange students and teachers. Teachers have access to in-service courses and training programmes free of charge. They have joined subject teacher networks to be constantly updated on important matters and actively exchange opinions. Schools offer variety of opportunities for using digital technology.

Estonian school curricula are based on the principle that students should have a broad worldview. Apart from languages, maths and science, many schools teach coding and robotics, starting already from the first grade. At the same time, the curriculum includes creative subjects such as music, art and physical education as mandatory subjects for all. Under physical education the school can decide to teach dance to all students or ski in the forest in case there is snow.

All schools have a technology class where they teach their children how to cook in well-equipped kitchenettes and knit a sock or cut woodwork. Usually all those practical skills are mastered by boys and girls alike in mixed groups. Making a dish, where everybody is responsible for certain ingredients and do their fair share during the preparation, is a good example of a typical common project that involves skills like cooperation, problem-solving, and everybody is goal oriented.

Estonians, by nature, are critical towards themselves as well as towards others. Their motto in life is that “it could always be better.” The criticism is probably the driving force for improvement.

Effective education is served with a subtle balance between tradition and innovation, rigor and freedom, group and the individual (Robinson and Aronica 2016). Estonian schools might be slightly more traditional, the cognitive skills as well as the development of the soft skills are very much valued. It is widely agreed that school should be a place where students feel safe, happy, challenged and motivated to become well equipped future citizens. Estonian policy makers have set out to invent the next education strategy for year 2030, which should be more geared towards individual learning paths.

Well, what is the secret of Estonian success? Marc Tucker put it this way: “The fact that Estonia is among the top performers in PISA does not appear to be the result of education policies pursued since Estonia gained its independence, but rather the result of hundreds of years of political, social and educational development which ended up supporting a strong commitment to education as well as a tradition of very high education standards, very demanding curriculum, high quality examinations built directly on the curriculum, highly educated teachers, and most of the other drivers of high performing national education systems” (Tucker 2015).

References

- Basic and Upper Secondary Schools Act*. Retrieved from <https://www.riigiteataja.ee/akt/131122015015>.
- Estonian Ministry of Education and Research. (2014). *The Estonian Lifelong Learning Strategy 2020*. Retrieved from https://www.hm.ee/sites/default/files/estonian_lifelong_strategy.pdf.
- Estonian Ministry of Education and Research. (2019). *Rahulolu haridusega*. Retrieved from <https://www.hm.ee/et/rahulolu>.
- HITSA strategy for 2018–2020*. Retrieved from <https://www.hitsa.ee/about-us/news/hitsa-strategy-for-2018-2020-supports-increase-in-digital-competencies-in-the-society-as-a-whole>.
- Lees, M. (2016). *Estonian Education System 1990–2016: Reforms and their impact*. https://4liberty.eu/wp-content/uploads/2016/08/Estonian-Education-System_1990-2016.pdf.
- Organisation for Economic Cooperation and Development. (2015). *Students, computers and learning: Making the connection, PISA*. Paris : OECD Publishing. <https://doi.org/10.1787/9789264239555-en>.
- Organisation for Economic Cooperation and Development. (2019a). *PISA 2018 results: What students know and can do* (Vol. 1). Paris: OECD Publishing. <https://doi.org/10.1787/5f07c754-en>.
- Organisation for Economic Cooperation and Development. (2019b). *PISA 2018 results: Where all students can succeed* (Vol. 2). Paris: OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>.
- Organisation for Economic Cooperation and Development. (2019c). *PISA 2018 results: What school life means for students' lives* (Vol. 3). Paris: OECD Publishing. <https://doi.org/10.1787/acd78851-en>.
- Organisation for Economic Cooperation and Development. (2019d). *TALIS 2018 results: Teachers and school leaders as lifelong learners* (Vol. 1). Paris: OECD Publishing. <https://doi.org/10.1787/1d0bc92a-en>.
- Robinson, K., & Aronica, L. (2016). *Creative schools*. UK: Penguin Books.
- Ruus, V.-R. (2002). The history of Estonian education—The story of the intellectual liberation of a nation. https://www.estonica.org/en/Education_and_science/The_history_of_Estonian_education_%E2%80%94_the_story_of_the_intellectual_liberation_of_a_nation/.
- Statistics Estonia. (2017). *Minifacts about Estonia 2017*. https://www.stat.ee/publication-2017_minifacts-about-estonia-2017.
- Tucker, M. (2015). *Tucker's Lens: Estonia: Unsung Heroine of the Baltic, but....* <https://ncee.org/2015/03/tuckers-lens-estonia-unsung-heroine-of-the-baltic-but/>.

Gunda Tire is PISA national project manager for Estonia since 2007. She works at Education and Youth Authority which is a government organization under the auspices of the Ministry of Education and Research. Her responsibilities include preparation, implementation and dissemination of PISA survey in Estonia. She has also been involved in other international educational projects. Gunda has been the editor for four national PISA reports, worked with schools and media as well as other stakeholders to disseminate PISA in Estonia and beyond. She has been a frequent speaker at educational conferences in Estonia and abroad, she has also received numerous delegations from all over the world and shared insights about Estonian education system.

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.



Finland: Success Through Equity—The Trajectories in PISA Performance



Arto K. Ahonen

Abstract The Finnish education system has gone through an exciting developmental path from a follower into a role model. Also on the two-decade history of PISA studies, Finland's performance has provided years of glory as of the world's top-performing nation, but also a substantial decline. This chapter examines Finland's educational outcomes in recent PISA-study and the trends across previous cycles. Boys' more unsatisfactory performance and the increasing effect of students' socio-economic background are clear predictors of the declining trend, but they can explain it only partly. Some of the other possible factors are discussed.

1 Finnish School System

It is still possible to identify a particular Nordic political philosophy entrenched in the Nordic model of society. The Nordic model emerges as a composite of two large European models: the Anglo-Saxon model's emphasis on economic liberalism and competition, and the Continental model's emphasis on a large public sector, social welfare and security (Telhaug et al. 2006). In the Nordic countries, social security still exists in the form of well-developed public services and a comprehensive well-functioning education system. The Nordic countries have invested more than most of the other nations in the education sector: the level of education is high, the state school is highly regarded, the principle of equal opportunities is adopted, and school standards are reasonably homogenous throughout the nations.

Basic education in Finland is provided free of charge for all age groups. If a pupil cannot attend school for medical or other reasons, the municipality of residence is obligated to arrange corresponding instruction in some other form. In most of the cases, students with special education needs are integrated on the mainstream classes, and only the students with very severe disabilities study in special education classes. These special education classes are in most cases located on regular schools, and there are only very few (70 in the year 2018) special education schools left as separate

A. K. Ahonen (✉)

Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland
e-mail: arto.k.ahonen@juu.fi

© The Author(s) 2021

N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_6

121

institutions. There are also some private schools in Finland, but the number of them is minimal. Altogether about 2300 schools provide comprehensive basic education. Ninety-five per cent of all schools are run by the communities and financed by the government. Also, the approximately 80 privately run schools accepted by the Ministry of Education receive their funding from the government. Private schools usually follow the general pedagogical core curriculum. Some are international or certain national (German, French, Russian). Also, some of them have a religious character or use a distinctive educational approach such as Montessori-, Freinet- or Steiner-pedagogy.

Compulsory education lasts for ten years, including a one-year compulsory pre-school class for 6-year-old pupils. In practice, all Finns complete nine-year comprehensive education. Following basic education, there are two main possibilities to choose from: upper secondary general education and vocational education, which both last three years. Both alternatives provide basic eligibility to continue studies at the post-secondary level (Fig. 1).

The network of comprehensive schools is supposed to cover the entire country. Free transportation is provided for school journeys exceeding five kilometres. Comprehensive school in Finland is legally one unit. However, due to former governance, it is still often divided into two levels: a lower level at grades 1–6 (primary) and grades 7–9 (lower secondary). Traditionally, class teachers instruct all subjects on the primary level. At the lower secondary level, the teaching is organised by subject teachers, who teach their major subject(s) only. There are also a growing number of comprehensive schools, where all the instruction is given in one school building by one group of staff. Nevertheless, the division on class teachers and subject teachers still exists, and their training is organised on separate programs in the universities.

About 95% of all the pupils that complete nine years of comprehensive school continue in upper secondary education (53% in general upper secondary education and 42% in vocational education). Both streams of upper secondary education are three-year programs, and they produce eligibility to continue on tertiary education. In practice, the majority of university applicants graduate from the general upper secondary schools. Meanwhile, the majority of students completing vocational education enter the workforce or continue their studies at the Universities of Applied Sciences.

2 Finland's Educational Outcomes in Comparison

2.1 Trend Across PISA Studies 2000–2018

According to the PISA 2018 survey, Finland still has a high level of competence in international comparison, as Finland represents the top of the European and OECD-countries together with Estonia (OECD 2019b). The top positions are dominated by the education systems of Asian countries, where the starting point for schooling is

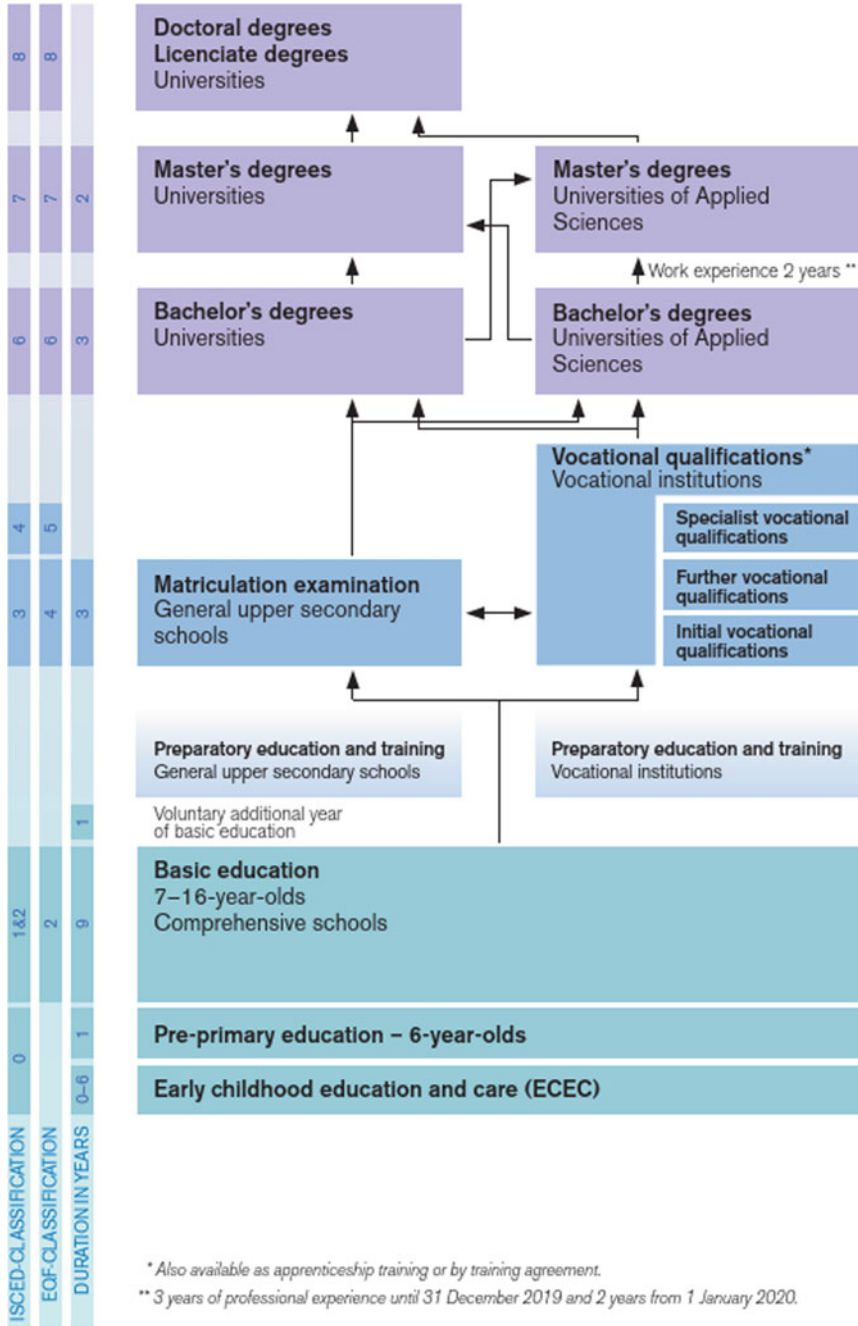


Fig. 1 Education system in Finland (Ministry of Education and Culture 2019)

very different from that of Finland (Sahlberg 2012). Some English-speaking countries, such as Canada and Ireland, run almost parallel to Finland. In the other Nordic countries, on the other hand, competence is lower than in Finland in all other assessment areas except mathematics. In PISA 2018, Finnish 15-year-olds were one of the best in reading literacy (mean score 520) in the OECD-countries together with Estonia (523), Canada (520), Ireland (518) and Korea (514). Among all the countries and economies, Finland was preceded by China's BSJZ (Beijing-Shanghai-Jiangsu-Zhejiang) area (555) and Singapore (549). The average scores of Macao-China and Hong Kong-China were also among those, whose scores did not differ statistically significantly from those in Finland. Finland's mean reading score fell by 6 points compared with PISA 2015, but the change was not statistically significant. A longer-term review also shows that the trend in reading literacy is declining not only in Finland but also in the OECD countries on average. Finland's mean score has dropped by 16 points relative to 2009 and by 26 points relative to 2000.

Mathematical literacy (mean score 507) was in PISA 2018 still well above the OECD average. Finland's ranking was between 7 and 13th among OECD countries and between 12 and 18 among all participating countries and economies. The Finnish average does not differ statistically from Canada (512), Denmark (509), Belgium (508), Sweden (502) and the United Kingdom (502). The European countries that outperformed Finland statistically significantly were Estonia (523), the Netherlands (519), Poland (516) and Switzerland (515). Although Finland's mean score dropped by 4 points from PISA 2015 the change was not statistically significant, so mathematical literacy effectively remained at its previous level.

The performance of Finnish students in science literacy (522) ranked among the third-best in the OECD countries immediately after Estonia (530) and Japan (529). The Finnish score did not differ statistically significantly from Korea (519), Canada (518), Hong-Kong-China (517) and Taiwan (516). Finland's score on science has fallen steadily, dropping by a total of 41 points since 2006 and statistically significant 9 points from 2015.

Compared to the previous PISA assessment in 2015, the average scores in different assessment areas in Finland had decreased statistically significantly only in science. Averages in reading and mathematical literacy have remained at almost the same level since 2012. However, a longer-term review (Fig. 2) shows that there has been a steady decline in Finland since 2006. In the recent PISA 2018 cycle, reading literacy was the main assessment area, which means that the most comprehensive assessment construct was obtained. By comparing the latest results with those years 2000 and 2009 of reading literacy being the main assessment domain, the averages have fallen clearly and statistically significant.

Over 14% of Finnish students had excellent reading proficiency at levels 5 and 6, which was roughly the same as in 2009 (15%). The number of top-performing students on level 6 even rose marginally from 2009, but the change was not statistically significant (Fig. 3). The number of low-performing readers (below level 2) increased by more than five percentage points in Finland compared with PISA 2009 and 2.5 percentage points compared with PISA 2015. Both are statistically significant

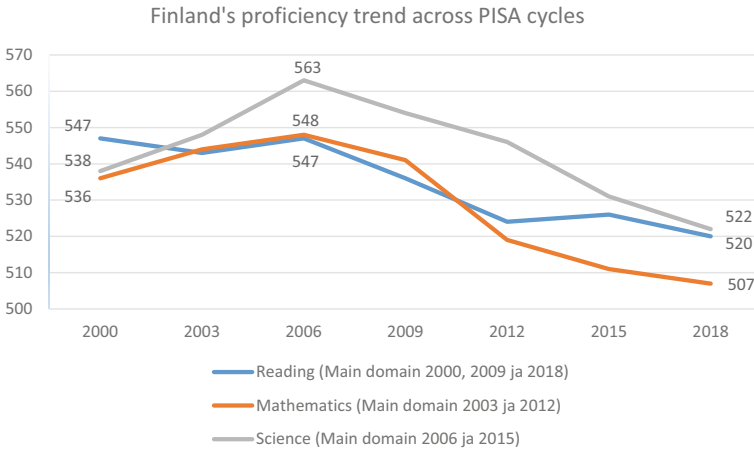


Fig. 2 Finland’s proficiency trends across PISA 2000–2018

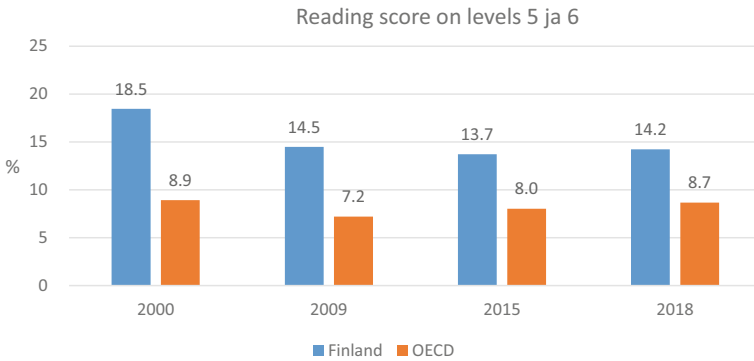


Fig. 3 Percentage of Finland’s 15-year-olds reading score on levels 5 and 6

changes. Level 2 proficiency, also referring to United Nations Sustainable Development Goals, has been identified as the minimum level of proficiency that each child should acquire by the end of their secondary education (OECD 2019a, p. 89). It is a serious concern that there are now, more than ever in the twenty-first century, young people whose reading proficiency is too weak for studying and participating in society. This is the situation both in Finland and across OECD countries (Fig. 4).

2.2 Gender Gap

In Finland, the gender gap in reading literacy performance has consistently been one of the highest in the participating countries. It was one of the highest in OECD

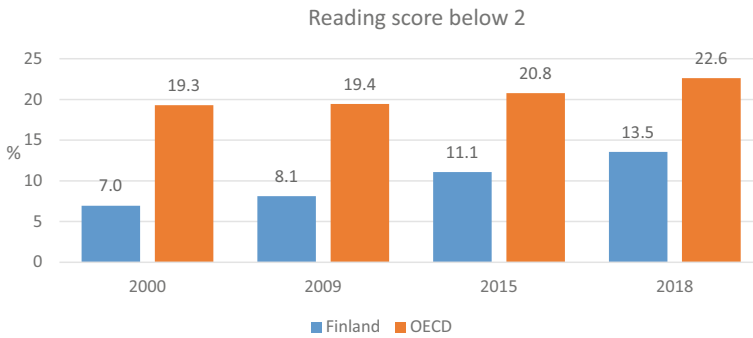


Fig. 4 Percentage of Finnish 15-year-olds reading score below 2

countries at this time too (OECD 2019b). The difference in favour of girls was 52 points, compared with an average of 30 points in OECD countries. Altogether 20% of Finnish girls but only 9% of boys ranked at the highest performance levels 5 and 6 (Fig. 5). Similarly, 20% of boys and 7% of girls were among the poorest performing readers. Among boys, the number of low-performing readers has increased by up to 7 percentage points since 2009, and among girls, the increase has been four percentage points.

For nearly two decades, the reading literacy performance of Finland has highlighted the substantial differences in skills between girls and boys. The difference in reading score among Finnish girls and boys was still the largest in the OECD countries. Also, in science, girls’ skill levels were higher than those of boys since 2009. In mathematics, the average for girls reached boys in 2012, after which girls have

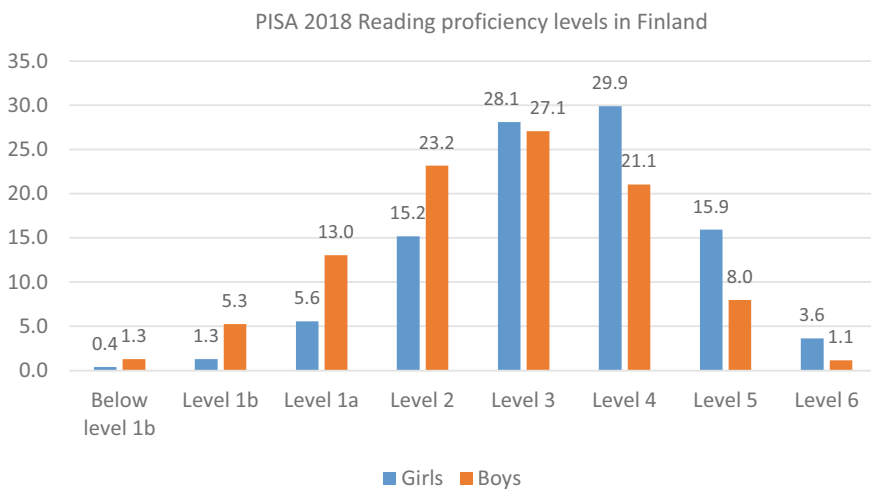


Fig. 5 PISA 2018 reading proficiency levels in Finland by gender

done better than boys in all the domain areas. Also, in science literacy, the gender gap in Finland was the largest in the OECD countries.

2.3 Socio-economic Background

The educational background and occupation of parents and family wealth (socio-economic background) linked to the reading proficiency of students in all participating countries (OECD 2019c). In Finland, the average difference in reading proficiency between the top and the bottom socio-economic quarters was 79 score points. In OECD countries, the corresponding difference was 88 score points. In Finland, the link between students' socio-economic background had become more marked since 2009 when it was 62 points. The poorer outcomes in the bottom quarter can explain this trend at least partly. In 2009, the average reading proficiency in the top quarter was 565 score points, remaining virtually unchanged in 2018 at 562 points. By contrast, the performance of the bottom quarter in 2018 (483 points) was 21 points lower than in 2009 (504 points).

3 Timely Changes, Trends and Explanatory Factors of PISA Proficiency in Finland

3.1 Long-Term Declining Trend

The longer-term decline in proficiency seems to be driven by the increase in the number of weak performers in all assessment areas in Finland. In terms of reading literacy, the share of excellent readers (levels 5 and 6) in the student population has remained unchanged since 2009. However, the share of weaker readers (below level 2) has increased by more than five percentage points. Currently, about 14% of young people in Finland do not reach a sufficient level of reading literacy to be prepared for further studies and life as a full member of society.

The average score of the most highly proficient students in reading literacy decile in Finland has remained practically the same since 2000. At the same time, the average reading literacy score of the lowest proficient decile has declined by about 9 points, which is a statistically significant change. The different development of deciles also reflects the more considerable variation in students' reading literacy scores. The gender gap in skills is also evident when looking at performance levels. There were more excellent female readers than male ones. Similarly, the number of weak male readers was significantly higher than that of the female.

In mathematics, the decline in results has been even. Compared to 2012, when mathematics was the main domain, the decline in mathematics competence is reflected in both a decrease in the number of excellent students (4 percentage

points) and an increase in the number of weak ones (3 percentage points). Since 2003, the average score of the best-performing decile in Finland has fallen by 9 points and the weakest decile by 10 points. When compared between genders, mathematical skills are somewhat equal. However, even in mathematics, a slightly higher proportion of boys were poorly qualified than girls.

In science, the share of top performers has fallen by nine percentage points since 2006, and the share of weak talents has risen to the same extent. As regards the variability of skills, the average drop in the performance of pupils with the lowest decile in science is 16 points since 2006. At the same time, the gap between the best and the weakest decile has also widened. Similarly to reading literacy, in science, the proportion of girls among the best performing decile was higher than the boys. Also, the proportion of boys was more substantial among the weakest performing decile.

For further education, postgraduate studies and working life, it is the weakest performing students who should be most concerned, because their level of competence is not sufficient for further studies and active participation in society. They are in great danger of being marginalised even after the completion of basic education. In light of the current results, the number of weak performers is in danger of further increasing, and a large proportion of them are boys.

3.2 Reading Engagement Strongly Linked with Reading Proficiency

As has been shown in the past in PISA studies, commitment to reading is a significant factor of literacy. Also, other international evaluation studies, such as PIRLS (Progress in International Reading Literacy Study) and TIMSS (Trends in International Mathematics and Science Study) have found an association between engagement and hobbyism and skill levels, whether measured in reading, mathematics or science (see Mullis et al. 2016, 2017; Martin et al. 2016).

Of all the countries participating in the PISA 2018 assessment, Finland was among the three countries where the interest in reading had decreased the most. More and more young people read only if they have to. Indeed, the joy of reading is currently one of the most critical goals in which pupils' parents and society as a whole can be involved. The decline in the interest in reading reflected the fact that the time spent on reading for pleasure was on average reduced. The time spent on reading explained 12% of the variation of reading literacy in Finland and 6% across OECD countries. The results show that even a small amount of daily reading has an impact on young people's literacy levels. The students who reported reading for pleasure half-an-hour daily outperformed those who did not read at all by 60 score points, and those who read one to hours daily outperformed no-readers by 95 score points.

In Finland, engagement with reading explains the variation in outcomes more so than in the OECD countries on average (Leino et al. 2019). In Finland, more students than before reported in PISA 2018 study a negative attitude towards reading. The

number of students who considered reading as their favourite hobby had decreased by nine percentage points since 2009. Correspondingly, the number of students who read only if they had to or only if they needed information had increased by 16 percentage points. In Finland, 15% of boys agreed or strongly agreed that reading was one of their favourite hobbies, whereas the corresponding figure for girls was 36%. In the OECD countries, the corresponding figures were 24% for boys and 44% for girls. What is particularly worrying is that as many as 63% of Finnish boys agreed or strongly agreed to the statement: “I read only if I have to.”

In Finland, reading-related variables as a whole were stronger explanatory factors of reading literacy than the socio-economic background of the pupil (Leino et al. 2019). Across OECD countries, on the other hand, the socio-economic background was stronger explanatory factor than several reading-related variables. Compared to OECD countries, the unique features of the Finnish PISA data were the relatively strong association between persistence, gender and level of reading performance. In Finland, perseverance explained 8% of the variation in literacy. Meanwhile, gender explained 7%. These degrees of explanation correspond to a magnitude correlation of 0.30. In OECD countries, perseverance and gender explained only 3% and 2% of the reading variance, respectively. In Finland, immigrants’ background association with reading literacy was also stronger than in the OECD countries on average. However, only 5% of the variation of reading literacy in Finland was explained by the immigrant background (2% in OECD countries).

4 Well-Being and Equity—The Cornerstones of Finland’s High-Quality Education

4.1 High Level of Life Satisfaction

The subjective well-being indicators of Finnish youth were at a reasonable level. 15-year-olds in Finland were somewhat satisfied with their lives (on average, 7.61 on scale 1–10). In terms of material and objectively measurable factors, Finland is of the wealthiest nations in the world; ahead of us were the other Nordic countries as well as Canada and Australia. When looking at the relationship between life satisfaction and knowledge, Finland stood out from other countries and education systems (Fig. 6). Finland was the only country with high levels of reading performance and life satisfaction. For example, in all Asian countries with high levels of knowledge, life satisfaction was low, and in countries with high levels of life satisfaction, reading proficiency was mostly weak. This begs the question of whether life satisfaction and knowledge are the opposite of a double-edged sword and is Finland only an exception to this phenomenon?

Pupils’ sense of belonging to their school was in Finland at the level of the OECD average, and pupils did not feel that they had much cooperation with their classmates. However, the experience of cohesiveness among Finnish students was strongly linked

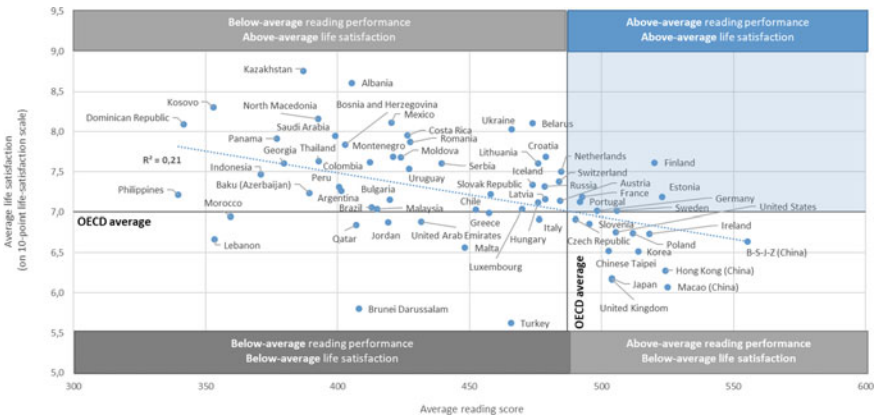


Fig. 6 Life satisfaction and reading performance across education systems (OECD 2019d)

to the experience of cooperation. In other words, working together and encouraging cooperation would increase the experience of cohesion and thus a more meaningful school for all. However, it seems that the happiness of our people, as found in other studies like The World Happiness Report (Helliwell et al. 2019) is also reflected in the lives of schoolchildren. We are knowledgeable and happy in our lives. This is a combination that must be one of the highest goals of all human life. It should be rejoiced.

4.2 Small Between-School Variation

The differences between Finnish school performance have always been small by international standards. The variation between Finnish schools was 7% of the total variation in reading proficiency. The previous represents the least variation among all the participating countries and economies, and it did not increase from the previous PISA survey. Disparities between schools did not increase, but differences in reading proficiency among students within individual schools were more substantial than ever in the history of Finland’s participation in PISA studies.

The differences in proficiency between sub-regions were not significant, but the location of the school seemed to be related to the level of competence. In the schools located in smaller and rural communities, the average scores were lower than in larger ones. What is noteworthy here, however, is that this phenomenon was only visible in the results of the boys, the results of the girls were at the same level regardless of the locality. The phenomenon was initially found in northern Sweden, but then also recognised at least throughout the Nordic countries. Known as the Jokkmokk effect, boys’ lives in the rural area contain values of nature and traditional occupations, which divert interest away from school (e.g. Ripley 2005). Often, the boys also

stay in their home towns. Instead, for girls, going to school often appears to be the only opportunity to pursue their endeavours. Studying also offers the opportunity to move away from home town. This finding suggests that such developments existed throughout Finland.

From the equality point of view, it is a negative result that the socio-economic background of the learner is still as strong as three years earlier when for the first time in the history of PISA, this correlation reached the OECD average (Vettenranta et al. 2016). Previously, the connection had been weaker in Finland than in the OECD countries on average. There has been no change in the gap between immigrant backgrounds and native pupils, and the gap remains the largest in Finland. Although the percentage of pupils with an immigrant background in the Finnish student population has increased slowly, it is still minimal, which is reflected in the small PISA sample size (5.8%) of them. Although there is thus much uncertainty about the results of pupils with an immigrant background, they are indisputably weaker than those of the native Finnish population. However, the result can be partly explained by language gaps and the socio-economic status of the families of immigrant pupils.

The sample of Swedish-speaking schools in Finland was also relatively small (approximately 7%), which makes it challenging to draw valid conclusions. When examining the materials of Swedish-speaking schools, the focus is on the better mathematics performance of the students studying there. The average score of students of Swedish-speaking schools in Finland was the best in the Nordic countries during the PISA 2018 round, and thus also better than those studying in Finnish-speaking schools. However, due to the small sample size, this difference was not statistically significant compared to students studying in Finnish-speaking schools and to Denmark and Sweden. It seems that, in Finnish-speaking schools, pupils' mathematics skills have systematically declined, and pupils in Swedish-speaking schools have maintained their standard. This is especially true for girls in Swedish-speaking schools, their mathematics score in this round was at the same level as in 2003, thus distinguishing themselves from Swedish-speaking boys and Finnish-speaking students on average.

In reading literacy, the difference between Finnish-speaking and Swedish-speaking schools was still significant and better for Finnish-speakers, although it has narrowed slightly from previous years. There has been a slightly steeper decline in the performance level of pupils in Finnish-speaking schools than in Swedish. The reading literacy performance of Swedish-speaking boys has continuously been alarmingly low. Their average score in all PISA studies has been below the OECD average, in every PISA cycle. In science, the gap has narrowed more than in reading: while in 2006, the difference was clear and significant for Finnish-speaking students, no significant difference was observed in 2015. The result was the same in 2018 as well.

5 Discussion

5.1 *Historical Improvement in Learning Outcomes*

Finnish students have received outstanding results in the PISA as well as PIRLS and TIMSS studies. Still, it is good to remember that Finland has not always been on the top of the international comparisons (Altinok et al. 2014). During two decades on the 1970s and 1980s Finnish students' achievement were rated below the global average and the step above the average was taken only as late as mid-1990s (Sahlberg 2011; Sahlgren 2015). By the end of the 1990s, the internal discussion and debate against the Finnish school system got more vigorous. There was a high demand for reforming the school system, claiming the present form was not producing good enough learning results (Simola et al. 2017). According to the many critical voices, the comprehensive schooling had a levelling effect, which gained more unsatisfactory results for all. When the results of the first PISA study appeared in 2001, the results were a genuine surprise for all in Finland. There were also some doubts about the study. Nevertheless, later it can be argued that the PISA study did save the Finnish comprehensive school system as the below citing from the second PISA national full report forewords show.

The outstanding success of Finnish students in PISA has been a great joy but at the same time a somewhat puzzling experience to all those responsible for and making decisions about education in Finland. At a single stroke, PISA has transformed our conceptions of the quality of the work done at our comprehensive school and of the foundations it has laid for Finland's future civilisation and development of knowledge. Traditionally, we have been used to thinking that the models for educational reforms have to be taken from abroad. This sudden change in role from a country following the example of others to one serving as a model for others reforming school has prompted us to recognise and think seriously about the special characteristics and strengths of our comprehensive school. (Väljjarvi et al. 2007)

The latest school reform in Finland was conducted in the mid-1970s. That reform's most significant change was the formation of comprehensive basic education. There was a switch from German tradition towards the Anglo-Saxon model, following especially Sweden. Before that, the students were divided on primary and grammar schools on early ages. Now Finland was the third nation to adopt a comprehensive school system after Sweden and DDR. The first curriculum for the Finnish comprehensive school was prepared carefully by the best expertise of that time, and the reform put into action gradually during the years 1972 and 1977. Shortly after the comprehensive school reform, new legislation for teacher qualifications was established. In 1979 Finland was a world's first nation to set a master's degree as a qualification for all teachers, also at the primary level of education. With that very same system, we are still operating at least through the decade of 2020. The national core curriculum for basic education in Finland has been renewed in approximately every ten years. During the existing history of PISA, there has been only one effective curriculum change in Finland, in the year 2004. The preceding core curriculum was from the year 1994, which gave the schools almost full independence to form their

local curriculum and teaching without any inspection and centralised control. The 2004 national core curriculum was a step towards more restrictive and centralised school policy but still without inspection or standardised testing.

The latest national core curriculum came to effect in the year 2016. The renewing process is split into two parts: a lesson frame and the actual curriculum. The lesson frame is subject to a parliamentary decision process. It is usually challenging to accomplish changes on it, and it did remain as such in the latest curriculum process. The number of lessons and subjects remained practically the same as previously. The curriculum renewal process is led by the National Agency of Education, conducted as office work, and it does need a political decision to come into action. In practice, the 2016 curriculum change did not have any effect on the 15-year-olds sitting the test in spring 2018, because it came into effect gradually.

5.2 Factors of the Declining Trend

The fall of Finland's proficiency since 2006 has been substantial, and it would be crucial to have a hint of the possible reasons behind that. Finland had not been alone on this declining trend, and the absolute fall in average proficiency is greater than the relative performance in comparison with other participating school systems. The average scores from the top year 2006 were so high that even after considerable absolute declines, Finland still ranks among the best of the OECD countries in Science and Reading. In mathematics, the drop in absolute proficiency has been substantial 41 points. Even though there has been a decline also in the other top-performing countries, Finland's performance drop is the greatest of all. Figure 6 shows that gap between this selected list of countries has narrowed along the years. When in the year 2003 PISA study the presented countries' mathematics proficiency varied from Poland's 490 average to Finland's 544, in 2018 study all these countries fit between 502 and 527 country averages. Figure 7 also presents that only in Poland the mathematics average has increased from the year 2003. Estonia has improved since its first participation in the year 2006 study.

Over the cycles, researchers have tried to examine the factors behind the decline, and it has become rather clear that the reasons cannot be found on the PISA data solely (Leino et al. 2019; Vettenranta et al. 2016; Välijärvi et al. 2007). Neither can they be located in the changes in schools, pedagogics or curriculum only. Simola (2014) and Simola et al. (2017) argues that the "Finnish miracle", especially referring on the top results on the first decade of twenty-first century, can be returned on the unique combination of firm beliefs in education, highly valued teacher profession and the pedagogical freedom of teachers without external inspections and testing. In his thorough monograph "Real Finnish Lessons", Sahlgren (2015) found that success is related to cultural and societal changes. Sahlgren (2015) also claims that the best results have been achieved based on the somewhat centralised schooling organisation rather than the de-centralised one. It is also evident that Finland's performance has been higher when the effect of the socio-economic background on students'

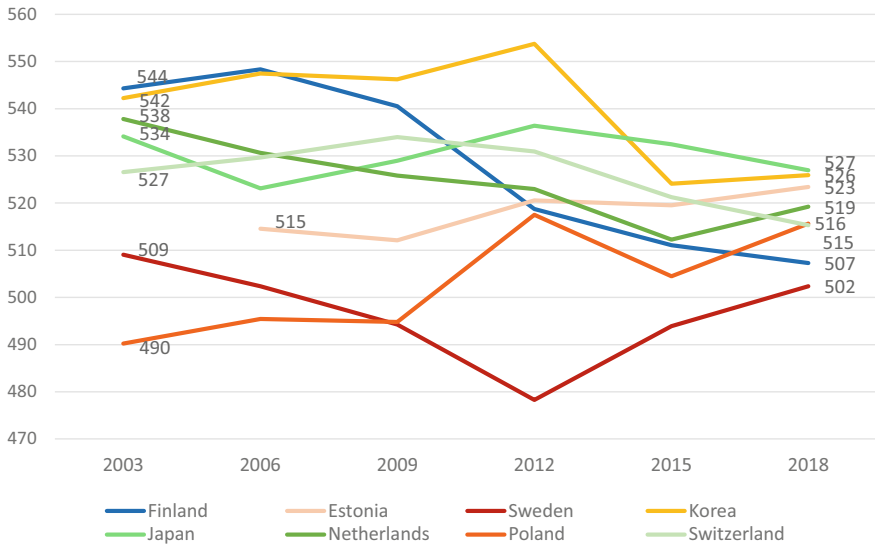


Fig. 7 Trends in PISA Mathematics

performance has been weaker, and along with the performance decline, the impact of the socio-economic gradient has grown stronger. However, we still have very little evidence to prove direct causal effects of Finland's performance trajectory. Still, it is essential to realise as Sahlgren (2015) notes: "Nothing happens overnight". Educational policy decisions and actions, if any, have far-reaching consequences, and the results can be recognised only by looking far enough in history.

References

- Altinok, N., Diebolt, C., & Demeulemeester, J.-L. (2014). A new international database on education quality: 1965–2010. *Applied Economics*, *46*(11), 1212–1247. <https://doi.org/10.1080/00036846.2013.868592>.
- Helliwell, J., Layard, R., & Sachs, J. (2019). *World happiness report 2019*. New York: Sustainable Development Solutions Network.
- Leino, K., Ahonen, A., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., Lämsä, J., Nissinen, K., Nissinen, V., Puhakka, E., Pulkkinen, J., Rautopuro, J., Sirén, M., Vainikainen, M.-P., & Vettenranta, J. (2019). PISA 18 ensituloksia. [First results of PISA 2018. National report in Finnish.] *Opetus- ja kulttuuriministeriön julkaisuja* 2019:40. <https://urn.fi/URN:ISBN:978-952-263-678-2>.
- Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in science*. <https://timssandpirls.bc.edu/timss2015/international-results/>.
- Ministry of Education and Culture. (2019). *Finnish education system*. <https://minedu.fi/en/education-system>. Accessed February 12, 2020.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Retrieved from <https://timssandpirls.bc.edu/timss2015/international-results/>.

- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/pirls2016/international-results/>. Accessed April 9, 2020.
- OECD. (2019a). *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing. <https://doi.org/10.1787/b25efab8-en>.
- OECD. (2019b). *PISA 2018 results. Volume I. What students know and can do?* Paris: OECD Publishing. <https://doi.org/10.1787/5f07c754-en>.
- OECD. (2019c). *PISA 2018 results. Volume II. Where all students can succeed*. Paris: OECD Publishing. <https://doi.org/10.1787/b5fd1b8f-en>.
- OECD. (2019d). *PISA 2018 results. Volume III. What school life means for students' lives*. Paris: OECD Publishing. <https://doi.org/10.1787/acd78851-en>.
- Ripley, A. (2005, March 7). Who says women can't be Einstein. *Time Magazine*.
- Sahlberg, P. (2011). The fourth way of Finland. *Journal of Educational Change*, 12(2), 173–184.
- Sahlberg, P. (2012). Finland. A non-competitive education for competitive economy. In *OECD: Strong performers and successful reformers—Lessons from PISA for Japan* (pp. 93–111). Paris: OECD.
- Sahlgren, G. H. (2015). *Real Finnish lesson. The true story of an education superpower*. London: Centre for policy studies. <https://www.cps.org.uk/files/reports/original/150410115444-RealFinnishLessonsFULLDRAFTCOVER.pdf>.
- Simola, H. (2014). The Finnish education mystery. *Historical and sociological essays on schooling in Finland*. London: Routledge. <https://doi.org/10.4324/9780203068762>.
- Simola, H., Kauko, J., Varjo, J., Kalalahti, M., & Sahlstrom, F. (2017). Dynamics in education politics. *Understanding and explaining the Finnish case*. London: Routledge. <https://doi.org/10.4324/9780203068793>.
- Telhaug, A. O., Mediås, O. A., & Aasen, P. (2006). The Nordic model in education: Education as part of the political system in the last 50 years. *Scandinavian Journal of Educational Research*, 50(3), 245–283.
- Väljjarvi, J., Kupari, P., Linnakylä, P., Reinikainen, P., Sulkunen, S., Törnroos, J., & Arffman, I. (2007). *The Finnish success in PISA—And some reasons behind it 2. PISA 2003*. Jyväskylä: Finnish Institute for Educational Research. <https://urn.fi/URN:ISBN:978-951-39-3038-7>.
- Vettenranta, J., Väljjarvi, J., Ahonen, A., Hautamäki, J., Hiltunen, J., Leino, K., Lähteinen, S., Nissinen, V., Puhakka, E., Rautopuro, J., & Vainikainen, M.-P. (2016). *PISA 2015: Ensiuloksia. Huipulla pudotuksesta huolimatta*. [PISA 2015 First results. National report in Finnish.] Opetus- ja kulttuuriministeriön julkaisuja 2016:41. <https://urn.fi/URN:ISBN:978-952-263-436-8>.

Dr. Arto K. Ahonen is a senior researcher working in the Finnish Institute for Educational Research at the University of Jyväskylä. He holds a position as a Finland's national project manager for OECD's Programme for International Student Assessment PISA-study for cycles 2018 and 2021. He has gained over 14 years of experience in educational policy-related research, analysis and management in two universities comprising positions of research fellow, research planner, project manager, senior research fellow, it-coordinator. His research paper Ahonen & Kinnunen (2015) "How Do Students Rank and Value 21st Century Skills?" was examining students' ideas on the importance of 21st skills. His most recent work on progress is about: "How do PISA proficiency levels, gender and parents' education and occupation predict transition to upper secondary education?" In addition to experience from academia, he has also gained experience as a schoolteacher and principal, making his professional experience reach 20 years. He holds a masters' degree in education from the teacher training program and a Ph.D. in Education from the University of Lapland, Finland.

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.



Poland: Polish Education Reforms and Evidence from International Assessments



Maciej Jakubowski

Abstract Over the last two decades, the Polish education system has been reformed several times, with the comprehensive structural reform in 1999, curriculum and evaluation reform in 2007, and early education reform introduced gradually until 2014. Student outcomes, as documented by PISA, but also other international assessments, largely improved over the last 20 years. Poland moved from below the OECD average to a group of top-performing countries in Europe. This chapter describes the reforms and research on their effects. It also discusses how it was possible to find political support for the reversal of changes that seemed to be highly successful. It provides three lessons from the Polish experience. First, the evidence should be widely disseminated among all stakeholders to sustain reforms. Second, the sole reliance on international studies is not sufficient. Additional investment into secondary analyses and national studies is necessary to develop evidence for better-informed political discussions. Third, some positive changes are more difficult to reverse. In Poland, increased school autonomy, but also external examinations, broader access to preschool and higher education, are among the changes that the new government could not alter.

1 Expansion of General Education as the Overarching Idea of the Polish School Reforms

This paper discusses how Polish education has changed over the last 20 years and the evidence on reform outcomes, mainly from the PISA assessment, but also from other international and national research. It also addresses the complex relations between research evidence, policymaking, and politics, providing some insights into recent changes in the Polish education system.

With the collapse of the communist system in 1989, Poland experienced a rapid transition from a centrally-planned to a market economy in the 1990s. The so-called

M. Jakubowski (✉)

Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

e-mail: mjakubowski@uw.edu.pl

© The Author(s) 2021

N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_7

137

shock therapy introduced in 1990 quickly transformed Poland into one of the fastest-growing economies in Europe. This rapid transition of the Polish economy was accompanied by a successful transition to parliamentary democracy. The first changes introduced in education focused on cleansing the textbooks and curricula of the political content inherited from communist times. Although the system was decentralized very early, still the key decisions remained with the Ministry of Education and central institutions. Only the responsibility for preschool education was transferred, almost entirely, to the newly established local governments—but the results were unsatisfactory due to insufficient resources (Jakubowski and Topińska 2009). It took a whole decade to prepare for more in-depth structural reforms of the Polish education system.

The education reforms started in 1999 and were continued until recently. The goals of the 1999 reform were to improve the quality of education and to increase educational opportunities for all students, but politically it was supported by those who wanted to break once and for all with an educational system inherited from the communist times. The reform also responded to the changing economy and increasing demand for skilled workers, which was driven by the fast-growing economy and increased integration with Western European markets.

For politically minded commentators, the education reforms in Poland were inconsistent and did not lead to substantial improvements. However, more careful analyses of policy objectives and outcomes suggest the opposite. Despite differences in opinion and the usual politics, the education reforms had one overarching idea behind them: to expand comprehensive education so as to provide learning opportunities for all students. The structural reform of 1999 replaced 8 years of primary school with 9 years of comprehensive education in primary and lower secondary schools. The curricular reform of 2008 introduced a new requirement for all vocational schools to cover at least a one-year equivalent of the core subjects taught in academic schools. In a way, it also completed the reforms begun in 1999 through the introduction of a consistent curriculum emphasising key competencies from preschool up to the end of upper secondary education. Finally, the reform of early education started in 2009 and was continued until 2015. It introduced compulsory education for 5-year-olds and extended the right to a preschool education to 3- and 4-year-olds. Overall, the reforms expanded the length of compulsory comprehensive education from 8 to 10 years. In 2015, general education started at the age of 3 and continued until the age of 16. Unfortunately, in 2016 these reforms were in large part reversed and now the period of general education is again shorter.

International assessments document large improvements in student outcomes over the last 20 years. In PISA, Poland has improved its performance from below the OECD average level to above-average performance. The latest results from TIMSS and PIRLS also show improved outcomes in primary education. Finally, the PIAAC assessments of adults show that only the youngest cohorts perform at or above the OECD average. In Europe, Poland is currently among the top performers in international assessment rankings; however, as we will discuss in this paper, this evidence has not been sufficient to convince those who dislike the changes introduced over the last 20 years.

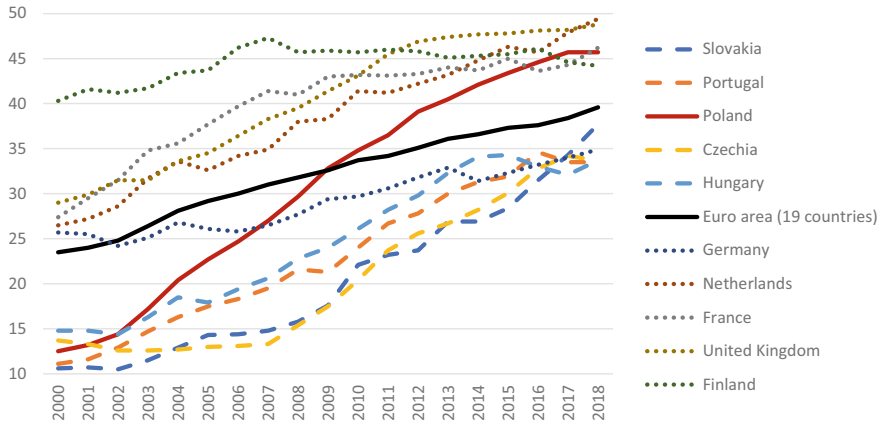


Fig. 1 Tertiary educational attainment—the percentage of the population aged 30–34. *Source* Eurostat, indicator SDG_04_20

Most importantly, the reforms reached the goals that were set in the 1990s, when they were planned in response to rapid structural changes of the economy and society. Many more students now go to general education upper secondary schools or general-vocational schools, which also provide access to higher education. Fewer go into basic vocational education or stop there without continuing on to higher secondary or tertiary degrees. In 1990, more than one-third of students went into basic vocational education. Now the figure is less than 10%.

The goal of the reforms was to encourage as many students as possible to continue education and to open a way to tertiary education for them. This was entirely successful. Figure 1 compares tertiary education attainment across selected European Union countries. In 2000, only 12.5% of Polish 30–34-year-olds benefited from having a tertiary degree. This was similar to the proportion of young people finishing tertiary education in other Eastern European countries (e.g., Slovakia, Czechia, Hungary) or to Portugal. However, tertiary attainment was two times higher in EU countries like the Netherlands, France, the UK, or Germany, and around 40% in Finland. Between 2000 and 2018, Poland experienced the largest increase across the EU in the proportion of young people obtaining a tertiary diploma (by 33 percentage points).

While there exists a widespread notion that an expansion of tertiary education is associated with a lowering of its quality, the market premium for a tertiary education diploma in Poland is comparable to the average across the EU or OECD countries. The rapid increase in the supply of people with tertiary degrees decreased the wage premium and increased the variability in salaries (see Gajderowicz et al. 2012), but the market valuation of these diplomas is still high. According to the OECD data, in Poland and across the OECD countries, earnings of 25-to-64-year-old adults with tertiary education are around 1.5 times higher compared to those with an upper secondary education (2017 data; Education at a Glance 2019, Table A4.1, OECD).

Young adults with tertiary degrees are also less affected by economic shocks like the crisis in 2009 and have much higher employment rates.

The direction for the reform of the Polish education system that was set in 1999 is not continued by the current government, which in 2016 reversed the reform that extended comprehensive education, while it promotes vocational education and limits its support for preschool education. This paper discusses evidence from international assessments on how this affected school reforms, but also the fact that the evidence from international and national studies was not sufficient to stop the reversal of the reforms.

2 PISA as a Necessary but not Sufficient Tool for Policy Evaluation in Poland

Before the first PISA study was conducted in 2000, there was not a single standardized assessment conducted in Poland that measured student knowledge and skills. The national examinations and university entrance exams were not standardized. The exam at the end of secondary education has the same set of questions for all students, but the results were evaluated differently in each school. Entrance exams for higher education varied between institutions and even between departments of the same university. The only international assessment in which Poland participated was IALS, in 1996, which documented the low level of adult skills in Poland at that time. That was not surprising considering that our society and economy were still in the process of transformation away from the communist system, in which access to higher education was limited, and the economy relied in large part on manufacturing and manual labour.

The results of PISA 2000 were not a surprise as most people expected to see a lower performance of students as compared with much more developed OECD countries. These results showed a dramatically low level of reading skills among students of basic vocational schools, with around 80% of them scoring below the basic proficiency level in PISA (Level 2). The results for academic upper secondary schools were much higher. This between-school variance, i.e., achievement differences between schools, was one of the highest across the OECD countries and similar to the results for Germany.

The PISA study became the main evaluation tool for the 1999 reform. This was the only standardized assessment providing comparative data at that time. Also, the implementation of PISA coincided with the implementation of the reform. Polish students tested in PISA 2000 were from the last cohort in the old structure of the school system. These students were still in unreformed upper secondary schools and had received only eight years of comprehensive education. The students tested in PISA 2003 were the first cohort that would go through a full three years of education at the newly established lower secondary schools, so this was the first cohort that

could benefit from the educational reforms. Thus, the comparison between PISA 2000 and PISA 2003 was the main evaluation tool for the 1999 reform.

Similarly, the extensive curriculum changes and introduction of school evaluation systems coincided with PISA studies between 2009 and 2012. Students tested in PISA 2009 were still following the old curriculum, while those tested in 2012 benefited from the new curriculum for the last years of their education prior to the test. School evaluations were also launched around that time. In the future, PISA will be used to evaluate the last changes, which reversed the 1999 structural reform. The latest PISA 2018 covered one of the last cohorts of students who followed nine years of comprehensive education and who studied in the lower secondary schools, which have now been liquidated. In PISA 2021, we will see what the results are of Polish students who have been through the new-old system with a shorter comprehensive education.

Although PISA is the only international assessment in Poland that is conducted repeatedly since 2000, the other international assessments provide helpful insights into the school system. PIRLS and TIMSS assessed 4th-grade students in 2016 and 2015, respectively (PIRLS was also conducted in 2011 in Poland, but for 3rd graders only). Both showed good results in reading, mathematics, and science, placing Polish students among the top performers in Europe. At the same time, the additional results confirmed other, less positive findings of student attitudes towards school and a sense of belonging. PIAAC, the only international study of adult literacy, showed that the youngest cohort tested in 2011 was the only one that performed at or above the average for the European Union. Also, in this study, the results for numeracy (mathematics) were lower than for literacy (reading). The PIAAC results suggested that only some students of higher education institutions were able to improve their skills after school. Differently than in other countries, the relative performance of adults started to diminish at a younger age, suggesting that the key competencies measured in PIAAC were not sufficiently developed after the completion of schooling (see Rynko 2013).

Figure 2 shows how the international assessments could be used to evaluate learning outcomes for student cohorts, which did or did not benefit from different education reforms. Obviously, the outcomes of international assessments only showed the association between student performance and reforms. Also, the reforms were quite complex, and it is difficult to disentangle various policies to show how they affected students. Finally, in a fast-changing economy and society like the Polish one over the last 20 years, there are numerous factors that could also influence student achievement. In the next section we will discuss how data from international assessments can be used to evaluate some reforms, especially those of 1999. We will also discuss how other sources of information, e.g., labour market data, can be used to complement these analyses and give insights into causal relations between policies and outcomes.

The collective evidence from large-scale international assessments demonstrates substantial improvements in student outcomes since 2000 and the implementation of the major 1999 reform. In the next section, we will review research using PISA, but also labour market data, using econometric methods for estimating the causal impact

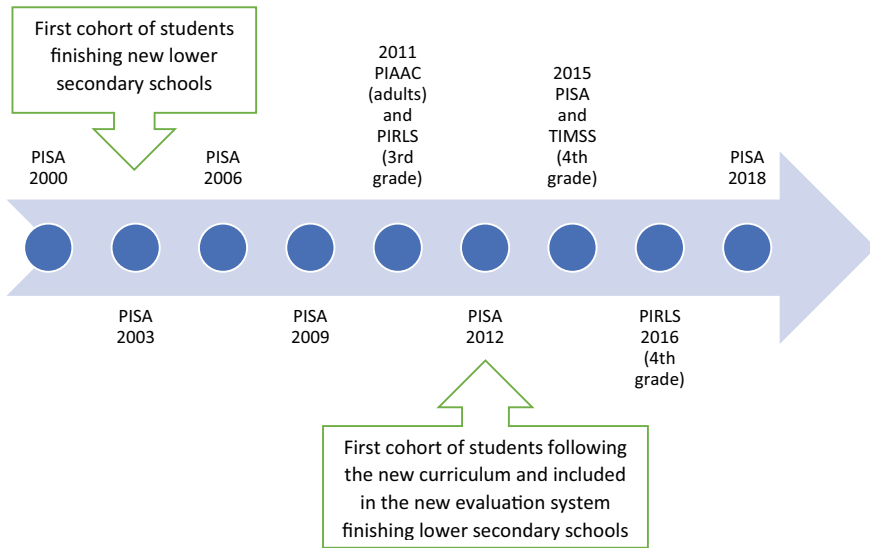


Fig. 2 Key international assessments and major reforms of the Polish school system

of policy changes. Overall, this research provides evidence that the 1999 reform had a sustainable impact on student achievement and later life.

The 1999 reform was planned using examples from other countries and the general notion that comprehensive education is of growing importance in modern economies. At that time, Polish experts and policymakers recognized these needs but were also politically motivated to break with the old system. The reform was implemented alongside other large reforms of the health system, pensions, and local administration. The education reform was implemented rapidly, without the agreement of the trade unions. This left many people with the impression that the reform was not well-prepared and had not been sufficiently discussed with key stakeholders. This might be a reason for the growth in the negative view of the old system that resulted in popular support for reversing the reforms in 2016.

The next large education reforms after 1999 were implemented between 2008 and 2009 and were also largely motivated by experience with international assessments and the opening of discussions about modernizing curricula and teaching methods. Part of the ministerial team that prepared the new curriculum reform took part in the implementation of PISA in Poland. The new curricula focused on expected learning outcomes rather than on the detailed description of what subject content teachers should cover. They also introduced cross-subject topics, emphasized applications, and left to teacher decisions about how some topics should be arranged over time. These changes also emphasized teacher autonomy, leaving more room for teachers to develop individual teaching programs and the use of various materials. The reform of school evaluation was not driven by international assessments, but the team implementing this reform used examples from other countries to build a national evaluation

framework and to plan the implementation of the school evaluation system. Finally, the reform of early education was driven by international comparisons showing that the simple fact that Polish students began school later and that too few children, especially in rural areas, participated in preschool education. Overall, the reforms were largely driven by international comparisons and expertise.

The knowledge of international assessment results and reliance on international expertise was, however, limited to a group of experts and researchers closely connected to the Ministry of Education (see Bialecki et al. 2017). The popular view of the reforms was driven by sentiment and politics in a way that is now common across different countries and in different policy areas. Opinion polls showed that new lower secondary schools were not very popular among older people and among those with lower education degrees. Among the young people who actually finished these schools, the opinion was more balanced, with a majority opting for keeping them. The lower secondary schools were blamed for behaviour problems and the “theory” of the negative impact of putting teenagers in separate schools became popular, despite the lack of any evidence to support it. In fact, international and national surveys showed that behaviour issues were not that common in these schools and did not increase after their implementation. The evidence on learning outcomes, student opinions, and surveys of behaviour problems had, however, limited impact on popular opinion. This opened up the political possibility of changing the system again, despite the overwhelming evidence but in line with popular sentiments and opinions.

We will now review the evidence used to evaluate the Polish reforms before discussing how it was possible to reverse the reforms despite the clear evidence of their success.

3 Polish Education Reforms and Evidence on Their Outcomes

3.1 PISA Results for Poland

The average performance of Polish students in PISA has improved since 2000 by more than 30 points in reading, which is one of the largest improvements across the OECD countries. Poland is the only country that has improved its performance to a level close to the best performers in Europe. Most other countries that have substantially improved their performance, e.g., Chile, started from a relatively low level and are still below the OECD average. From the national perspective, improvement in mathematics is especially significant as, until recently, international assessments like PISA and PIAAC demonstrated relatively lower numeracy skills when compared to literacy. Since 2003, mathematics performance has improved in Poland by around 25 points, and now the results in all subjects are similar.

Changes in the average PISA performance of Polish students are shown in Fig. 3, which documents the progress in all three domains measured in PISA: reading, mathematics, and science. It compares average performance in Poland to the current OECD average, which is slightly below 490 points (487 points in reading, 489 in mathematics, and 489 in science). The results are compared with the first assessment from which reliable trends can be established. Reading has been compared since 2000, but mathematics since 2003 and science since 2006. The results in mathematics and science in 2000 were similar to those for reading, but the assessment frameworks have changed, and the test scores are not directly comparable.

PISA 2018 data provide a unique opportunity to compare reading achievement across seven editions of the PISA survey and between 2000 and 2018. The most reliable comparisons are between 2000 and 2009 and 2009 and 2018. In every edition, PISA measures one domain in a more detailed way, meaning that all students answer questions in that domain and that it is covered by a large number of test questions (more than 100). Reading was the main domain in 2000, 2009, and 2018. Thus, it is possible for the first time in PISA to compare achievement changes across two longer periods. It should be taken into account, however, that the reading assessment framework has slightly changed and that assessments in 2000 and 2009 were conducted on paper, while the 2018 assessment was computer-based (see OECD 2019 for details on the assessment framework and the reliability of the measurement of changes over time).

Poland was one of a few countries that made significant progress in learning outcomes between 2000 and 2018. Moreover, when looking at the low-achieving students, Poland experienced the most substantial improvement across the OECD countries. Figure 4 shows that across the OECD countries, the performance of low-achieving students declined between 2000 and 2018 by around 10 points. Most of this

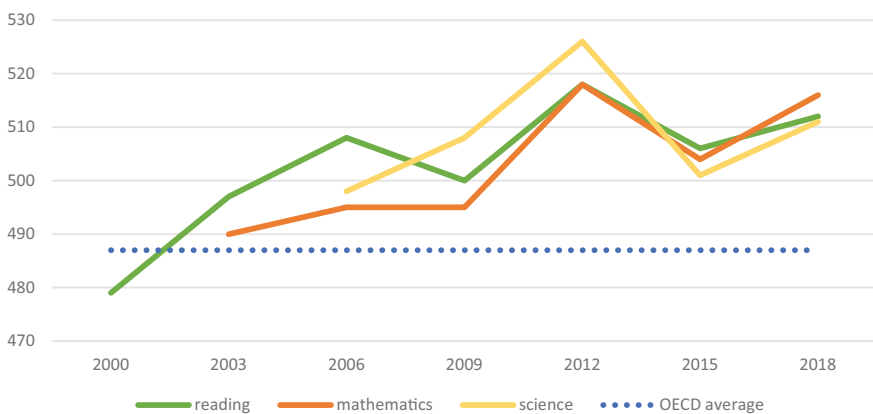


Fig. 3 Achievement trends in PISA average performance in Poland. *Source* OECD 2019

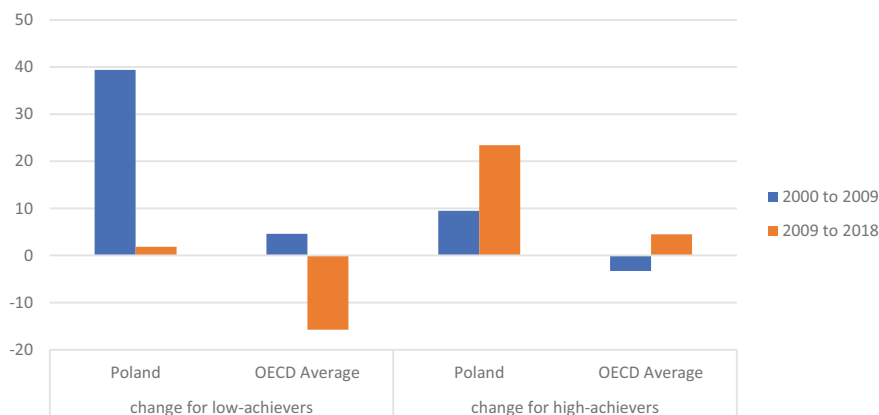


Fig. 4 Changes in reading performance for low- and high-achieving students in Poland and across the OECD countries. *Source* Table I.B1.13, OECD 2019. Performance of low- and high-achieving students is measured as the 10th and 90th percentile of the performance distribution, respectively

decline was between 2009 and 2018. In Poland, the opposite happened. The performance of low-achievers improved by more than 40 points, and most of the improvement occurred between 2000 and 2009, while there was no change in performance between 2009 and 2018.

The performance of the high-achieving students also improved in Poland, but mostly between 2009 and 2018. Figure 4 shows that across the OECD countries, the performance of high-achievers did not change between 2000 and 2018. A small decline between 2000 and 2009 was counterbalanced by a modest improvement between 2009 and 2018, and the overall change is insignificant. In Poland, the performance of high-achievers improved by 9 points between 2000 and 2009, which was followed by a 23-point increase between 2009 and 2018.

To sum up, the key findings from PISA regarding performance trends in Poland are:

- An overall large improvement of the average performance of Polish students in reading, but also in mathematics;
- A large improvement among low-achievers between 2000 and 2009 (in fact mostly between 2000 and 2003);
- A small improvement between 2000 and 2009 among high-achievers, followed by a large improvement between 2009 and 2018.

These results should be seen from the perspective of the overall decline in the performance of low-achieving and average students across the OECD countries and stable results for the high-achieving students. We will now discuss the details of the Polish reforms and how evidence from PISA can be used to evaluate them.

3.2 *The 1999 School Reform and Evidence on Student Outcomes*

The foundations for a modern, effective school system were established by the 1999 reform, which revolutionized the situation in Poland. It not only modified the structure of schools but also increased school and teacher autonomy, freed-up the textbook market, introduced standardized national exams, changed the professional development scheme for teachers, introduced a new financing system that allocated funds according to the per-pupil formula, further decentralized the system and initiated changes in curriculum.

The 1999 school reform had three primary goals:

- (a) To improve teaching quality
- (b) To increase educational opportunities
- (c) To improve efficiency.

The reforms achieved all these goals in the ways anticipated by their authors. Teaching quality was improved, mainly through the introduction of a core curriculum and increased teacher autonomy. Transparency and efficiency of education spending as improved through a formula-based system of resource distribution. The establishment of new, better-equipped, lower secondary schools in rural areas narrowed performance differences between schools. Finally, the reform improved access to tertiary education, and the number of higher education students started to increase rapidly.

The most revolutionary change was related to the restructuring of the school system. The eight-year basic primary education was limited to six years, followed by three years of comprehensive lower secondary education. The selection between different educational programmes was thus postponed by one year, to the age of 16. In this new system, all students followed the same curriculum for 9 instead of 8 years. Upper secondary education was shortened by one year. As before the reforms, only the basic vocational school does not provide direct access to higher education.

Figure 5 compares the system before and after the 1999 reform, in a transition period from 2008 to 2015 (showing the targeted system at the end of the changes), and after 2016 with most changes reversed. The system introduced after 1999 extended the period of comprehensive education by one year and later introduced a compulsory “zero class” for six-year-olds. Thus, before 2008 it had already added two additional years of compulsory education with the common curriculum. Furthermore, the 2008 reform made it obligatory for basic vocational schools to cover one year of the common curriculum, adding one additional year of general education. Later, compulsory education for five years olds was introduced, and the school starting age was lowered to six. Finally, the government introduced new regulations that guaranteed places in preschool education for 3- and 4-year-olds. Since 2016, the lower secondary schools have been abolished, the upper secondary school curriculum has changed, and compulsory preschool for 5-year-olds has also been abolished.

Age	3-4	5	6	7	8	9	10	11	12	13	14	15
Before 1999	Voluntary preschool with no government guarantees		"0"	Primary								
After 1999	Voluntary preschool with no government guarantees		"0"	Primary						Lower secondary		
Changes 2008 to 2015	Place guaranteed	Compulsory	Primary						Lower secondary		One year of common curriculum	
Since 2016	Place guaranteed	"0"	Primary school									

Fig. 5 Changes in the provision of preschool education and compulsory education with the general curriculum (in green)

Overall, when looking at Fig. 5 it is clear that until recently, Poland has consistently followed the path of expanding general education. While before 1999, students had only eight compulsory years of the same general program, those who started education in 2015 have had a right to preschool education from the age of 3 and will be obliged to follow the same general program from the age of 5 until the age of 16. This makes 11 years of general education material obligatory for all students. Using all the opportunities provided from the age of 3, students will benefit from 13 years of general education rather than the 8 available in the 1990s. It is also worth noting that compulsory education in Poland ends at the age of 18, so all students are also required to continue education until this age.

The 1999 reform introduced external national examinations. The first exams were launched in 2002 and monitored student outcomes at the end of every stage of education (they were conducted after the 6th, 9th, and 12th/13th grades; since the abolition of the lower secondary schools they are now conducted at the end of primary and secondary school). The exams are standardized, so all students answer the same questions, and their results are evaluated centrally to assure fair and comparable judgments. Individual results are available to all students and teachers. The results at the school level are also available to the public. Based on the exam results, the so-called value-added measures of student progress in the lower and upper secondary schools were developed and are now publicly available. The external examination system creates incentives for the improvement of teaching quality. It is quite difficult to punish or reward teachers for the exam results, as they are not linked to individual teachers but to schools only. On the other hand, the results are publicly available, which creates social and political pressure to achieve good outcomes at the school level. The results are also important for lower and upper secondary students as they decide whether students will get to better secondary or tertiary programmes.

Altogether, the external assessments of learning outcomes and the large degree of autonomy enjoyed by Polish schools in terms of teaching seem to provide the right mix of freedom and external monitoring. The market for textbooks was already established in the 1990s but was regulated by the Ministry. School autonomy was extended in 1999 by allowing teachers to decide themselves which textbooks and

teaching methods to use. The reform also introduced a new system of teacher professional attainment with four levels. The system created incentives to participate in professional development and was also used to increase teacher salaries, as every level comes with better remuneration.

Finally, the 1999 reform changed the governance and financing system. This was further decentralized, with the ownership of schools transferred to local governments, and a new per-student formula to distribute resources was introduced. The reform introduced proper incentives for the rationalization of school networks, which helped schools survive the substantial demographic decline that started in the 1990s. It also increased overall efficiency, as local governments were better at managing schools and their finances than the central government agencies. Currently, local governments are partly responsible for financing education, although most of the funds are still transferred from the central budget, and teachers' salaries are still in large part regulated centrally.

3.3 Outcomes of the 1999 Reform

We have already discussed large improvements in student achievement in Poland, as documented by PISA performance trends (see Figs. 3 and 4). More in-depth studies suggest that these improvements are associated mainly with the extension of comprehensive education by one year, which benefited mostly the former basic vocational students and to lesser degree students in vocational secondary education. We will now review the results of an analysis of PISA results that shows large benefits for low-achieving students. Other studies include research using labour market data that follows the cohorts affected and unaffected by the reform. We briefly discuss the results of these studies at the end of this section.

The variations created by the policy change of 1999 can be used to see how the reform affected the reading skills of 15-year-olds. Jakubowski et al. (2016) used a difference-in-differences model that compares the change in test scores of the likely vocational school students who were able to study the general academic curriculum because of the reform. The group of "likely vocational students" is constructed, using the propensity score matching method, by comparing 2003 comprehensive school students with similar characteristics (e.g., gender, socio-economic background) to students who were in vocational schools in 2000. This quasi-experimental approach attempts to show a causal link between the reform and student outcomes, mainly for those who without the reform would likely go to vocational schools directly after primary education.

Table 1 presents the actual results from the PISA 2000, 2003, and 2006 studies and counterfactual averages constructed from samples of matched students. The overall achievement of Polish students increased significantly between 2000 and 2003, with additional improvement between 2003 and 2006. The most interesting question is, however, whether the reform affected students in general secondary, vocational secondary, and basic vocational schools similarly.

Table 1 Factual and counterfactual scores of students in different upper secondary tracks

Reading achievement	PISA 2000 factual weighted score	PISA 2003 factual weighted score	PISA 2003 matched counterfactual score	PISA 2006 factual weighted mean score	PISA 2006 matched counterfactual score
All schools	479.1	496.6	483.1	507.6	504.8
Basic vocational	357.6	–	453.3	–	473.5
Vocational secondary	478.4	–	478.5	–	498.2
General secondary	543.4	–	516.4	–	532.0

Source Jakubowski et al. (2016)

Table 2 compares the score improvement among 2003 and 2006 15-year-olds likely to have gone to different types of older secondary school in 2000. In other words, these estimates assess trends in performance for all students and across groups of students who, without the reform, would be in different secondary tracks. Again, there is an overall improvement in average performance among 15-year-olds in Poland. The score improvement for all students is remarkable, around 26 points from 2000 to 2006. Crucial estimates concern the hypothetical performance improvement from 2000 in different tracks. Performance improvement for potential students of former basic vocational schools is simulated to be slightly below 100 points from 2000 to 2003 and 116 points from 2000 to 2006. This is more than one standard deviation of PISA scores in OECD countries, which is a dramatic improvement. These estimates are statistically significant, supporting the hypothesis that 15-year-old students who, without the reform, would have been placed in vocational tracks benefited greatly from the reform. However, the benefits for students in other tracks are not so evident. Students in vocational secondary schools have similar scores in 2003 and improved scores—by 20 points—in 2006. Students in the general track would potentially have lower scores in 2003 and similar performance in 2006.

Table 2 Propensity-score matching estimates of score change for students in different upper secondary school tracks

Reading achievement	Score change: PISA 2003—PISA 2000	Score change: PISA 2006—PISA 2000
All schools	3.9 (5.2)	25.6 (5.1)
Basic vocational	95.6 (8.4)	115.9 (7.1)
Vocational secondary	–5.5 (7.8)	19.7 (7.5)
General secondary	–27.0 (7.6)	–11.4 (7.0)

Source Jakubowski et al. (2016). Standard errors in parentheses

We now turn to additional results from the national study that used PISA instruments to test 16- and 17-year-old students in upper secondary schools. These results show that students in vocational schools perform better after the reform when compared to students from vocational schools before the reform. This suggests that the reform has a lasting influence on their achievement. On the other hand, these additional data show that after finishing their general education, these students do not gain much more in vocational education in terms of general skills like reading or mathematics.

These findings were confirmed by studies that followed the cohorts of students on the labour market. These studies use a research design that allows for the estimation of the causal impact of the reform, mainly the extension of general education and the labour market outcomes. In general, both studies compare earnings or unemployment among former students who were affected and unaffected by the reform due to small differences in birth dates. Drucker and Horn (2016) compared income and employment probability among adults of nearly the same age but who were born in different months and thus benefited from 8 or 9 years of compulsory general education. According to their quasi-experimental analysis, the reform improved income by around 3–4% and employment probability by 2–3%. The positive outcomes were larger for the lowest-educated workers, which is in line with the findings from PISA. Using different datasets and focusing on people who finished basic vocational education, Liwiński (2018) found positive labour market effects of the 1999 reform for former male students.

3.4 The Second Wave of Reforms—Curriculum and Evaluation Reforms from (2008 and 2009)

The 1999 reform established, as we have seen, the foundations for a modern education system; however, the curriculum was still too prescriptive, focusing on what teachers should teach rather than what knowledge and skills students should acquire. Rapid changes in Polish education, economy, and society called for its modernization. The results of the OECD PISA study were also used to motivate the reform, as until 2012, they showed that Polish students performed relatively worse in analytical or reasoning tasks. The reform started in 2007 with a consultation process, including all major stakeholders in the system. After one-year-long discussions, a new curriculum was passed in 2008. It was implemented gradually, with the last changes affecting upper secondary schools several years later.

The new curriculum was developed according to these principles:

- Describe the expected learning outcomes for each stage of education,
- Indicate the main objectives of teaching for each school subject,
- Define the requirements of central assessments,
- Constitute a coherent part of the Polish Qualifications Framework.

As one of the authors of the reform writes (Marciniak 2015): “*The curriculum has two layers. The basic layer comprises 3–5 general requirements for each subject, which defines the main objective for teaching a given subject at a given education level. For example, for mathematics at lower secondary school the general requirements include mathematical modelling, strategic thinking, and mathematical reasoning and argumentation. This implies that the primary goal of the teaching process as a whole should be oriented towards developing these skills. The second layer consists of detailed requirements, describing the specific knowledge and skills to be mastered by students, e.g., “a student can solve a system of two linear equations.” However, these particular requirements serve only as a tool in achieving more general aims, as defined by the general requirements.*”

The new curriculum has strengthened teacher autonomy but also the responsibility for the learning process, as it defines learning outcomes only and in a relatively broad manner. It emphasized the general goals of teaching in each subject, leaving a lot of space for teachers in terms of developing their programs and the choice of resources to use. The national exams were aligned with the new curriculum and attempt to focus more on reasoning than on fact-checking.

In fact, the improvement of Polish students between the PISA 2009 and PISA 2012 studies was mostly driven by better responses to items measuring more complex, analytical thinking, at least in mathematics. The reform also emphasized cross-curricular skills and teamwork, while those elements are still only partly implemented in classroom practice.

The new curriculum introduced two substantial changes in upper secondary schools. Firstly, some subjects were combined into interdisciplinary blocks for students who specialize in other subjects (e.g., students focusing on physics would have history classes combined into social science blocks). This change was misinterpreted and criticized as limiting the teaching of history, while, in fact, the number of hours devoted to history teaching was the same but organized differently. The second was the obligation to cover at least an equivalent of one year of the general curriculum in basic vocational schools. While this change was in line with the idea of extending the coverage of general education to all students, in practice, it did not bring the expected results due to the limited teaching capacity in basic vocational schools, low motivation of students, and probably also a too-short period of implementation. In general, these two changes were the most criticized and problematic and were reversed with the new curriculum introduced in 2016.

Another large change was the introduction of school evaluations that replaced the old inspection-type system (see Mazurkiewicz et al. 2014). When developing this system, researchers and decision-makers reviewed several school evaluation systems in other countries, mostly looking at those that performed well in international assessments, e.g., Finland, Scotland, and the Netherlands. Thus, this development was motivated by international comparisons, while the final system decided upon was unique, incorporating only some ideas from other countries.

Prior to the reform, inspections focused on checking administrative and legal issues, sometimes pretending to assess teaching quality but without any substantial investigations involved. The new system was evidence-driven, with the aim of

providing feedback that could help to improve teaching quality. From 2009 the school evaluation system was incorporated into the supervisory structures as a basic method of pedagogical supervision and separated from administrative or legal control. The evaluation was based on a set of requirements that should be addressed by each educational facility. The requirements covered a broad set of activities like delivery of the core curriculum, the development of selected student attitudes and social skills, cooperation with parents, organization of work, and analysis of examination results.

Importantly, the whole process is transparent and inclusive. The evaluation reports are published and discussed with school staff and stakeholders. The goal is to encourage open evidence-based reflection about teaching and school organization. While the system has been modified several times already, and the current administration is not highly supportive of it, it has become an important part of the education system. Currently, both school evaluations and examination results are publicly available at the school level.

3.5 Early Education Reform

International comparisons clearly pointed to one key deficiency of the Polish education system—the relatively low participation rate in preschool education and the later starting of school education (and the later finishing of tertiary education as a result). Changes began in 2007, with support for preschool education from the European Structural Funds, which was followed in 2009 with a government guarantee for places for 5-year-olds in preschool education. This was then replaced in 2011 with compulsory preschool education for children at this age. The plan also assumed a shift in compulsory primary school arrangements, with the starting age changing from 7 to 6 in 2012. However, the latter change was postponed and never fully implemented due to protests and the change of government.

Before the reform, Poland had one of the lowest preschool participation rates in Europe. Figure 6 shows that in 2000 around 58.3% of Polish children participated in preschool education before starting primary school, which can be compared to 85.5% across the 28 EU countries. This participation rate started to increase slowly until 2007 and then more rapidly, mostly thanks to large developments in rural areas. In fact, the preschool participation rate of 3- to 5-year-olds in rural areas was below 25% before 2007. With the help of European Structural Funds and later the support of the central government, the rural local governments could fund new places for preschool education, and in 2014 the participation of 3- to 5-year-olds in preschool education increased to 70% in rural areas.

Overall, Poland almost closed the gap in preschool participation in 2016, mainly due to regulations making preschool compulsory and thanks to additional support for local governments. In 2013, a law was passed that lowered the costs of preschool education for families thanks to grants from the central budget. The government also introduced a guarantee of preschool places for 3- and 4-year-olds that was financed by the central budget and introduced gradually. A recent small decline in preschool

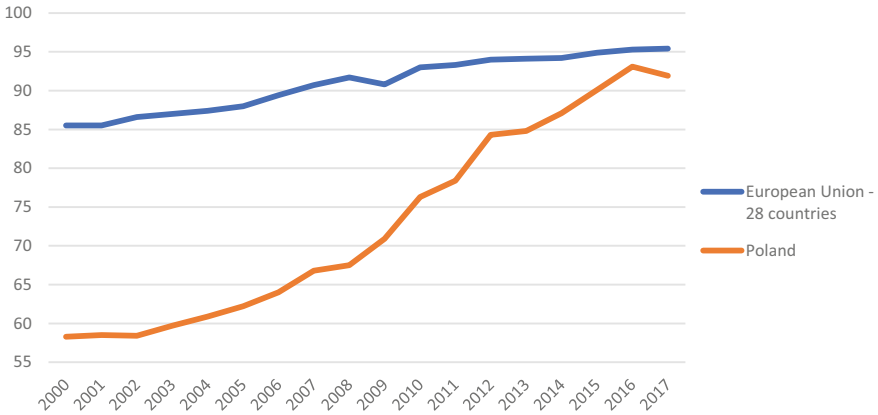


Fig. 6 Participation in preschool education–Poland and EU average. *Source* Eurostat, % of the age group between 4-years-old and the starting age of compulsory education. Indicator SDG_04_30 retrieved at 8th of January 2020

enrolment is the effect of the current government’s abolition of compulsory preschool for 5-year-olds, but still, the participation rates are now close to the EU average and much larger when compared to 2007 when the reform started.

One can see this reform of early education as the final step in building a comprehensive education system in Poland. Thanks to the reforms, all students now have access to comprehensive education from the age of 3 until the age of 15, with additional, obligatory one-year comprehensive material that needs to be covered in all types of upper secondary schools. Thus, for the Polish students who want to fully benefit from the system, free comprehensive education can now last for 13 years.

4 Evidence, Public Sentiment, Politics, and Top-Down Reforms

The reforms implemented between 1999 and 2013 extended compulsory general education, increased preschool participation and modernized the Polish school system. International assessments like PISA, but also PIAAC, TIMSS, and PIRLS, demonstrate how successful these changes were. The latest PISA results show the cumulative effect of improved quality at the primary and lower secondary levels, with Polish students ranking among the top performers in Europe in all subjects. Mathematics was typically the weakest domain for Polish students, but it has been heavily reformed in recent years. Now, in Europe Polish 15-year-olds are outperformed only by Estonians and scored better than young Finns, whose school system is often celebrated in Poland as of higher quality. The evidence shows large gains for low-achieving students and a large decline in between-schools differences after the 1999 reform. These positive outcomes are also confirmed by labour market studies

showing that the extension of compulsory general education can be directly linked to earnings and a lower probability of unemployment, especially for people with vocational degrees. The improvements after 2009 were larger for higher-achieving students. We still need to wait for more in-depth research analysing the impact of the reforms in this period.

The overwhelming evidence in support of the 1999 reform was not widely discussed until the current government decided to change the system. Large protests against the reversal of this reform, which were driven by teacher trade unions, but also supported by groups of parents, researchers, and education experts, did change the views of some people. Surveys of public opinion showed that positive or negative views were closely related to political preferences (for example, see CBOS 2018). The reversal of the reforms was positively viewed by voters of the current ruling party and negatively by those who support opposition parties. Also, the older generations viewed more positively the system inherited from the communist times, while younger people, including those who were actually educated in the lower secondary schools, would as majority support keeping the system as introduced in 1999.

Clearly, the politics of this reform are not related to evidence but have more to do with sentiment and political views. The issue of education reforms was heavily politicized. At the same time, it did not matter that the current party, which reversed these reforms, was among the first to support them some time ago as the opposition party. What matters for the opinion of most people who are not directly involved in education is the current political battle and not the school reform itself.

At the same time, while politics and sentiment seem to be key factors in understanding the support for the reversal of the reforms, they cannot explain everything. Among people who were against the reversal of the 1999 reform, many share negative opinions about its implementation in 1999, have negative views of the current education system, and did not support the early education reforms, especially the lowering of the starting school age.

First, the perception of the 1999 reform is generally negative, and even for people who can recognize successful outcomes, the popular notion is that this reform was chaotic and implemented by force. The implementation of the reform was partly motivated by a wish to dismantle the remaining elements of the education system inherited from the communist times. The reform was introduced in the package of four large reforms (pension reform, health reform, administration reform), which added up to an enormous burden during the implementation. Overall, the reform was difficult for three reasons: it was extremely ambitious and had a list of components that would be sufficient for several reforms (change of school structure, decentralization, change of financing, change of curriculum and textbooks); political opposition; and additional issues caused by the almost-parallel implementation of the other three reforms. Teachers and parents were overwhelmed by the number of changes introduced. In effect, although in the beginning, public opinion expressed more positive views on the reform, the negative opinions increased soon after the implementation and are still present nowadays. While many teachers saw the reform as a chance to improve their situation, they were also afraid of the changes and expressed negative views on the implementation. Negative views on the reform were also amplified

by negative receptions of the other three reforms, which, except for administrative reform, were highly unpopular and have also been reversed recently to a large degree.

It is doubtful, however, that these needed reforms could have been implemented in a less controversial way or with a longer period of consultation and consensus-forming. The government collapsed shortly after the implementation of the education reform, and some changes were quickly reversed by the succeeding governments (for example, an obligatory standardized mathematics exam at the end of upper secondary school was postponed for nearly ten years). Also, important changes have remained even after the recent reform reversal, including not only system-level solutions but also the attitudes of students and teachers. Attempts to re-centralize education are difficult to imagine now, even if the current government tries to limit the autonomy of local governments and schools. Despite difficulties in the beginning, local governments managed educational facilities more efficiently, while teachers used the freedom given by increased autonomy to improve learning outcomes. Fewer students are in basic vocational education now, and enrolment in tertiary education institutions is still high.

Second, the negative view of the reforms and the current education system might be related to the limited role played by parents. Reformers tried to implement regulations giving parents and students more saying in the management of schools, but attempts to formalize the role or set up parents' councils in schools did not succeed, and they still play a mainly advisory role. Similarly, students and other stakeholders, e.g., local NGOs or employers, play limited roles in shaping local schools. This lack of representation of key stakeholders at the local and national levels is often criticized and results in harmful tensions.

Finally, there is criticism of the Polish education system coming from education experts and opinion leaders, which claims it is an old-fashioned system that also does not deal properly with inequality. Many claims about curricula and innovative teaching methods are disputable, as the proposed approaches are rarely new and are often unsupported by research evidence (see Christodoulou 2013, for a discussion of education myths popular in the UK, but also repeated in Poland). The so-called teaching of 21st-century skills is disputable, and one can easily say that Polish schools are able to find a good balance between innovations and traditional teaching, securing very good outcomes for most students. The other criticism is related to inequality and is often linked to the discussion of how socio-economic background is related to student performance in PISA. It is true that the overall effect of socio-economic background on achievement has not changed much, but that is mainly due to the fact that large improvements among low-achieving students were accompanied by similarly large improvements among top-achievers. The differences between students of different backgrounds are average or below average in Poland. The expansion of general education and increased support for preschool education are probably the two most effective policies aimed at limiting the impact of socio-economic background on students. The belief that schools can become the big equalizer in a society as divided as Poland can use international assessments to show that the improvements made in Poland are substantial from a relative and not from the absolute perspective.

5 Conclusions

Since 2000 the Polish education system has been reformed several times, with the most recent wave of reforms reversing key changes implemented earlier. The 1999/2000 reform restructured, decentralized, and introduced standardized national examinations. The reforms around 2007/2008 introduced a new core curriculum and school evaluation system. Later preschool and early education were reformed. The most recent changes partly reversed earlier reforms, but it seems that the system follows its dynamic based on substantial teacher and school autonomy.

Student outcomes, as documented by PISA, but also other international assessments, largely improved over the last 20 years. Thanks to the expansion of compulsory general education, Poland managed to narrow differences between schools and to improve the performance of its low-achievers significantly. Second waves of reforms improved the overall quality of the system, with a larger degree of improvement obtaining among top-achieving students and in mathematics.

There are three lessons from the Polish experience with the implementation of the reforms and the recent reversal of most of them. First, the evidence is not sufficient to support and sustain reforms. The Polish success in PISA was not widely known or sufficiently promoted by education experts and leaders. Even the improving results from the other international assessments did not convince the public that the school system performs very well and that Poland is among the top-ranking countries in Europe. This evidence should be discussed more widely, common misconceptions or invalid criticism should be addressed, and a feeling of being proud of the achievements of the school system should be promoted and celebrated. But nothing like this has happened in Poland. Thus, it is not surprising that many people still do not believe the results from PISA are true or that the results of international comparisons are sufficient for them to change their negative views.

Second, it is not sufficient to rely on international assessments and international reports, while it is necessary to develop a reflective culture of policy and practice, which would support continuous research efforts. International assessments do provide reliable benchmark comparisons allowing for the evaluation of the overall outcomes of the education system in each country. However, the most important questions related to policy outcomes need to be addressed using the secondary analysis of international data and national research. Moreover, the results of these analyses need to be more widely disseminated and popularized among education experts, teachers, parents, and policymakers. Too often, the research community and ministries invest a lot into data collection, only to rely later solely on international reports. This is a serious under-utilization of international assessments and limits their impact on education policy and practice.

Third, not all of the changes stemming from 1999 and the later reforms have been reversed, and some of them became a solid foundation for the Polish education system. Maybe the most important changes are in the attitudes of students and teachers. Most families and students believe nowadays that it is possible to achieve a

tertiary or another diploma that provides better labour market opportunities. Teachers have a feeling of substantial pedagogical autonomy and focus on student outcomes. These and other changes have had a large influence on system outcomes.

References

- Bialecki, I., Jakubowski, M., & Wisniewski, J. (2017). Education policy in Poland: The impact of PISA (and other international studies). *European Journal of Education*, 52, 167–174.
- CBOS. (2018). Ocena systemu edukacji po roku od wprowadzenia reformy. Komunikat z badań 122/2018, Centrum Badań Opinii Społecznej.
- Christodoulou, D., (2013). *Seven myths about education*. Routledge.
- Drucker, L. F., & Horn, D. (2016). *Decreased tracking, increased earning: Evidence from the comprehensive Polish educational reform of 1999*. Budapest Working Papers on the Labour Market 1602, Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences.
- Gajderowicz, T., Grotkowska, G., & Wincenciak, L. (2012). Premia płacowa z wykształcenia wyższego według grup zawodów. *Ekonomista* (5).
- Jakubowski, M., & Topinska, I. (2009). The impact of decentralization on education in Poland. In E. Ahmad & G. Brosio (Eds.), *Does decentralization enhance service delivery and poverty reduction?* Chapter 9, Edward Elgar Publishing.
- Jakubowski, M., Patrinos, H., Porta, E., & Wiśniewski, J., (2016). The effects of delaying tracking in secondary school: Evidence from the 1999 education reform in Poland. *Education Economics*, 24(6), 557–572.
- Liwiński, J. (2018). *The impact of compulsory schooling on earnings*. Evidence from the 1999 Education Reform in Poland, No 253, GLO Discussion Paper Series, Global Labor Organization (GLO)
- Marciniak, Z. (2015). Reviewing polish education reform in the late 1990s—possible lessons to be learned, The World Bank.
- Mazurkiewicz, G., Walczak, B., & Jewdokimow, M. (2014). Implementation of a New School Supervision System in Poland, OECD Education Working Papers No. 111, OECD Publishing, Paris.
- OECD. (2019). *PISA 2018 assessment and analytical framework*. PISA: OECD Publishing, Paris.
- Rynko, M. (Ed.). (2013). *Umiejętności Polaków – wyniki Międzynarodowego Badania Kompetencji Osób Dorosłych (PIAAC)*. Warszawa: Instytut Badań Edukacyjnych.

Maciej Jakubowski is a policymaker, advisor, and researcher in education and labour market policy. He served as an under-secretary of state (deputy minister) at the Polish Ministry of National Education between 2012 and 2014. In 2014, he established the Evidence Institute in Poland to promote evidence-based policymaking and teaching practice. Before joining the government, he worked in the OECD PISA team in Paris (between 2008 and 2012). He worked as a consultant for the World Bank, UNDP, UNESCO, OECD, and several other governments and international institutions. Mr. Jakubowski holds a Ph.D. degree in economics and MA in sociology from the University of Warsaw where he still works at the Faculty of Economic Sciences. His research focuses on statistical methods for policy evaluation and the development and analysis of student assessments.

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.



Portugal: The PISA Effects on Education



João Marôco

Abstract From the bottom of the league table for PISA 2000 Portugal has raised to the OCDE average being the only OECD member that showed, up to PISA 2018, consistent growth in reading, mathematics, and science. This chapter gives a brief description of the Portuguese Education system and how PISA outcomes have shaped Portuguese education policies. It identifies the policies that probably explain the improvement in PISA and pinpoints weakness of the Portuguese education system through the lenses of PISA.

1 Introduction

Accountability is a key feature of modern education systems since countries invest a significant portion of their resources educating their young. In the early 1960s, the International Association for the Evaluation of Educational Achievement (IEA) demonstrated the feasibility of large-scale studies and cross-country comparisons of student achievement in key school subjects (mathematics, reading, and science). During the last decades of the twentieth century the Organization for Economic Cooperation and Development (OECD), following in IEA's footsteps, identified the need to regularly collect reliable and valid educational indicators that could be used to compare its member countries' educational systems and inform policymakers on the outcomes of education policies. The Programme for International Student Assessment (PISA) was the OECD's response to address some limitations of the IEA's studies, namely insufficiencies in education quality measurement and limited international coverage (Breakspear 2014). PISA would fill this perceived gap by providing a measure of comparative educational systems outcomes not focused on curricular knowledge, like the IEA studies, but in terms of what students can do with the knowledge, skills, and competencies learned in school to solve everyday problems and be active citizens. On OECD's own words "PISA represents a new commitment by the government of OECD countries to monitor the outcomes of education systems

J. Marôco (✉)
ISPA—Instituto Universitário, Lisboa, Portugal
e-mail: jpmaroco@ispa.pt

© The Author(s) 2021
N. Crato (ed.), *Improving a Country's Education*,
https://doi.org/10.1007/978-3-030-59031-4_8

in terms of student achievement regularly and within a common framework that is internationally agreed upon. PISA aims at providing a new basis for policy dialogue and for collaboration in defining and operationalizing educational goals. (...) PISA can assist countries in seeking to bring about improvements in schooling and better preparation for young people as they enter an adult life of rapid change and deepening global interdependence” (OECD 2001, p. 3).

The first relevant international assessment of the Portuguese education system took place in 1995 with the IEA’s Trends in International Mathematics and Science Study (TIMSS). The Portuguese results were so poor that policymakers at the time argued that the study failed to assess the country’s students’ knowledge and skills. As a result, Portugal withdrew from TIMSS until 2011 (Barroso 2010; Marôco 2020). However, TIMSS 1995 revealed that the poor performance of Portuguese students went much deeper into the Portuguese social structure, with the education system echoing structural problems that went far beyond schools, teachers, or policymakers (Justino 2010). The aftermath of the TIMSS 1995 shock set the seed for the external evaluation of the Portuguese education system and evidence-based policy reforms.

Portugal, as a founding member of the OECD, participated in the first edition of the OECD’s PISA study in 2000. Like in TIMSS 1995, the Portuguese students were ranked at the bottom of the table of the ordered OECD participants in mathematics, science, and reading. These PISA results finally set the stage for the much-needed education reforms that took place in the following years. In 2018 Portuguese students stabilized their position on the PISA OECD average, being the only OECD country with a positive trend in all three domains during the 2000 to 2018 PISA life frame (OECD 2019c). According to Andreas Schleicher, the OECD Director of Education and skills, “Portugal is Europe’s biggest success story at PISA” (Tavares 2017). This chapter briefly reviews both the Portuguese education system, the Portuguese educational reforms driven by PISA, and its effects on the PISA national results.

2 The Portuguese Education System

The Education pace in Portugal, from the beginning, was set by the Catholic Church with the predominant action of the Jesuits Order. The 1826 Constitutional Bill of Rights was the first official document setting the trend for the gratuity of a primary education focusing on reading, writing, and mathematics (MEC-OEI 2003; Ramos 2004). The implementation of a republican regime in 1910 led to the expulsion of religious orders from Portugal and brought the first republican education reform. This reform was especially aimed at improving the very low literacy rates of the population, namely that from the rural areas, with emphasis on the importance of reading at an early age (Candeias et al. 2007; MEC-OEI 2003).

Post-WWII reforms in Europe found the Portuguese education systems still lagging behind its European neighbors (Justino 2010). Salazar’s fascist regime promoted an inward-looking education system that focused on the Portuguese mainland and its African colonies and very little exposure to external influences. It was

not until 1960 that the 3 or 4-year schooling mandatory for all children (girls and boys, respectively) was imposed. By that time, the OECD's sponsored Mediterranean Regional Project set the first attempt at aligning the Portuguese education with the international education frameworks as part of an effort to meet economic dynamic growth needs (Alves 2012; Mendonça 2011). The Mediterranean project found a country largely deprived of education. In 1970, 18% of the Portuguese population was illiterate, 66% of 15-year olds had not completed any level of formal education, and only 0.9% of the total population had a higher education degree (Crato 2020). A large set of reforms were therefore proposed, in the early 1970s, to all levels of basic, secondary and university studies. Amongst the introduced measures was the mandatory attendance of school for at least eight years. However, the 1974 military coup that ended the almost 50 years of fascist regime aborted these reforms and set the country into revolutionary fervor. A major reorganization of schooling cycles and curricula reforms occurred at a furious pace according to social constructivist views of education. In 1986, a new "Basic Law of the Educational System" ensured the right of education and culture to all children promoting the training required for active citizenship, equality of opportunities, and freedom of learning and teaching. This reform also expanded mandatory schooling to nine years. Vocational and professional tracks aimed at accessing a profession or higher education were introduced in parallel with the regular sciences and humanities tracks during the late 1990s, early 2000s. The mandatory exams that were abolished during the years following the 1974 Portuguese revolution were slowly reintroduced after 1996 for grade 12 certifying the terminus of secondary education and, from 1998 on, ranking the students' access to higher education. National high-stake Portuguese language and Mathematics exams were introduced for grade 9 in 2005, and in 2009 the Parliament decided to extend mandatory schooling to 12 years of basic (grades 1–9) and secondary (grades 10–12) education. This was put in place gradually in the 2012–2015 period. Universal access to preschool for 5-year old children, but not mandatory enrollment, was also introduced in 2009. In 2015, the preschool age group was extended to the 4-year olds.

Despite the extraordinary evolution observed since the 1970s, and the 12-year mandatory schooling, the completion of secondary and higher education remains a challenge. As of 2018, the actual schooling rate was 90% for pre-school education, 95% for elementary (grades 1–4), 89% for primary (grades 5–6), 88% for lower-secondary (grades 7–9) and 79% for upper-secondary (grades 10–12) (PORDATA 2019b) (see Fig. 1). Around 41% of 19–20 year-old young adults were enrolled in higher education and only 25% of adults (25–64 years old) held a university degree (the OECD average is 40%) (OECD 2019b). The Portuguese Basic and Secondary Education system is also characterized by being the one with the oldest teaching professionals in the OECD. More than 40% of teachers are 50 years old or older and only 1% are below the age of 30 (OECD 2019b). The average number of students per class is 21/22 for primary and lower-secondary and the expenditure is 11 k USD per student (the total expenditure on education is 3.6% of the GDP) (OECD 2019b; PORDATA 2019a). About three-quarters (72.9%) of schools are public, and most are

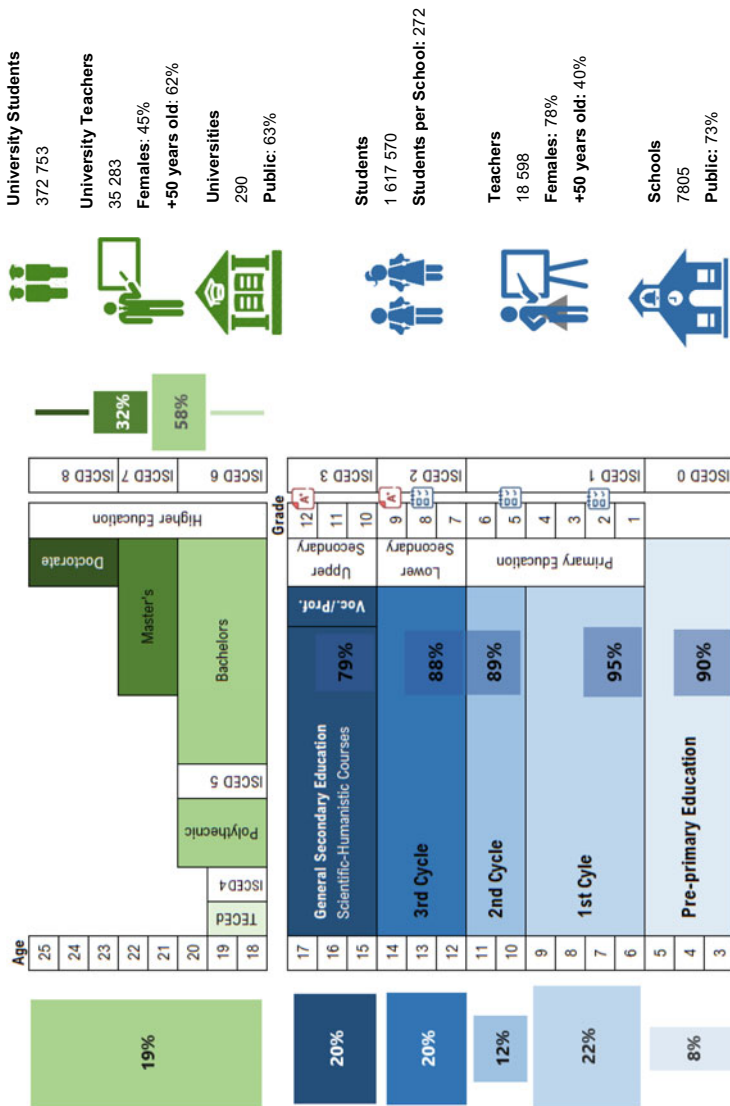


Fig. 1 Structure of the Portuguese education system as of 2018. The percentages on the left-hand side are the distribution of students in the three cycles ($n = 1\ 617\ 570$). Percentages within the cycles are the actual schooling rate. Percentages on the right-hand side are the students' distribution through the different higher education degrees ($n = 372\ 753$). represent high-stake census assessment; represent low-stake census assessment. Adapted from PORDATA (2019c), sources in the text, and Marôco (2020)

grouped in school clusters offering grades 1–12. A full description of the Portuguese education system can be found in Eurydice (2019). Figure 1 summarizes the actual structure of the Portuguese Education system.

3 The Portuguese PISA Trends

Portugal, as a member of the OECD, has taken part in all the editions of PISA. In the first edition, back in 2000, the Portuguese students ranked at around the antepenultimate positions for the three domains out of the 28 OECD countries who had their results reported the following year (OECD 2001). Approximately one out of four (26.3%) Portuguese students who took the PISA test in 2000 did not reach the minimum acceptable proficiency level in reading (level 2). That was more than twice the OECD average. Only 4% were able to reach the advanced level in reading literacy (level 5), about half of the OECD average (OECD 2001). It took 15 years to see the Portuguese students raise to the OECD average in mathematics literacy and significantly above the OECD average in reading and science (OECD 2016) (Fig. 2). In 2018, the Portuguese position at the OECD average was again confirmed for reading, science, and mathematics (OECD 2019c). Even so, 20.2% of the Portuguese students did not reach the minimum acceptable proficiency level (level 2) in reading, a value that nevertheless is slightly lower than the OECD average (22.6%). Portugal’s trends in PISA contrasts with the overall OECD trends. In the OECD Secretary-General Angel Gurría’s own words “only seven of the 79 education systems analyzed [in PISA 2018] saw significant improvements in the reading, mathematics and science

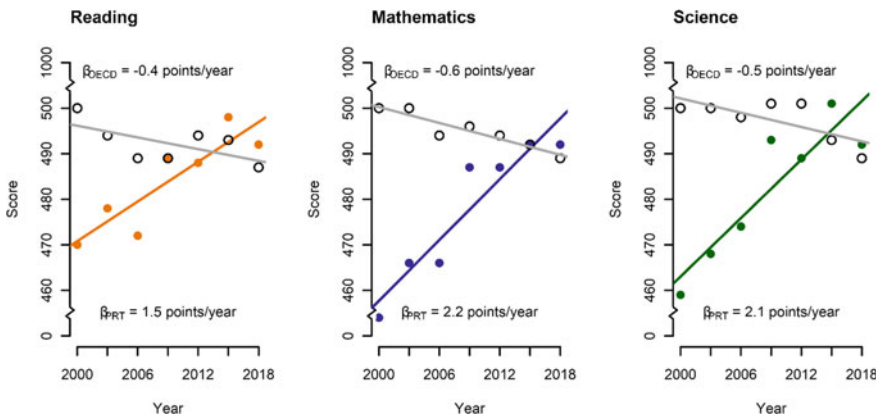


Fig. 2 Trends in reading, mathematics, and science literacy for Portugal (closed symbols) and the OECD average (open symbols). The β coefficients are the slopes of the linear OLS regression lines displayed. Data were retrieved from the OECD PISA reports (2001–2019) (The PISA scale has an average of 500 points and a standard deviation of 100 points for reading, science and mathematics established on the cycle were the domain was for the first time the major domain.)

performance of their students throughout their participation in PISA, and only one of these, Portugal, is a member of the OECD” (OECD 2019c, p. 3). Despite some ups and downs, Portuguese students’ performance in PISA has continuously improved. The average growth rate was 1.5 points per year for reading, 2.2 points per year for mathematics, and 2.1 points per year for science. The corresponding trends for the OECD were -0.4 points per year for reading, -0.6 points per year for mathematics, and -0.5 for science (see Fig. 2).

As compared to students who took the PISA test in 2000, the 2015 and 2018 cohorts have improved, on average, and for the three PISA domains, by 31 points (0.3 standard-deviations on the PISA scale). This improvement corresponds to about one school year (OECD 2009, p. 23). As the OECD report on PISA 2015 points out: “Macao (China) and Portugal were able to ‘move everyone up’ in science, mathematics and reading performance over the past decade by increasing the number of top performers while simultaneously reducing the number of students who do not achieve the baseline level of skills. Their experiences demonstrate that education systems can nurture top performers and assist struggling students simultaneously.” (OECD 2016, p. 266). However, and despite the positive overall evolution, the Portuguese PISA data reveals strong regional asymmetries. An analysis of PISA 2015’s major domain, science literacy, revealed that the difference between the highest and lowest achieving Portuguese NUTS III regions was equivalent to almost two and a half school years (about three-quarters of a PISA standard deviation) (Marôco et al. 2016, Marôco 2017, 2020). Analyses for reading, mathematics, and science literacies in 2018 revealed similar asymmetries (Fig. 3). The differences between the region with the highest statistically significant difference above the national mean and the lowest ranking region ranged between 59 points (for reading) to 72 points (for mathematics). These differences correspond to about two to two and a half PISA school years. The years of schooling gap was reduced from PISA 2015 to PISA 2018. However, this reduction was due to a significant drop in the science results, and non-statistically significant drops in reading and mathematics (see Fig. 2) rather than an improvement of low-achieving regions.

Although concerns were raised regarding the validity of PISA to assess the literacy of Portuguese students (see, e.g., Cristo 2017), subsequent research work has shown the concurrent and content validity of PISA, and also TIMSS, with the Portuguese national high-stake exams for mathematics (Marôco and Lourenço 2017). Figure 4 summarizes the overlap between major curricular domains of the 9th grade high-stakes mathematics exam (1st call of 2015) and the mathematics content domains in PISA 2015, and also the correlation observed between mathematics content literacy in PISA and the national exam score of the students who participated in both tests (Spearman’s $r = 0.64 \pm 0.01$, $p < 0.001$). It is worthwhile to note that, despite the different objectives of the PISA test and the national exams, the correlation between PISA and the national exam is almost equal to the correlation between the students’ teachers final assigned grade and the national exam grade (Spearman’s $r = 0.62 \pm 0.01$, $p < 0.001$) (see also Marôco 2020).

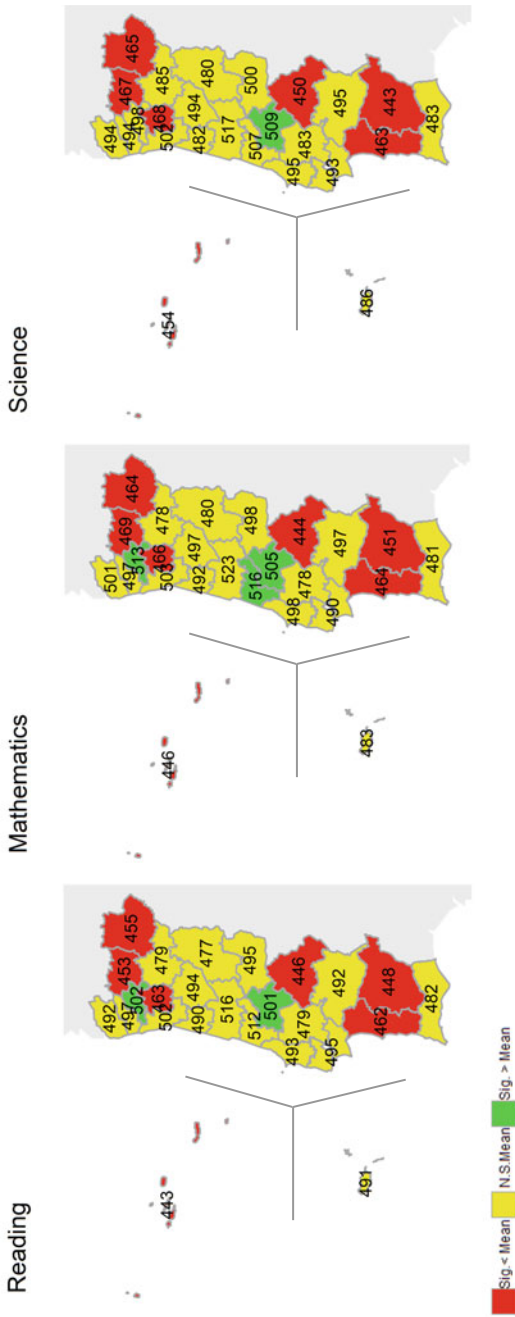


Fig. 3 Mean scores per NUTS III regions in Portugal for reading, mathematics, and science on PISA 2018. The yellow color identifies regions not statistically significantly different from the national average (492 points for reading, mathematics, and science), the red color regions where the means are significantly lower than the national average, and the green color regions with means scores that are significantly higher than the national average. Data retrieved from the PISA PRT database (OECD 2019a). Statistical comparisons of the mean for regions vs national mean were weighted by student weights, and $p \leq 0.05$ for large sample unilateral t-tests was assumed for statistically significant differences

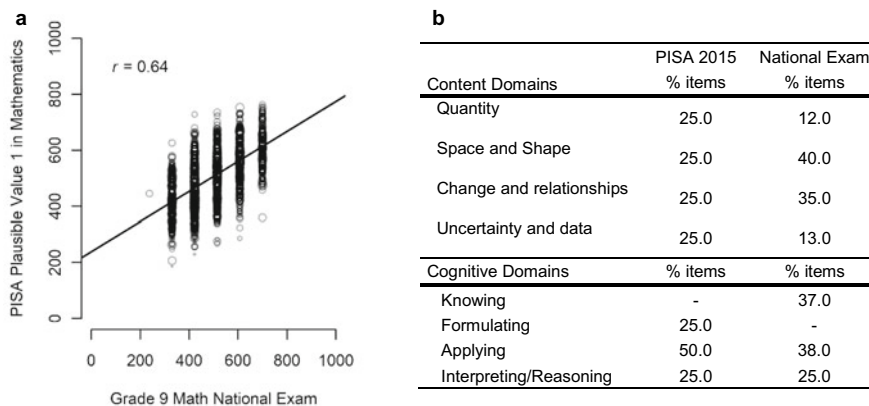


Fig. 4 Concurrent (a) and content (b) validity of the national mathematics exam at grade nine and the PISA 2105 math literacy. National exam scores (ordinal scale ranging from 1 to 5) were converted to the PISA scale for illustration purposes. The correlation coefficient was calculated with the 10 plausible values for math literacy weighted by the final trimmed nonresponse adjusted student weight using an SPSS syntax produced by IEA’s IDB Analyzer corrected to calculate Spearman’s rho Adapted from Marôco (2020)

4 PISA’s Education-Driven Policies

Portugal’s policymakers have looked at international assessments of educational systems since the first OECD Mediterranean Regional Project diagnosis in the late 1960s. Before PISA, Portugal participated in the Second International Assessment of Education Progress (IAEP II 1991) and IEA’s Third International Mathematics and Science Study (TIMSS 1995). While IAEP II went relatively unnoticed, TIMSS 1995 was the first large scale comparative assessment that showed Portuguese students considerably lagging behind similar age peers from the 26 countries who sampled 4th and 8th-grade students. At that time these very poor results were dismissed because policymakers felt that TIMSS was not a valid measure of Portugal’s students’ specific knowledge and skills that were not aligned with the TIMSS curricula framework (Barroso 2010; Carvalho et al. 2017). Despite the TIMSS 1995 insights being unfavorably received, the 1995 large scale assessment set the seed to assessment policy changes and mathematics and science curricular reforms. It was also a turning point in the acknowledgement of the need to not only further assess Portugal’s educational system according to international standards but also to pay more attention to results in basic subjects (Crato 2020; Marôco 2020).

As an OECD member, Portugal participated in the first edition of PISA (2000) all the others that followed. In the words of the minister of Education Nuno Crato, in office from 2012 to 2015, PISA in Portugal is mainly “seen as a mirror” reflecting where the country stands in comparison to other countries in the PISA picture

(Carvalho et al. 2017) but has nevertheless allowed education policymakers to propose evidence-based policies changes as follows.

The PISA 2000 debacle and the publication of its results in 2001 set the stage for the endorsement of a series of ongoing measures aimed at out of class students support (accompanied studies) and the reformulation of upper secondary curricula by the minister of education Júlio Pedrosa (Carvalho 2009; Carvalho et al. 2017). The next minister, in office from 2003 to 2004, David Justino, implicitly recognized the poor PISA and TIMSS results to promote the re-emergence of national assessments in Portugal, first as a low-stakes, in 2003, and, in 2005, as high-stakes exams for mathematics and Portuguese language at the end of grade nine (see Justino and Almeida 2017 for further details). The next explicit mention of PISA was done by Minister Carmo Seabra in an address to the national parliament based on “the last OCDE clear results” (Afonso and Costa 2009). Carmo Seabra brought the PISA 2003 results to the political agenda to promote curricular changes and the need to prioritize the learning of the Portuguese language, mathematics, and science.

The next education minister, Maria de Lurdes Rodrigues, in office from 2005 to 2009, identified PISA as a major source of statistical data on Portuguese students’ literacy and its importance to support evidence-driven policies. The evidence provided by PISA 2000 and 2003 that Portuguese students were performing poorly in terms of reading, mathematics and science literacies drove the introduction of a series of programs and strategic plans consolidating educational policies started in the early 1990s after the publication of the 1986’s Basic Law of the Educational System (Fernandes et al. 2019; Fernandes and Gonçalves 2018). These included the Training Program in Experimental Science Teaching (2006), the National Program for Portuguese Language Teaching (2007), aimed at primary education, the Mathematics Action Plan (2006), and the National Reading Plan (2007) (Afonso and Costa 2009; Carvalho et al. 2017). During her tenure, Lurdes Rodrigues quoted the association between poor results in PISA and families’ cultural and socioeconomic status to enlarge the economic support for students from low-income families; to facilitate the access to internet and computers for primary education (the 2007 Technological Education Plan); and to reorganize the Priority Intervention Educational Territories (TEIP) Program for schools located in economically depressed areas (Afonso and Costa 2009).

The next big impact of PISA results on the Portuguese education policies came with minister Nuno Crato, in office from 2012 to 2015. He took the early reforms and the apparent stagnation of the PISA results from 2009 to 2012 to reinforce curricular “targets” and learning outcomes on the basic and secondary education, curricular structure revisions (more teaching hours) for Portuguese, Mathematics, and Sciences (2013–2014), school autonomy, and to push for better teachers’ initial training. He paid particular attention to vocational high-school tracks and made them part of the compulsory schooling diversifying the schools’ offers to cover different students’ interests (Crato 2020). He also implemented the end of cycle high-stakes exams for mathematics and Portuguese in grade 6 (2012) and grade 4 (2013) as well as the assessment of teachers’ qualifications and certification required for teaching, proposed by Lurdes Rodrigues, in 2014–2015. Both grades 4 and 6

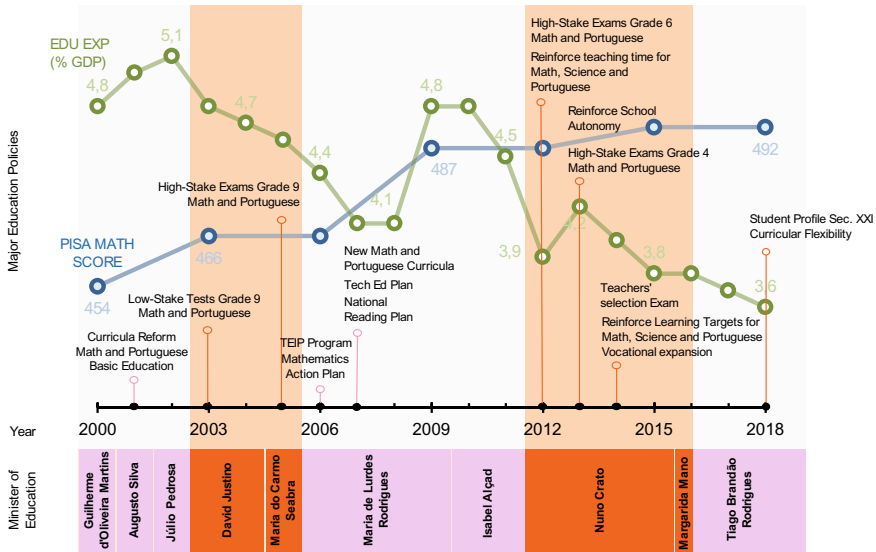


Fig. 5 Major educational policies explicitly justified or supported by PISA outcomes (the pink color indicates socialist governments and the orange color social-democrat governments.) Updated from Marôco (2020) and sources in the text. Educational Expenditure (EDU EXP) as a percentage of GDP was retrieved from PORTDA (2019a)

exams and the teachers' examinations were terminated in 2016 by the newly elected government. The curricular reforms of Crato also pursued the international alignment of the national curricula with the ones inferred by PISA and, especially, TIMSS frameworks (Marôco 2020). Figure 5 summarizes the main educational policies that were justified explicitly with PISA outcomes.

5 What May Explain the Portuguese Evolution in PISA

The concurrent validity of the PISA 2015 mathematics results with the Portuguese national exams and its trends have been demonstrated elsewhere (Marôco 2018; Marôco and Lourenço 2017) and thus it is possible to safely use PISA as a reasonable proxy for the Portuguese education system. However, it must be acknowledged that PISA is a correlational study. PISA uses a complex test design and statistical methods to impute students' missing by design responses (see e.g., OECD 2009). In each PISA cycle, a different cohort of students is sampled, and trends are estimated from items that are common to two or more editions of the test. Henceforth, although causal inferences may be suggested by PISA data, there is no way to ensure that the data support causal effects since correlation does not imply causation. Furthermore, the

PISA test has also drawn criticism from several sources both on the importance of the limited subjects covered by the test for students' and economies' development (see e.g., Schult and Sparfeldt 2016), to the lack of transcultural invariance when comparing countries' results (Rutkowski and Svetina 2014). Despite not being free from criticism, PISA is generally accepted by policymakers and the public in general as a valid and reliable instrument to benchmark the performance of education systems and facilitate education reforms both abroad (Breakspear 2012; Phillips and Jiang 2015) and locally (Carvalho et al. 2017; Fernandes et al. 2019; Justino 2010). A lag between policy changes and results observed in the PISA test as well as cumulative effects must also be considered when linking policies with PISA results. And again, inferring causality from correlation may just be a form of statistical fantasy.

Portuguese students' performance has improved significantly in the PISA test and PISA outcomes have supported Portuguese education policy changes. In every edition and following the release of the PISA results, commentators, from journalists to academics to policymakers, profusely give their accounts of what causes the evolution of Portuguese results. Carvalho et al (2017), regarding PISA 2015, reviewed all the opinion articles published after the public release of the PISA 2015 report. As far as education policies are concerned, a consensus emerged about the causal effects of the extension of pre-school education, differentiation of pedagogical practices, improvement of schools infrastructures, a culture of 'exigency' supported on the reinforcement of curricula aligned with international frameworks, high-stake exams, and increasing offer of vocational/professional courses targeted at students with a lesser interest in the regular track. Although, for some (see e.g., Fernandes et al. 2019, p. 42) these PISA effects on policy were just a reflection and a continuation of the education policies set almost 15 years before PISA by the 1986's Basic Law of the Educational System. Ferreira et al. (2017) looking at PISA results from 2000 to 2015, with a major emphasis on PISA 2012, have identified, as follows, the principal features that explain Portugal's evolution: (1) the overall expenditure in education per capita in line with other OECD member states, in spite of Portugal being a relatively poor country in OECD terms [the 2018 GDP per capita for Portugal was 32.4 k USD versus 43.5 k USD for the OECD average (OECD Stat 2019.)]; (2) Pre-school coverage close to 100%; (3) Teachers appropriate specific and pedagogical training, competence and motivation towards teaching; (4) Students' support by parents and teachers, motivation and persistence; (5) Schools in less favored economic regions performing above the expectation and schools' educational projects aligned with the community; and (6) Improvement in Parents' education.

However, PISA also shows that there is still much need for improvement in the Portuguese education system. At system level, there is an urgent need to promote measures aimed at the reduction of grade retention, increase parental education, renew aging teachers, and improve schools' autonomy, especially as far as teacher recruitment is concerned (Ferreira et al. 2017). It is noticeable that despite the economic crisis of 2008–2013, when national GDP was reduced by 8% (Perez and Matsaganis 2018) and the overall negative evolution of the education expenditure (−0.07% of GDP per year from 2000 to 2018), Portugal was still able to increase its overall PISA scores. Indeed, the correlation between the Portuguese expenditure in

education (as % of GDP, see Fig. 5) and the PISA results from 2000 to 2018 is $r = -0.72$. The same correlation for the OECD is $r = 0.50$. At student level, secondary analysis of data from PISA 2105 (Marôco 2017) as well as PISA 2018 (Gomes et al. 2019; Marôco 2019) reveals that students expectations on their future occupation and the families' socioeconomic and cultural status are still major determinants of the Portuguese students' performance.

6 Concluding Remarks

Like in any other country and economies that take part in PISA, the Portuguese media, the public, educators, and policymakers accept PISA as a robust and legitimate proxy for the Portuguese education system. Although the bottom-of-the-table PISA 2000 results were not received in Portugal with as much "shock" as they were in other poor performing countries (e.g. Germany), PISA has nevertheless produced data and evidence that has been used by the Portuguese education policymakers to justify and promote reforms at different levels of the system. First, and foremost, the poor PISA results were the evidence required to promote the always controversial curricula restructuration in key disciplinary areas, namely Portuguese language, mathematics, and natural sciences, both at the basic and secondary education levels. National programs were aimed at the promotion of reading habits and the increase of time for teaching Portuguese language and mathematics. Teachers' requirements and training, as well as increased teaching times, were also promoted based on comparisons with other PISA participants. Although no causal effects of policies motivated by PISA can be undoubtedly defended due to the correlational nature of the study, Portuguese students' improvements in PISA were aligned with some key education policies changes. The biggest jump in the Portuguese PISA results was observed from 2006 to 2009 and the temporal coincidence with the introduction of the 9th-grade mathematics and Portuguese language exams in 2005 is undeniable. Also, regarding the evolution of PISA from 2015 to 2018, Science was the only subject with a statistically significant drop, coincidentally this is the PISA domain that does not have a high stake national assessment. The effect of high-stake assessments on PISA scores has also been observed in several other countries (Bergbauer et al. 2018). The introduction of high-stake assessments in the Portuguese system is probably the policy with a larger effect on the Portuguese PISA story. Strengthen of curricula, learning targets, and structural teaching and class changes in response to PISA may also play an important role, and those have been consistently pursued by both socialist and social democrat ministers up to 2015 (see Fig. 4). This has, however, changed with the last cycle of governance which brought the extinction of grade four and six national exams and the teacher screening exams. The high-stake exams for grades four and six were exchanged by low-stake diagnostics tests for grades two, six and eight in Portuguese language, mathematics and several other rotating subjects. The new minister also introduced curricular flexibility and curricula trimming to "essential learning targets" at public schools, as well as the continuity of measures aimed at the

reduction of grade retention to the OECD levels. A short-term effect of these policies may well just surface in the PISA 2018 results. The devaluation of external high-stake assessments and the suggestion for trimming of learning targets may reduce the effort and engagement of students with low-stake tests like PISA. Indeed, three out of four Portuguese students reported expending less effort on the PISA test than if the test counted towards their marks, the same figure for the OECD was 68% (OECD 2019c). Also, and for the first time, the participation rate of the Portuguese students (76%) was below the PISA standard of 80% (OECD 2019c).

One recurrent criticism of PISA effects on education is the funneling of school's subjects to the PISA domains—see, e.g. the open letter from academics from all over the world to Dr. Andreas Schleicher published by The Guardian in May 2014 (Various 2014). The new education policies in place since 2016 acknowledge this and other criticisms. According to minister Tiago Brandão Rodrigues, in office since late 2015, “PISA recommendations are embodied in the current Government's program” (Bourbon 2016). Its major effects will, however, only be seen in the next PISA cycles, once the policy-lag effect is overcome.

Despite the praised evolution of the Portuguese students in PISA, a trend with no companion in the OECD, PISA reveals that student performance is strongly asymmetric within the country. An equivalent of two school years separates the highest and the lowest-achieving regions of the country. PISA also shows that schools have failed consistently to serve as social elevators. Students' expectations and families' socio-economic social status are the major determinants of Portuguese students' results. These effects have been present in all PISA editions, including the last.

From an epistemic lag in the last century, the Portuguese education system has raised to the level of its OECD counterparts as measured by PISA. Education requirements are changing at a pace faster than ever before, and education policies are changing in accordance to meet the need for the so-called XXI century skills. In the coming waves, PISA will tell us whether Portugal is still moving in the right direction.

References

- Afonso, N., & Costa, E. (2009). The influence of the Programme for International Student Assessment (PISA) on policy decision in Portugal: The education policies of the 17th Portuguese Constitutional Government. *Sísifo. Educational Sciences Journal*, 10(set/dez), 53–64. <https://sisifo.fpce.ul.pt>.
- Alves, L. A. M. (2012). *História da Educação - Uma Introdução*. <https://ler.letras.up.pt/uploads/ficheiros/10021.pdf>.
- Barroso, C. F. de A. (2010). *Os estudos PISA e o ensino das ciências físico-naturais em Portugal: a comparabilidade dos resultados nacionais e as implicações para a política educacional*. [Universidade Nova de Lisboa]. <https://run.unl.pt/bitstream/10362/5279/1/carlosbarroso.pdf>.
- Bergbauer, A. B., Hanushek, E. A., & Woessmann, L. (2018). Testing. In *NBER working paper* (Vol. 24836). <https://doi.org/10.1021/ja01163a101>.

- Bourbon, M. J. (2016). *Recomendações do PISA estão plasmadas no programa do atual Governo*. Expresso. <https://expresso.pt/sociedade/2016-12-06-Recomendacoes-do-PISA-estao-plasmadas-no-programa-do-atual-Governo>.
- Breakspear, S. (2012). The Policy Impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. *OECD Education Working Papers*, 71, 1–31. <https://doi.org/10.1787/5k9fdqffr28-en>.
- Breakspear, S. (2014). How does PISA shape education policy making? Why how we measure learning determines what counts in education. *Centre for Strategic Education: Seminar Series*, 240, 16. <https://simonbreakspear.com/wp-content/uploads/2015/09/Breakspear-PISA-Paper.pdf>.
- Candeias, A., Paz, A., & Rocha, M. (2007). *Alfabetização e escola em Portugal nos séculos XIX e XX: os censos e as estatísticas*. Fundação Calouste Gulbenkian.
- Carvalho, L. M. (2009). Pisa and educational public policies: studies in six European countries. *Sisifo Educational Sciences Journal*, 10, 3–84. <https://sisifo.fpce.ul.pt/pdfs/Journal10ENG.pdf>.
- Carvalho, L. M., Costa, E., & Gonçalves, C. (2017). Fifteen years looking at the mirror: On the presence of PISA in education policy processes (Portugal, 2000–2016). *European Journal of Education*, 52(2), 154–166. <https://doi.org/10.1111/ejed.12210>.
- 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*. Springer.
- Cristo, A. H. (2017). Os três problemas dos exames nacionais. *Observador*. <https://observador.pt/especiais/os-tres-problemas-dos-exames-nacionais/>.
- Eurydice. (2019). *Portugal*. Retrieved December 18, 2019, from https://eacea.ec.europa.eu/national-policies/eurydice/content/portugal_en.
- Fernandes, D., & Gonçalves, C. (2018). Para Compreender O Desempenho Dos Alunos Portugueses No PISA (2000–2015). In Maria Isabel Ramalho Ortigão (Ed.), *Políticas de Avaliação, Currículo e Qualidade: Diálogos sobre o PISA – Volume 3* (pp. 39–68). EDITORA CRV. <https://doi.org/10.24824/978854442369.1>.
- Fernandes, D., Neves, C., Tinoca, L., Viseu, S., & Henriques, S. (2019). *Políticas educativas e desempenho de Portugal no PISA (2000–2015)*. https://afc.dge.mec.pt/sites/default/files/docs/RelatorioFinal_PolitEdu_PISA2000_2015.pdf.
- Ferreira, A. S., Flores, I., & Casas-Novas, T. (2017). *Porque Melhoraram os Resultados PISA em Portugal. Estudo Longitudinal e comparado (2000–2015)*. Fundação Francisco Manuel dos Santos. <https://www.ffms.pt/FileDownload/9857244f-4dfb-48ad-b196-0448dc444865/porque-melhoraram-os-resultados-pisa-em-portugal>.
- Gomes, J. F., Hormigo, I., & Marôco, J. (2019). *Primeiras lições do PISA: dois países a seguir e Portugal no mundo*. Teresa e Alexandre Soares Dos Santos Iniciativa Educação. <https://www.iniciativaeducacao.org/pt/ed-on/ed-on-estatisticas/primeiras-liceos-do-pisa-dois-paises-a-seguir-e-portugal-no-mundo>.
- Justino, D. (2010). *Difícil é Educá-los*. Fundação Francisco Manuel dos Santos.
- Justino, D., & Almeida, S. (2017). International assessment, curriculum policy induction and instruction time management. *European Journal of Curriculum Studies*, 4(2), 671–691. <https://pages.ie.uminho.pt/ejcs/index.php/ejcs/article/view/162/94>.
- Marôco, J. (2017). Assimetrias Educacionais em Portugal: Através das Lentes do PISA. In Conselho Nacional de Educação (CNE) (Ed.), *Estado da Educação 2016* (pp. 254–274). Conselho Nacional de Educação (CNE). https://www.cnedu.pt/content/noticias/CNE/CNE-EE2016_web.pdf.
- Marôco, J. (2018). *E pur si muove!* – *Observador*. <https://observador.pt/opiniao/e-pur-si-muove/>.
- Marôco, J. (2019). *PISA 2018: Portugal, é ainda a maior história de sucesso na Europa?* Teresa e Alexandre Soares Dos Santos Iniciativa Educação. <https://www.iniciativaeducacao.org/pt/ed-on/ed-on-estatisticas/pisa-2018>.
- Marôco, J. (2020). International large-scale assessments: Trends and effects on the portuguese public education system. In H. Harju-Luukkainen, N. McElvany, & J. Stang (Eds.), *Monitoring student achievement in the 21st century european policy perspectives and assessment strategies*. Springer. <https://www.springer.com/gp/book/9783030389680/aboutBook>.

- Marôco, J., Gonçalves, C., Lourenço, V., & Mendes, R. (2016). *PISA 2015 - Portugal. Vol. I: Literacia Científica, Literacia de Leitura & Literacia Matemática*. Lisboa: IAVE.
- Marôco, J., & Lourenço, V. (2017). *Search of concurrent validity of large-scale international students assessments (LSISA) and high-stakes national exams: the portuguese case study*. Paper presented at the 2017 European Conference on Educational Research: Network 9 Assessment, Evaluation, Testing and Measurement, 126. https://www.iave.pt/images/FicheirosPDF/Estudos_Internacionais/ECER2017_InSearchOfConcurrentValidity.pdf.
- MEC-OEI. (2003). *Breve Evolução Histórica Do Sistema Educativo*. <https://www.oei.es/historico/quipu/portugal/historia.pdf>.
- Mendonça, A. (2011). *Evolução da Política Educativa em Portugal*. Retrieved September 20, 2018, from <https://www3.uma.pt/alicemendonca/conteudo/investigacao/evolucaodapoliticaeducativaemPortugal.pdf>.
- OECD.Stat. (2019). *Level of GDP per capita and productivity*. https://stats.oecd.org/index.aspx?DataSetCode=PDB_LV.
- OECD. (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. OECD Publications. <http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33691596.pdf>.
- OECD. (2009). PISA data analysis manual: SPSS. In *Analysis* (2nd edn). OECD. <https://doi.org/10.1787/9789264056275-en>.
- OECD. (2016). *Pisa 2015 results: Excellence and equity in education: Vol. I*. OECD Publishing. <https://doi.org/10.1787/9789264266490-en>.
- OECD. (2019a). *2018 Database—PISA*. <https://www.oecd.org/pisa/data/2018database/>.
- OECD. (2019b). *Education at a Glance 2019* (OECD Indic). OECD Publishing. <https://doi.org/10.1787/f6dc8198-es>.
- OECD. (2019c). *PISA 2018 results (Volume I): What students know and can do: Vol. I* (O. Publishing (ed.)). <https://doi.org/10.1787/5f07c754-en>.
- Perez, S. A., & Matsaganis, M. (2018). The political economy of austerity in Southern Europe. *New Political Economy*, 23(2), 192–207. <https://doi.org/10.1080/13563467.2017.1370445>.
- Phillips, G. W., & Jiang, T. (2015). Using PISA as an international benchmark in standard setting. *Journal of Applied Measurement*, 16(2), 161–170. <https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2015-26021-004&lang=pt-br&site=ehost-live&scope=site>.
- PORDATA. (2019a). *Government expenditure on education as a share of GDP*. <https://doi.org/10.1787/9789264301900-graph65-en>.
- PORDATA. (2019b). *PORDATA—Actual schooling rate*. <https://www.pordata.pt/en/Portugal/Actual+schooling+rate-987-7864>.
- PORDATA. (2019c). *PORDATA—Education*. PORDATA. Base de Dados de Portugal Contemporâneo. <https://www.pordata.pt/en/Theme/Portugal/Education-17>.
- Ramos, R. (2004). Para uma história política da cidadania em Portugal. *Análise Social*, 39(172), 547–569. <https://doi.org/10.2307/41012031>.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>.
- Schult, J., & Sparfeldt, J. R. (2016). Reliability and Validity of PIRLS and TIMSS. *European Journal of Psychological Assessment*, 1–12. <https://doi.org/10.1027/1015-5759/a000338>.
- Tavares, P. S. (2017). Andreas Schleicher: Portugal é a maior história de sucesso da Europa no PISA. *Diário de Notícias*. <https://www.dn.pt/portugal/interior/andreas-schleicher-portugal-e-a-maior-historia-de-sucesso-da-europa-no-pisa-5659076.html>.
- Various. (2014). OECD and Pisa tests are damaging education worldwide—academics. *The Guardian*, 06.05.14. <https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>.

João Marôco is an associate professor at ISPA-IU and invited professor at the NOVA University of Lisbon where he teaches statistics, data analysis, and research methods. He is also a consultant for Educational Assessment and Statistics at the World Bank. Between 2014 and 2018 he served on the board of directors of IAVE, I. P. (the Portuguese National Assessment Agency) where he coordinated PISA, TIMSS, TIMSS Advanced, PIRLS, and ICILS and e-Assessment pilot projects. He is the author/co-author of more than 350 research papers related to statistics, psychometrics and education and four books in Statistics and Psychometrics. According to Google Scholar, his academic works have been quoted more than 27500 times ($H = 55$, $i10 = 179$).

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.



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



Montse Gomendio

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

1 The Value of International Comparisons: Uses and Misuses

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

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

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

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

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

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

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

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

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

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

2 The Spanish Case: Shedding Light on the Darkness

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

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

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

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

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

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

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

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

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

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

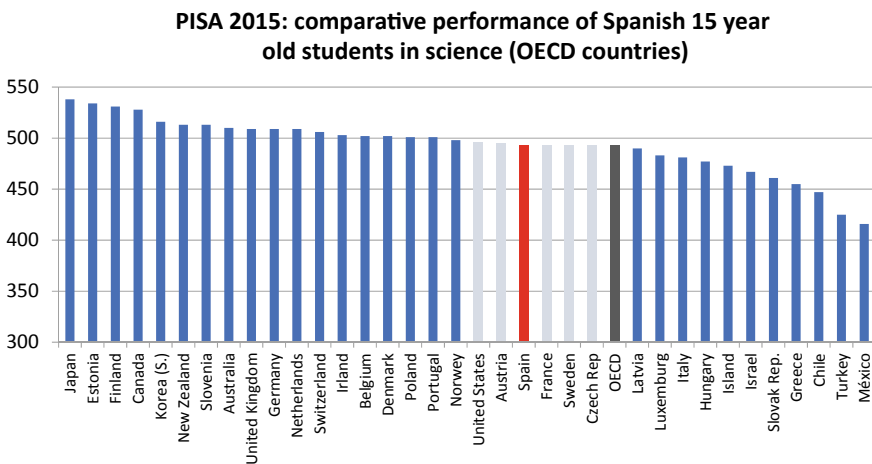


Fig. 1 PISA 2015 scores for the main domain (science)

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

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

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

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

2.2 What ILSAs Tell Us About Trends Over Time

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

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

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

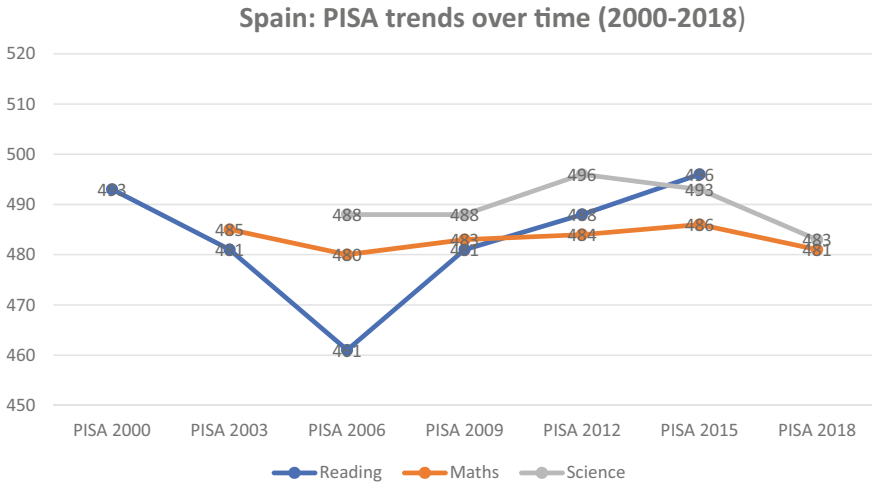


Fig. 2 PISA scores for Spain from 2000 until 2018

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

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

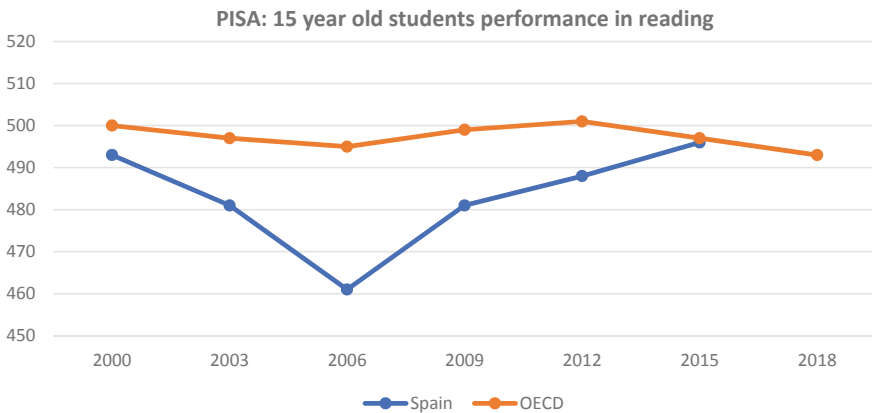


Fig. 3 Spain versus OECD: reading performance over time (PISA)

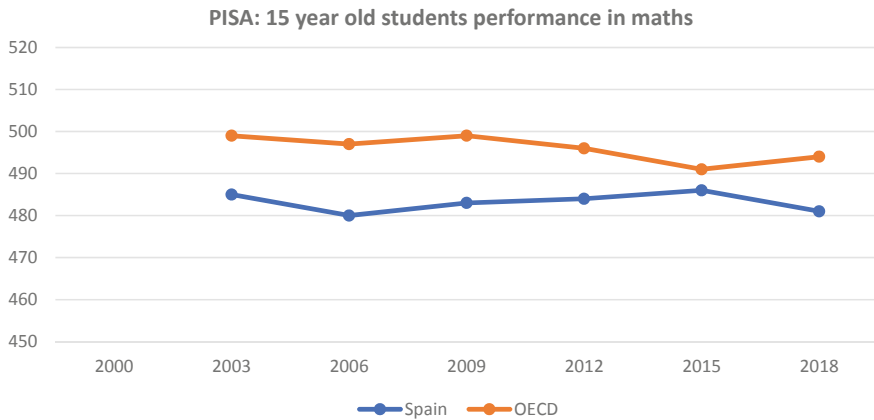


Fig. 4 Spain versus OECD: maths performance over time (PISA)

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

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

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

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

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

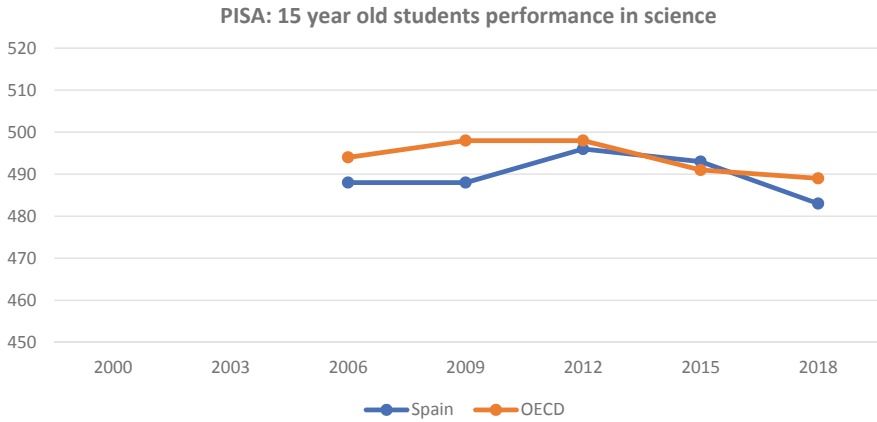
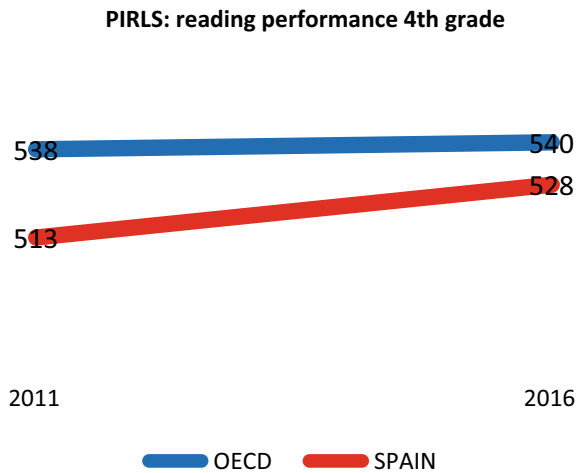


Fig. 5 Spain versus OECD: science performance over time (PISA)

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



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

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

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

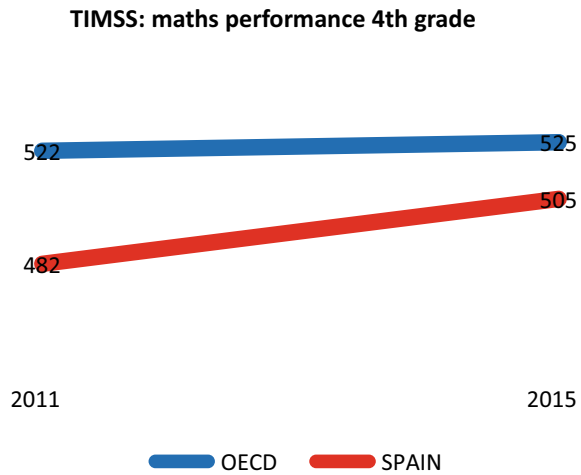
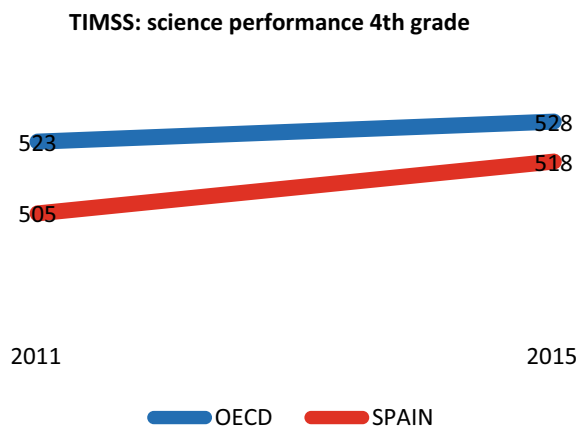


Fig. 8 Spain versus OECD: reading performance over time (TIMSS)



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

2.3 What PISA Reveals About Regional Differences

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

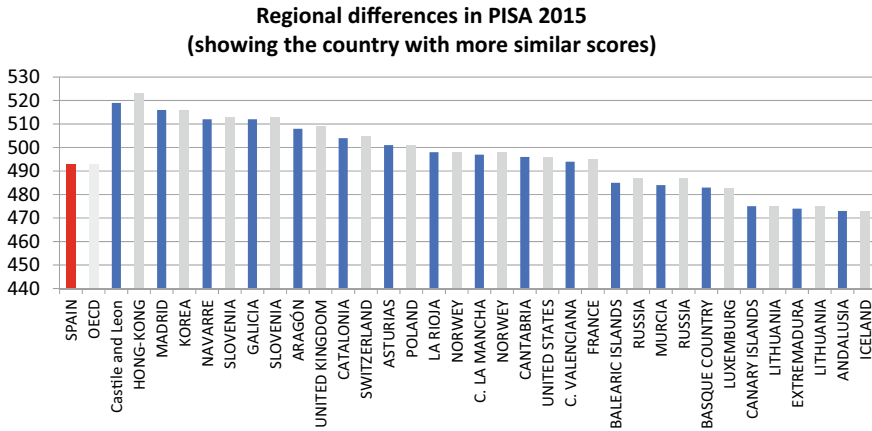


Fig. 9 Regional differences in Spain according to PISA 2015

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

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

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

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

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

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

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

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

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

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

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

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

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

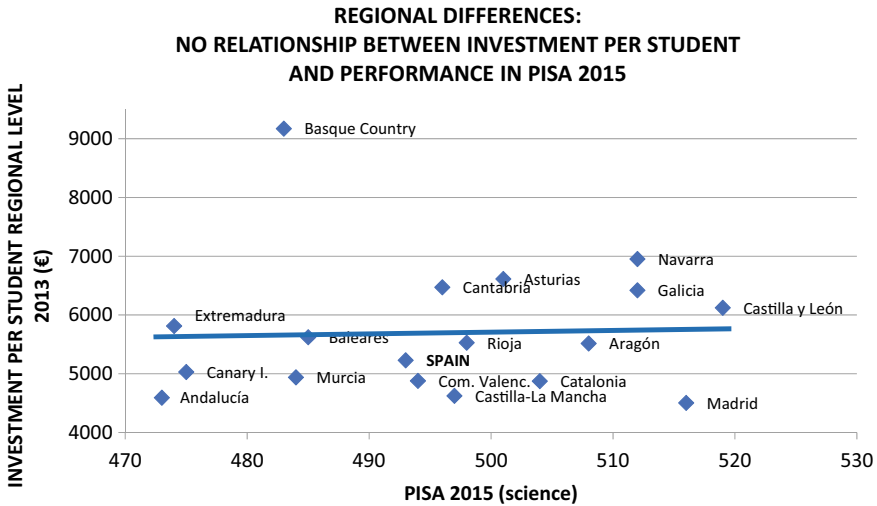


Fig. 10 Relationship between investment per student in each region and student performance according to PISA 2015 (modified from Wert 2019)

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

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

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

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

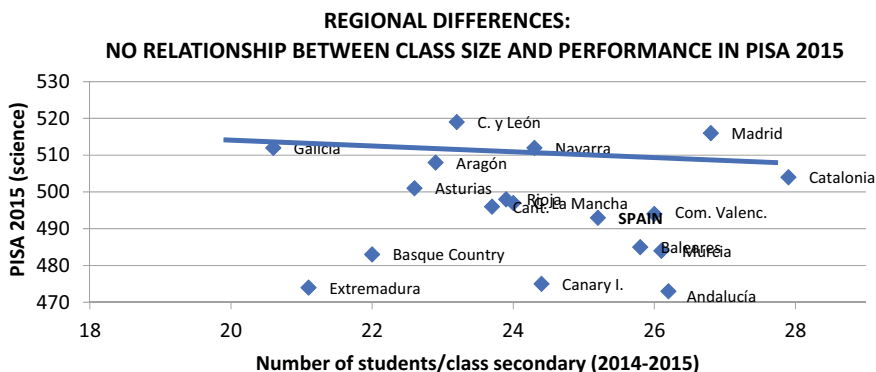


Fig. 11 Relationship between class size in each region and student performance according to PISA 2015 (modified from Wert 2019)

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

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

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

2.5 The Impact of Education Policies

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

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

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

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

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

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

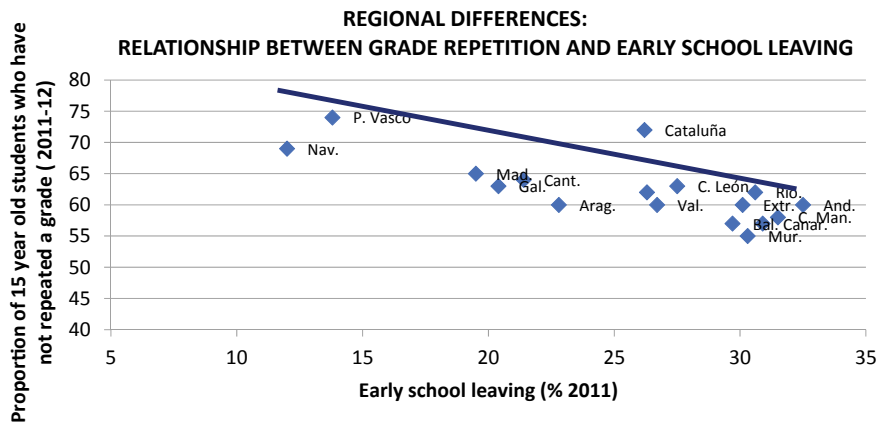


Fig. 12 Relationship between grade repetition and early school leaving (modified from Wert 2019)

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

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

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

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

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

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

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

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

3 What Happened in PISA 2018: A Broken Thermometer?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

4 Conclusions

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

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

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

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

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

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

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

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

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

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

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

Authors' ADDENDUM (27 July)

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

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

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

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

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

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

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

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

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

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

References

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

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

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

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

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

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

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

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

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



Taiwan: Performance in the Programme for International Student Assessment



Su-Wei Lin, Huey-Ing Tzou, I-Chung Lu, and Pi-Hsia Hung

Abstract Taiwan has, from 2006, participated in five Programme for International Student Assessment (PISA) surveys. This chapter discusses Taiwan's performance in PISA and its implications. At first, the education system and the process of educational reform in Taiwan were described. Then Taiwan's performances for reading, math, and science in PISA were delineated. Taiwanese students have had consistently excellent performance for math and science; its reading performance, although not as outstanding as those for math and science, has improved significantly from 2009 to 2018. The gender gap in reading, in favour of female students, has narrowed, and the gender gap in math and science has been small. Educational equity, especially between rural and urban students, has also improved from 2006 to 2018. The proportion of high performers in reading and the proportion of low performers in reading, math, and science has increased from 2006 to 2018, while the proportions of top performers in math and science have decreased. These findings are interpreted from the perspectives of cultural beliefs, changes in the education system and national assessment, government investment in the related domains, and the nature of the PISA assessment.

1 Introduction

The Programme for International Student Assessment (PISA), organized by the Organisation for Economic Co-operation and Development (OECD), is a cross-national survey conducted once every 3 years. PISA assesses a country's performance profile with respect to the core competencies required for 15-year-old students to participate in future society.

S.-W. Lin · H.-I. Tzou · P.-H. Hung
Department of Education, National University of Tainan, Tainan City, Taiwan

I.-C. Lu (✉)
Department of Education, National Pingtung University, Pingtung City, Taiwan
e-mail: ichung@mail.nptu.edu.tw

Specifically, through evaluating the performance of students in foundational competence domains and gathering information on students, teachers, and schools, PISA provides an overall score that describes how well a country's students are performing. This score aids countries in the adjustment of their educational policies, and educational decision makers attach great importance to PISA: the results allow such decision makers to compare their students' performance with those in other PISA-participating countries and economies, thus helping them better understand the future competitiveness of their students. Many participating countries publish their own PISA reports, and public discussions often cite PISA materials. Although the PISA survey cannot confirm that educational input has a causal relationship with the PISA results, these results are valuable because they allow educators, policy-makers, and the general public to understand the similarities and differences between education systems. Furthermore, PISA attracts much attention from global media, indicating the PISA's impact. Some countries have also started to develop and implement PISA-related assessments as additional projects or as part of their national assessments.

Taiwan has participated in the PISA survey five times, beginning from PISA 2006. The PISA 2018 survey focused on reading—with science and mathematics as minor evaluation areas—where it specifically examined students' attitudes toward and motivation for reading.

This chapter is organized as follows: First, the education system and PISA-related educational policies in Taiwan are briefly reviewed. Subsequently, Taiwan's results in the five PISA surveys (2006–2018) are discussed, with regard to trends, gender difference, social equity in learning outcomes, and changes in top and low performers. It then concludes with implications and policy recommendations.

2 The Education System in Taiwan

The present-day education system in Taiwan has a 6-3-3-4 structure. It was established in 1949, then having only 6 years of compulsory primary school education. To reduce competitive pressure in middle school admissions and because a more highly skilled workforce was needed for Taiwan's national development, Taiwan's 9-year compulsory education system was implemented subsequently in 1968, which had 6 years of primary school followed by 3 years of junior high school. The compulsory education system was free of charge, and it was in place for more than four decades until 2014, where it was extended to the present-day 12-year basic education system. This extension was aimed at developing a more highly skilled workforce for future economic growth. Although early childhood education (i.e., preschool) is not part of Taiwan's compulsory education system, the government has actively invested resources targeted at reducing the burden of financially disadvantaged families from sending their (specifically, 5-year-old) children to preschool (Ministry of Education [MOE], nd).

Students in Taiwan's education system have two important choices. This first choice comes after graduation from junior high school, where students can either choose to go to a senior secondary school or a 5-year junior college, depending on their interests as well as performance in the required Comprehensive Assessment Program (CAP). In general, senior secondary education in Taiwan comprises four school types: general, skill-based, comprehensive, and specialized senior secondary schools. Thus, to aid students in making this decision, in addition to the regular curriculum, technical arts education is included in junior high schools to offer students a greater diversity of learning opportunities. Thus, students do have the opportunity to better understand what vocational education will look like and explore future career options.

The second choice comes after graduation from senior secondary school, where students choose which college to go to. Most of these graduates will have taken the General Scholastic Ability Test (GSAT), and they obtain admission into a college of their choice through two paths (personal application and school recommendation) based on their GSAT results. For students without a place in a college of their choice or with unsatisfactory GSAT results, they can still take the Advanced Subject Test (AST) and obtain college admission based on their AST results as well as their preference list. Meanwhile, graduates of skill-based senior secondary schools can take the Technological and Vocational Education Joint College Entrance Examination to get admission into technical colleges or technical universities.

Figure 1 presents educational statistics for 2018. In particular, the enrolment rates for preschool, elementary school, junior high school, senior high school, and university or college were 63%, 97%, 98%, 94%, and 77%, respectively. The gross enrolment ratio and average years of schooling were 94% and 12.2 years. In Taiwan, primary and secondary school teachers were relatively young, and the proportion of teachers older than 50 years was approximately 20%. The average class size, for both primary and secondary schools, was approximately 26 students, and the education expenditure per student was more than NT\$203,000 (US\$6700). Total education expenditure accounted for 5.08% of Taiwan's GDP. More than half (56%) of primary and secondary schools in Taiwan were public but only 31% of higher education institutions were public. A detailed description of Taiwan's education system can be found in the 2019 edition of Education Statistics of the Republic of China (MOE 2019) and on the website of the Department of Statistics, MOE (http://stats.moe.gov.tw/files/ebook/Education_Statistics/108/108edu_ODF.htm).

Since the 1990s, Taiwan's MOE has been steadily engaging in educational reform. Initial reforms focused on ensuring that all students have access to a high-quality education, whereas recent reforms have focused on developing teacher capacity to foster the critical thinking and literacy skills needed in a fast-changing global economy. These recent reforms are part of Taiwan's response to criticism that its education system, in focusing too heavily on standardized tests, rewards rote memorization rather than the creative application of knowledge. This section outlines the evolution of educational policy in Taiwan over the past 30 years.

On April 10, 1994, several nongovernmental organizations in Taiwan organized a march and formed the 410 Alliance of Education Reform. The alliance made

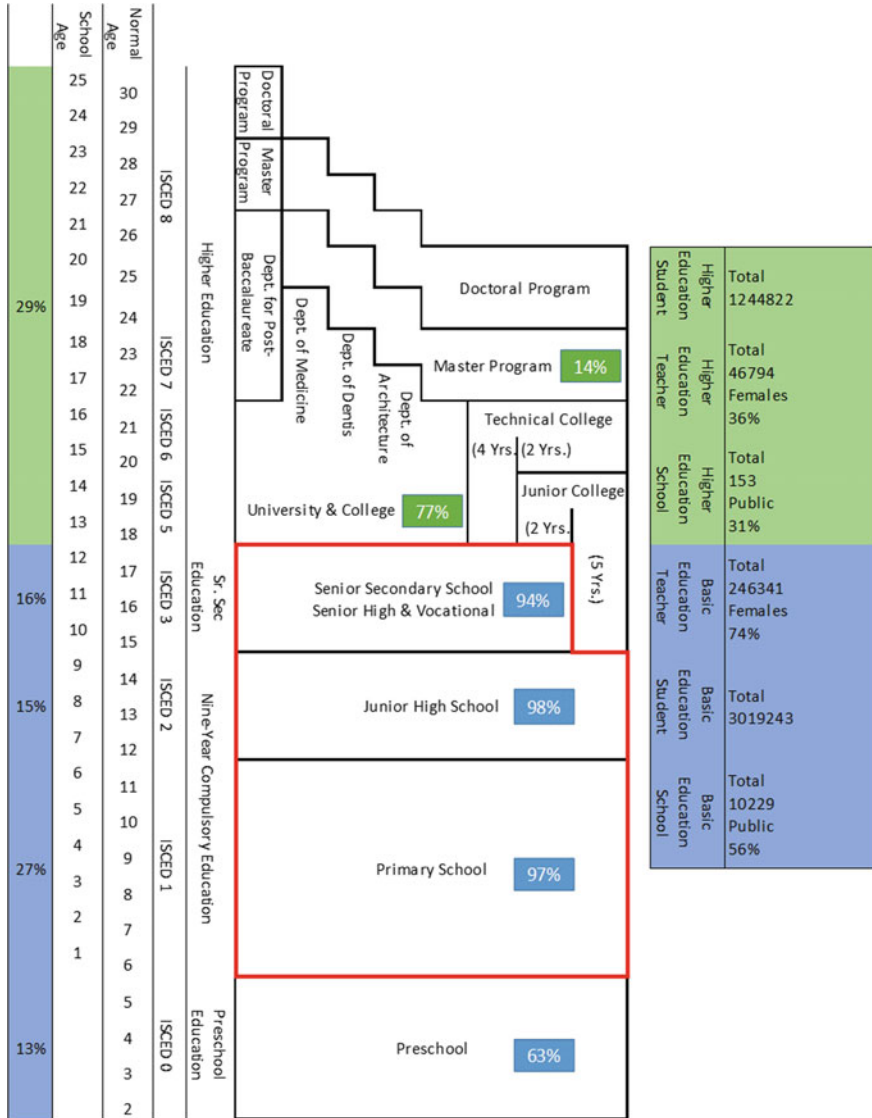


Fig. 1 Structure of the Taiwanese education system in 2018. Data Source MOE (2019)

four appeals for education reform: (1) to establish more senior high schools and universities, (2) to reduce the class- and school-size in primary and junior high schools, (3) to promote the modernization of education, and (4) to formulate the *Educational Fundamental Act*. This march was regarded as the birth of Taiwan’s education reform, and the four appeals became the primary axis of education reform.

In response to the public's demand, the Executive Yuan pledged to set up the Consultation Committee of Education Reform. In December 1996, the Committee proposed the *Consultants' Concluding Report on Education Reform*, which sought to relax the limitations in the education system, to take good care of every student, to accelerate school entrance paths, to enhance teaching quality, and to build a lifelong learning society. To implement the aforementioned proposal, the MOE proposed the *Action Plan for Educational Reform*, which was designed to implement 12 policy items within 5 years with a budget of more than NT\$157 billion. These policy items included increasing the education budget, strengthening education research, enhancing primary and junior high school education, universalizing early childhood education, improving preservice and in-service teacher education, promoting diversity and refinement in technical and vocational education, and making further education more accessible.

In 2010, the MOE published the *Education Report of the Republic of China*, which outlined the educational development blueprint for Taiwan over 2011–2020. This report proposed three visions (new era, new education, and new promise) and four goals (refinement, innovation, fairness, and sustainability). To fulfil these visions and goals, ten strategies were formulated: (1) promoting 12-year basic education and integrating kindergartens with nursery schools, (2) improving the education system and reinforcing education resources, (3) refining preservice teacher education and teachers' professional development, (4) promoting the transformation and development of higher education, (5) innovating the education industry and cultivating talent for the knowledge-based economy, (6) developing the literacies of diverse modern citizens, (7) promoting sports and a healthy lifestyle for all, (8) promoting respect for cultural diversity and the rights of disadvantaged groups as well as those who need special education, (9) expanding cross-strait, international, and overseas Chinese education, and (10) deepening lifelong learning and cultivating a learning society (MOE 2010).

Because humans are the most important resource and their talent is key to national development, *MOE Talent White Paper* was published in December 2013, which proposed a 10-year blueprint for cultivating talent over 2014–2023. The proposed blueprint is illustrated in Fig. 2. It included two visions of “cultivating excellent and creative people” and “improving Taiwan's international competitiveness,” in addition to 12 themes of administration.

Beyond the *Action Plan*, *Education Report*, and *White Paper*, Taiwan's education underwent two significant innovations in the past two decades. The first was the 2001 replacement of the Joint High School Entrance Examination with the Basic Competence Test (BCtest) for junior high school students. The BCtest was a required test for all junior high graduates and evaluated competence in five subjects: Mandarin, Mathematics, English, Science, and social studies. All BCtest items were multiple choice except for the Writing Assessment, which was included in 2007. Students could take the BCtest twice in 1 year and use their better score for admission into a senior secondary school. However, as part of the implementation of Taiwan's 12-year basic education policy, the BCtest was replaced with the CAP in 2014. Unlike the BCtest, the CAP was only administered once a year, and a section evaluating English

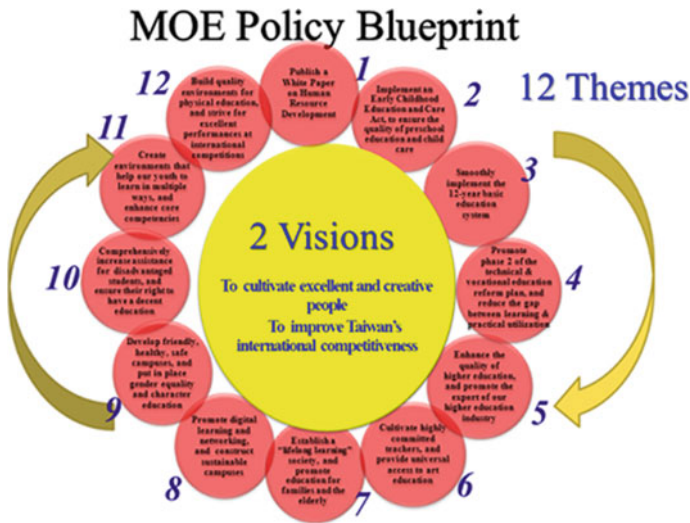


Fig. 2 MOE's 2014 policy blueprint

listening comprehension was added. Rather than a scaled score, the test results were reported in three levels—proficient, basic, and improvement needed.

In addition to changes in the examination system, curricular reform was another milestone in Taiwan's educational reform. In keeping with global 21st century trends of educational reform, the MOE initiated curricular and instructional reforms in primary and junior high school education based on the *Action Plan for Educational Reform*. Because the curriculum is foundational to schooling and instruction, the MOE prioritized the development and implementation of grade 1–9 curriculum. The *General Guidelines of Grade 1–9 Curriculum* was promulgated in 1998. Meanwhile, the MOE decided to introduce the Grade 1–9 Curriculum gradually, beginning from the 2001 academic year. At its core, the Grade 1–9 Curriculum was student-centred and focused on life experiences to cultivate students' 10 basic competencies. The new curriculum had five new features: (1) replacing knowledge with basic competency in students, (2) providing English instruction in primary education, (3) emphasizing the integration of learning areas, (4) focusing on school-based curriculum design, and (5) integrating instruction and assessment.

Subsequently, in response to the implementation of the 12-year basic education system, Taiwan's MOE released the *Curriculum Guidelines of 12-Year Basic Education—General Guidelines* in November, 2014. The Curriculum Guidelines were based on the newly adopted concepts of *taking initiative*, *engaging in interaction*, and *seeking the common good* to encourage students to become spontaneous and motivated learners. The visions of the new curriculum were to develop talent in every student—nurture by nature—and promote lifelong learning. To implement the ideas and goals of the 12-year basic education policy, core competencies were used as the basis of curriculum development to ensure continuity between educational stages

and encourage integration between domains as well as subjects. The concept of core competency underscores how learning should not be limited to the knowledge and skills taught in school, where learning should instead engage real-life scenarios and emphasize holistic development through action and self-development. The notion of lifelong learning constituted the heart of the core competencies in 12-year basic education. Figure 3 illustrates the aforementioned concept of core competencies.

Because the PISA survey is a low-stakes assessment for Taiwanese students, the PISA survey does not directly implicate education policy in Taiwan. Before the PISA survey, the 1999 *Third International Mathematics and Science Study-Repeat* (TIMSS-R) was the first international large-scale assessment (ILSA) taken up by Taiwan. Since then, Taiwan has been involved in many large-scale international assessments, such as the TIMSS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study), PISA, ICCS (International Civic and Citizenship Education Study), and TALIS (Teaching and Learning International Study). Taiwanese students have had outstanding performances in both the 1999 TIMSS-R and 2003 TIMSS. However, there was a large gap between low- and high-performers in the 2003 TIMSS. To shorten the achievement gap and improve educational equity, the MOE promoted the *After-School Alternative Program* (ASAP) in 2006 (Lin et al. 2013).

As the performance of Taiwanese students in the PISA reading assessment was unsatisfactory, the MOE, National Science Council (present-day Ministry of Science

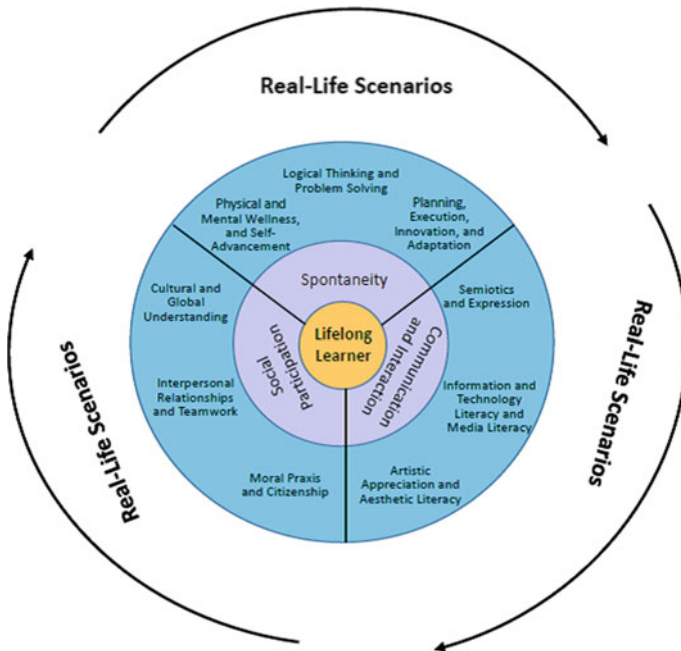


Fig. 3 Wheel-in-action diagram of core competencies. Source MOE (2014). Figure 1

and Technology), and local government education bureaus have allocated additional funds for reading instruction, hoping to improve reading performance. Research on effective reading instruction has been encouraged, and professional learning communities on reading instruction among teachers have also been promoted. The MOE also combined the ASAP and *Educational Priority Areas Project: Learning Guidance* into a new project called *Project for Implementation of Remedial Instruction* in 2013. Students who have not passed the remedial education screening test were the target students of this program.

3 Taiwan's PISA Performance

The first PISA survey joined by Taiwan was the 2006 PISA. Taiwanese students performed very well in mathematics (549 points) and science (532 points) in the 2006 PISA and have maintained such performance.

This impressive performance in mathematics and science can be explained by several factors. The first factor is cultural norms regarding learning. Influenced by Confucianism, Taiwanese society places a premium on education, holds teachers in high esteem, emphasizes student discipline and attention in the classroom, and prioritizes both repeated practice and a firm grasp on foundational knowledge (Tan 2015a, b, 2017). Furthermore, effort, instead of innate ability, is emphasized as the basis for achievement (Stevenson et al. 1993), and academic achievement is believed to be the key to future success (Wei and Eisenhart 2011). Parents, especially mothers, are also highly involved in their children's education, and they demand effort and good grades from their children (Fejgin 1995). As a result of these cultural beliefs toward education and parenting, relative to US students, Taiwanese students spend more time on homework, receive more help from family members with homework, and have more positive attitudes toward homework (Chen and Stevenson 1989). Meanwhile, because of this emphasis on effort, many Taiwanese students attend *buxiban*, which are private after-school programs that help students attain high grades in standardized tests. Mathematics and science are especially popular subjects focused on in *buxiban*. In addition to these cultural beliefs, a highly competitive education system contributes to Taiwanese students' excellent performance in mathematics and science. As mentioned earlier, grade 9 students are required to take the CAP in order to get into different types of senior secondary school. Taiwanese students are therefore under much pressure to perform well in the national examination.

Taiwanese students had unsatisfactory reading performance in the PISA 2006 (496 points). For all five PISA surveys Taiwan has participated in, Taiwan's mean readings scores were consistently and considerably lower than its mean scores in mathematics and science. This gap is unusual because reading is foundational to learning, regardless of the subject matter. Several reasons may explain this gap. First, the belief that mathematics and science are more important than language arts is prevalent in Taiwanese society, especially among parents and teachers. Students are encouraged to study further in science, technology, engineering, and mathematics (STEM)-related

areas and discouraged to pursue a career in arts and humanities areas. When students enrol in *buxiban* outside of school, they rarely enrol in classes on Chinese/Mandarin. Second, many in Taiwan, even school teachers, believe that mathematics and science are more difficult than Chinese/Mandarin. In addition to the focus on mathematics and science in *buxiban* classes, junior high schools are also more likely to offer remedial instruction in mathematics and science than in Chinese/Mandarin. Third, except for the writing test, the national assessment comprised only multiple-choice items before 2014. Thus, Taiwanese students have had little experience with constructed-response items, which constitute approximately 40% of items in the PISA reading assessment. All these factors may contribute to Taiwan’s relatively poor performance in the PISA reading assessment.

In the PISA 2018, Taiwan’s scores for reading, mathematics, and science were 503, 531 and 516, respectively. Comparing to the performance in PISA 2009, in which year the major domain was reading as well, the reading score of 503 constituted an 8-point improvement for Taiwanese students. Taiwan has also come to score better relative to the OECD average in reading: Taiwan scored 2 points higher in 2009 but 16 points higher in 2018, improving from the “not significantly different from the OECD average” group in 2009 to the “significantly higher than the OECD average” group in 2018. However, Taiwan’s PISA performance in mathematics and science has decreased in 2018: Taiwan’s mathematics and science scores were, respectively, 11 and 16 points lower than its corresponding scores in PISA 2015.

At present, Taiwan has had results from five PISA cycle surveys. Figure 4 displays the 12-year evolution in the three domains of reading, mathematics, and science for Taiwan and the OECD average. The evolutionary trends between the three domains differed for Taiwan. Reading performance had a hump-shaped trend, due to both Taiwan’s outstanding performance in the PISA 2012 as well as similar performance among the other cycle surveys. Mathematics performance had a gradual downward trend, although it remained outstanding among the participating countries. Science

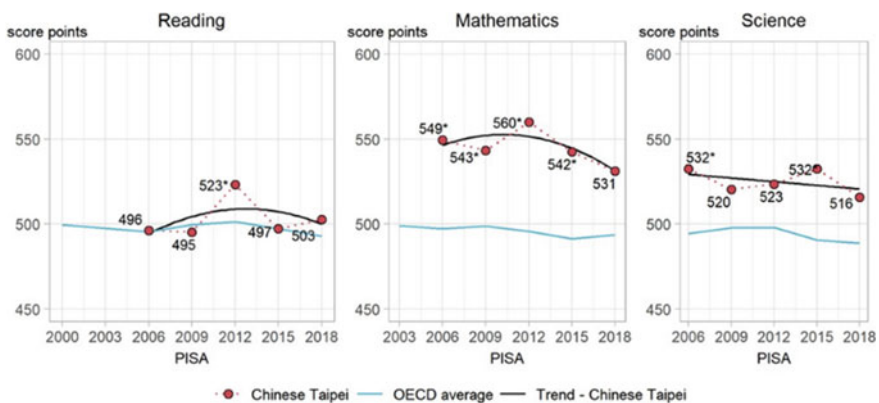


Fig. 4 Trends for Taiwan and the OECD average in reading, mathematics and scientific literacy. Source OECD (2019a), Chinese Taipei—Country Note—PISA 2018 Results, Fig. 2

performance remained stable, although the mean science performance in the latest PISA 2018 was the lowest ever.

The downward trend in Taiwan’s mathematics performance is a worrying finding. The authors can’t help but ask, has the mathematics competence of Taiwan students really declined? Taiwan has the CAP for grade 9 students, but the CAP does not employ a common scale and its testing results over the years can’t be compared directly, nor can the results be used to construct an evolutionary trend. Furthermore, the CAP is a comprehensive exam that is closely related to Taiwan’s national curriculum, which measures achievement in curricular knowledge that differs from those measured in the PISA assessment. Thus, the CAP provides only limited information for elucidating the downward trend in PISA mathematics performance in Taiwan.

We speculated that three reasons may explain this downward trend in Taiwanese students’ PISA mathematics performance. The first is testing fatigue. In addition to PISA, Taiwan has participated in several international assessments, such as TIMSS, PIRLS, ICCS, and TALIS (see Fig. 5). Including field trials and main studies, since 2006, Taiwan has conducted one to two large-scale assessments almost every year. The Taiwanese public was initially interested in Taiwan’s performance in these large-scale assessments, but frequent testing resulted in testing fatigue among schools, teachers, and students. Furthermore, these assessments were low-stakes tests for students, and they could not attract the sustained attention of the Taiwanese public. Teachers, students, and parents thus preferred to put their effort into high-stakes tests, such as the CAP, rather than these international assessments.

The second reason is student unfamiliarity with the PISA’s computerized testing format. The computerized testing was adopted since the PISA 2015 survey. The emphasis was placed on the use of technological tools for solving mathematics literacy-related problems. The CAP, by contrast, is a paper-and-pencil test and focuses on assessing the student’s acquisition of foundational knowledge. The proportion of improvement needed students (low performers) in the CAP math test has declined year by year, indicating that Taiwanese students’ mathematical competence has improved. However, technology is not widely used in mathematics classrooms in Taiwan and students receive little instruction from teachers on how

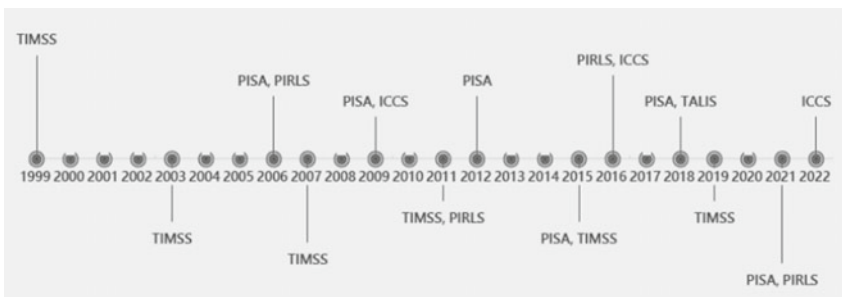


Fig. 5 Timeline of Taiwan’s participation in large-scale international assessments

to use technology to solve mathematical problems. Therefore, the improvement of Taiwanese students in the CAP math test was unable to reflect on the performance in PISA math assessment.

A less competitive education system in Taiwan over time may be the third reason. Because Taiwan has had a low birth rate, the number of examinees in the CAP decreased from 285,295 in 2015 to 215,219 in 2018. This large decrease has meant that fewer students are competing for admission into any given high school, which decreases the incentive, especially among lower-performing students, to study hard. This potentially explains the increased number of low-performing students. Furthermore, because the results for the CAP, a criterion-referenced test introduced in 2014, are of only three levels—proficient, basic, and improvement needed, high-achieving students have little incentive to aim for a perfect score and to do their best effort. That is may be one of the reason why, in PISA 2018, the proportions of high-performing Taiwanese students in mathematics and science were lower relative to previous PISA surveys.

3.1 Top and Low Performers

To aid interpretation of student scores, PISA divides student performance into several proficiency levels, where levels 5 and 6 indicate high performance and below level 2 indicate low performance. In PISA 2018, 10.9% of Taiwanese students were at or above level 5, which was more than double the corresponding figure of 5.2% in PISA 2009 (Table 1). This revealed that Taiwan’s number of top performers in reading has increased significantly over the past 9 years. However, in PISA 2018, 17.8% of Taiwanese students did not reach level 2, which although lower than the 2018 OECD average (22.7%), was higher than the corresponding figure for Taiwan in PISA 2009 (15.6%). Therefore, Taiwan’s proportions of top performers and low performers in reading increased obviously between 2009 and 2018, explaining the

Table 1 Percentage of low and top performers in reading, mathematics, and science (2006–2018)

Year	Reading		Math		Science	
	Below level 2	Level 5 or above	Below level 2	Level 5 or above	Below level 2	Level 5 or above
	(<407.47)	(≥625.61)	(<420.07)	(≥606.99)	(<409.54)	(≥633.33)
2018	17.8	10.9	14	23.2	15.1	11.7
2015	17.2	6.9	12.7	28.1	12.4	15.4
2012	11.5	11.8	12.8	37.2	9.8	8.3
2009	15.6	5.2	12.8	28.6	11.1	8.8
2006	15.3	4.7	12	31.9	11.6	14.6

Source OECD (2019b, c), PISA 2018 Database, Table I.B1.7
 OECD (2007a, b), PISA 2006 Database, Table 2.1a, 6.1.a and 6.2a

reason that Taiwan's mean performance in reading did not improve significantly during this time period.

This 2009–2018 increase in the proportion of top performing Taiwanese students in reading is consistent with expectations. This is because the MOE has invested much resources into reading education after Taiwan's poor reading performance in the 2006 PISA and 2006 PIRLS. Nevertheless, the increase in the proportion of low performers in reading also suggests that educational resources alone are insufficient to improve the reading abilities of low-performing students, who require individualized remedial instruction on how to read adaptively and strategically. Reading instruction in general should also be individualized and reading skills for the new information age should be cultivated; professional development programs that hone teachers' abilities to conduct such reading instruction are urgently needed.

As for mathematics performance, in PISA 2018, 23.2% of Taiwanese students (and 10.9% of OECD students) were at levels 5 or above, and 14% were below level 2. The proportion of low-performing Taiwanese students has been stable across the past five PISA surveys, although the 2018 figure of 14% was the highest ever. Conversely, the proportion of Taiwanese students who were top performers decreased from 37.2% in 2012 to 23.2% in 2018, which explains the reasons that Taiwan's mean scores for mathematics performance decreased from 2012 to 2018.

As for science performance, in PISA 2018, 11.7% of Taiwanese students (and 21.9% of OECD students) were at levels 5 or above and 15.1% were below level 2. The 2018 figure of 15.1% for low performers was the highest ever, and the proportion of top performers decreased by 3.7% from 2015 to 2018. This decrease in the proportion of top performers and increase in the proportion of low performers explain the reason that Taiwan's mean performance in science in PISA 2018 was the lowest ever.

The decreased proportion of top performers in mathematics and science between PISA 2018 and previous PISA surveys may be attributable to changes in the aforementioned scale of national assessment. In the BCtest, a national assessment that preceded the CAP, performance for each subject was scored at a maximum of 80 points. Because a single point increase for a subject may decide a student's admission into a more elite school, top performers tended to study diligently to attain a perfect score. By contrast, as mentioned earlier, the declining birth rate and three-tier scoring system for the CAP has made Taiwan's education system less competitive, which has decreased the incentives for low and top performers to study hard.

3.2 Gender Differences

The PISA results have consistently indicated that in reading, female students outperform male students in most countries. This has also been the case in Taiwan. As indicated in Table 2, In PISA 2018, the difference between average female and male reading scores (average female reading score – average male reading score) was 22 points, which was significantly lower than the OECD average of 30 points. This gender gap narrowed significantly from 37 points in 2009 to 22 points in 2018. This

Table 2 Gender differences for Taiwan and OECD average in reading performance (2006–2018)

Year	Taiwan			OECD average		
	Girls	Boys	Gender differences	Girls	Boys	Gender differences
2018	514	492	22	502	472	30
2015	510	485	25	504	477	27
2012	539	507	32	516	478	38
2009	514	477	37	511	472	39
2006	507	486	21	511	473	38

Note Statistically significant values are indicated in bold

Source OECD (2019b, c), PISA 2018 Database, Tables II.B1.7.27–II.B1.7.42. OECD (2007a, b), PISA 2006 Database, Tables 6.1C and 6.2C

decreased gender difference was due to improvements in male students' reading performance and female students' reading performance remaining the same.

The gender gap in PISA performance can be further elucidated by considering top and low performers (see Table 3). For Taiwanese students in PISA 2018, 9.8% of male students and 11.9% of female students read at levels 5 or above; these figures were greater than the corresponding figures of OECD average in PISA 2018. These figures were also significantly greater than those for Taiwanese students in PISA 2009, with the increase for male students being far greater than that for female students. Regarding low performers, for Taiwanese students in PISA 2018, 21.3% of male

Table 3 Percentage of girls and boys at each proficiency level in reading (2006–2018)

	Girls					Boys				
	2006 ^a	2009 ^b	2012 ^b	2015 ^b	2018	2006 ^a	2009 ^b	2012 ^b	2015 ^b	2018
Below level 1c	2.2	0.1	0.2	0.4	0	5.2	0.3	1	1.6	0.2
Level 1c					0.5					1.9
Level 1b		1.5	1.1	3.1	3		5.5	3.8	5.6	6
Level 1a	9.7	7.9	5.4	10	10.7	13.2	14.9	11.6	13.6	13.2
Level 2	23.2	22.2	16.9	21.1	21.1	25.4	27	19.4	23.8	22.5
Level 3	34.7	36.2	30.1	32.3	28.9	33.3	31	29.7	30.4	26
Level 4	24.1	24.9	31.5	24.6	23.8	19.3	17.2	25.8	19.6	20.3
Level 5	6.1	6.5	12.8	7.7	9.9	3.5	3.1	7.9	4.9	8.6
Level 6		0.6	1.9	0.8	2		0.2	0.9	0.4	1.2

Note ^aThe lowest reading level of 2006 is only classified as level 1a, and the highest is only classified as level 5

^bThe lowest reading level for 2009, 2012 and 2015 is only classified as level 1b

Source OECD (2019b, c), PISA 2018 Database, Tables II.B1.7.2, II.B1.7.4 and II.B1.7.6

OECD (2016a, b), PISA 2015 Database, Tables B2.I.3, B2.I.7 and B2.I.11

OECD (2013, 2014), PISA 2012 Database, Tables I.2.2a, I.4.2a and I.5.2a

OECD (2010a, b), PISA 2009 Database, Tables I.2.2, I.3.2 and I.3.5

OECD (2007a, b), PISA 2006 Database, Tables 2.1b, 6.1b and 6.2b

students and 14.2% of female students read below level 2; these figures were lower than the corresponding 2018 OECD averages of 27.6% and 17.5%, respectively. However, these figures were also greater than those for low-performing Taiwanese students in PISA 2009: 20.7% of male students read below level 2 in 2009 compared with 21.3% in 2018, and 9.5% of female students read below level 2 in 2009 compared with a significant increase to 14.2% in 2018.

The greater improvement of male students in reading from 2009 to 2018 may be attributable to the new computerized format of PISA 2018. Previous studies have found that Taiwanese male students had lower motivation for printed reading than female students (e.g. Sung et al. 2003). The PISA 2018 survey also showed that Taiwanese male students had less interests in reading and spend considerably more leisure time on using ICT than their female counterparts. These findings seem to suggest that digital reading may be more attractive than printed reading for male students. They may have tended to be more interested and engaged in the PISA test when it was administered through a computer as opposed to through paper-and-pencil. If this explanation is correct, then teachers can use digital reading to encourage reading in male students.

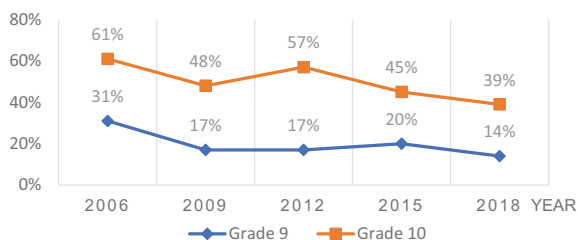
In contrast to the large gender gap in reading performance, the gender gaps among Taiwanese students in mathematics and science were smaller. In PISA 2018, male students slightly outperformed female students by 4 points in mathematics and 1 point in science, albeit nonsignificantly so. Such non significance was also noted for previous PISA surveys, except for PISA 2006, where male students outperformed female students by 13 and 7 points in mathematics and science, respectively.

This small gender gap in mathematics and science performance among Taiwanese students may be attributable to the following reasons. First, the Taiwan government has adopted a policy of cultivating female talent in science and technology. To address the gender disparity in STEM fields and to attract more young women into STEM, Taiwan's Ministry of Science and Technology has consistently invested money into promoting female role models as well as hands-on STEM activities and research opportunities in university laboratories for female secondary school students who have an interest in STEM. Second, the stereotype of STEM being a male field has become less entrenched in Taiwanese society, particularly among parents and teachers; girls are encouraged to pursue any field they may be interested in. Consequently, more female high school students have expressed interest in pursuing further study in a STEM field.

3.3 Social Equity in Learning Outcomes

Interschool disparities due to social stratification may affect students' learning opportunities and, by implication, educational outcomes. Educational systems with low interschool variability typically have high education equity. For example, in Finland, interschool variability in reading performance constituted less than 10% of the total variability for the country (OECD 2019d). Figure 6 presents the trend for interschool

Fig. 6 Variation in reading among schools by grade (2006–2018)



variability in Taiwan with regard to PISA reading performance. In Taiwan, 15-year-old students are mainly in grade 9 or 10. Grade 9 students are mostly in junior high schools in their neighbourhood, whereas grade 10 students are in senior high schools or vocational high schools, as primarily determined through the aforementioned admission system. Therefore, interschool differences in grade 10 are affected by the CAP examination, resulting in higher variation in grade 10 than in grade 9. As indicated in Fig. 6, interschool variation for reading in 2018 were the lowest ever. The total variation for the grades 9 and 10 decreased respectively from 61 and 31% in 2006 to 14 and 39% in 2018. This result suggests an improvement in educational equity in Taiwan from 2006 to 2018.

High educational equity can also be indicated by a low correlation of socioeconomic status with educational attainment in general and literacy performance in particular. The PISA index of economic, social, and cultural status (ESCS) enables a comparison between students and schools of different socioeconomic profiles. The ESCS slope of Taiwan in 2018 was equal to the OECD average, where a one-unit increase in ESCS was associated with a 37-point increase in the PISA reading score. ESCS explained 11.4% of the variance in Taiwanese students' reading performance, which was slightly lower than the corresponding figure of 11.7% in 2009.

Due to the increase in the standard deviation of Taiwanese students' reading performance across the PISA surveys, the total variation (sum of between- and within-school variation, see Table 4) gradually increased. However, variation in reading performance among schools did not change considerably over the years, while the proportion of between-school variation decreased from 47% in 2006 to 29% in 2018. This indicated that the performance difference between schools in Taiwan has been decreasing and the equity in education has been improved. Conversely, from Year 2006 to Year 2018, a one-unit increase in student ESCS level was associated with an increase 14–18 score points in reading performance. This suggested that the relationship of student ESCS and reading performance remained relatively stable from 2006 to 2018. Meanwhile, if the school ESCS level increased one unit, the school reading performance could increase 84–96 score points. In other words, the relationship between school ESCS and school reading performance was stronger than that between student ESCS and student reading performance.

The variation between schools and variation within schools in reading performance could be further explained by student ESCS and school ESCS. For example, student ESCS level explained 8.1% (2015) to 10.8% (2012) of the variation in reading

Table 4 Gradient decomposition of the PISA index of ESCS into inter- and intra-school components in reading assessment (2006–2018)

Year	Total variation	Variation between schools (%)	Variation within schools (%)	Change in reading performance associated with a one-unit increase in student ESCS level	Change in school reading performance associated with a one-unit increase in school ESCS level	Explained variance in student ESCS level (%)			Explained variance in student and school level ESCS (%)		
						Total variation	Variation between schools	Variation within schools	Total variation	Variation between schools	Variation within schools
2018	10,370	3042 (29)	7328 (71)	17	90	8.7	23.1	2.7	23.0	72.1	2.7
2015	8668	2904 (34)	5764 (66)	15	95	8.1	19.8	2.2	24.6	69.0	2.2
2012	8214	3201 (39)	5013 (61)	18	94	10.8	22.4	3.4	29.5	70.4	3.4
2009	7357	2515 (34)	4842 (66)	17	84	9.1	20.4	3.2	23.5	62.4	3.2
2006	7060	3322 (47)	3738 (53)	14	96	8.2	14.7	2.5	28.5	57.7	2.6

Source OECD (2019b), PISA 2018 Database; OECD (2016a), PISA 2015 Database; OECD (2013), PISA 2012 Database; OECD (2010a), PISA 2009 Database; OECD (2007a) PISA 2006 Database

Table 5 Mean score and variation in reading performance by urbanization level

Year	Urban		Suburban		Rural	
	Mean	SD	Mean	SD	Mean	SD
2018	524	102	495	99	475	96
2009	510	85	484	82	454	86

Source OECD (2019b), PISA 2018 Database; OECD (2010a), PISA 2009 Database

performance. If it was coupled with school ESCS, the explained proportion of total variation increased to 23.0% (2018) to 29.5% (2012). Addition of school ESCS amplified the variation in reading performance between schools. As seen in Table 4, the explained variance ranging from 14.7% (2006) to 22.4% (2012) was increased to a range from 57.7% (2006) to 72.1% (2018). Apparently and not surprisingly, school ESCS can explain the reading performance variation between schools. To further understand the impact of school ESCS, we narrowed our focus on results of 2009 and 2018 as reading literacy was assessed in both years. The total variation in reading performance explained by student ESCS was 9.1% and 8.7%, respectively; together with school ESCS, the total variation explained was increased to 23.0% and 23.5%, respectively. The increase by school ESCS in explaining the total variation was about 14%, which was roughly the same in both years. In terms of the variation in reading performance between schools, however, the explained variance was increased from 20.4 to 62.4% in 2009 and from 23.1% to 72.1% in 2018, when school ESCS was added to the model after student ESCS was already in. The 42% increase in 2009 and the roughly 50% increase in 2018 indicated that school ESCS played a more dominant role in explaining school differences in reading performance, after nearly ten years. To sum up, the results above implied that ESCS was an essential factor to equity in Taiwan's education, especially school ESCS.

The urban–rural gap also implicates educational equity. Table 5 presents the 2009 and 2018 reading performance of students from regions of different urbanization levels in Taiwan. According to the research report about the classification of levels of urbanization by Academia Sinica (Hou et al. 2008), we classified Taiwan's PISA-participating schools into three urbanization levels: urban, suburban, and rural. The reading scores for all three urbanization levels improved significantly between 2009 and 2018, by 14, 11, and 21 points for urban, suburban, and rural areas, respectively, where the greatest increase was for students in rural areas. Furthermore, the urban–rural gap in reading performance narrowed from 56 to 49 points, both in favour of urban students, from 2009 to 2018.

The increasing mean reading scores for students from the three urbanization levels and the decreasing urban–rural gap in reading performance indicate the effectiveness of government investment into reading educational resources for the past 10 years. The MOE has, since 2001, implemented a series of reading education projects. For example, the 2006–2008 *Reading Promotion Project for Schools in Rural Areas* was aimed at providing library resources, teacher training on reading instruction, and reading promotion activities for schools in less wealthy rural areas. The 2008

Reading 101: Reading Promotion Project and *2017 Promotion of Reading Education Project* were also focused on teachers’ professional development in reading instruction, in addition to training teachers librarians, developing reading materials, improving schools’ reading resources, improving the books and equipment of school libraries, providing funding to schools in rural areas for reading resources, and presenting awards to teachers with excellent performance in reading promotion.

4 Taiwan’s Performance in TIMSS and PIRLS

Taiwan has also participated in other ILSAs. Tables 6 and 7 detail the performance of Taiwanese students in PIRLS and TIMSS, respectively. As indicated in Table 6, grade 4 students in Taiwan have improved consistently and significantly from PIRLS 2006 to PIRLS 2011 and 2016. In 2016, PIRLS implemented a computer-based assessment. The score of Taiwanese students in the computer-based assessment was 546 points, which was significantly lower than the 559 points in the paper-based assessment.

Table 6 Taiwanese students’ performance in PIRLS and ePIRLS

Year	PIRLS	ePIRLS
	Mean (SE)	Mean (SE)
2006	535(2.0)	
2011	551(1.8) ▲	
2016	559(2.0) ▲	546(2.0) ■

Note ▲More recent year significantly higher, ■ Difference in PIRLS and ePIRLS statistically significant

Source Mullis et al. (2017)

Table 7 Taiwanese students’ performance in TIMSS

Year	Mathematics		Science	
	Grade 4	Grade 8	Grade 4	Grade 8
	Mean(SE)	Mean(SE)	Mean(SE)	Mean(SE)
1999	–	585 (4.2)	–	569 (4.2)
2003	564 (1.8)	585 (4.6)	551 (1.8)	571 (3.5)
2007	576 (1.8) ▲	598 (4.6) ▲	557 (2.0) ▲	561 (3.6) ▼
2011	591 (2.0) ▲	609 (3.2) ▲	552 (2.2)	564 (2.3)
2015	597 (1.9) ▲	599 (2.4) ▼	555 (1.8)	569 (2.1)

Note – only eighth graders in TIMSS 1999 surveys, ▲More recent year significantly higher, ▼More recent year significantly lower

Source Mullis et al. (2016)

TIMSS did not implement a computer-based assessment until TIMSS 2019, and the present results are thus based on the results of the paper-based TIMSS assessment. As shown in Table 7, Taiwan’s grade 4 students have continued to improve in mathematics and maintained their performance in science, and Taiwan’s grade 8 students had the best mathematics performance in 2011, with a slight decline in 2015. Taiwanese students’ performance in science have remained relatively stable, except for TIMSS 2007. Compared with the results in 2003, the grade 4 students’ science scores in 2007 improved significantly, while the grade 8 students’ scores decreased significantly.

Although the sampling targets are different in the three ILSAs, the target populations are all within the range of basic education in Taiwan. Also, Taiwan has accumulated 3 or 5 survey results in these ILSAs. Figure 7 attempts to present the evolutionary trend of Taiwanese students’ performance on these three ILSAs. The evidence provided by PISA on students’ performance does not seem to align with the results from PIRLS and TIMSS. As for reading, PISA has a hump-shaped trend and PIRLS has a positive, but flattening trend. The positive trend of the PIRLS results fails to replicate and extend to PISA. One important reason might be that the 15-year-old students in PISA put most of their effort into the high-stakes test—CAP and they did not try their best in PISA as the grade 4 students did in PIRLS.

As for math, PISA has a gradual downward trend and TIMSS has a hump-shaped trend in grade 8 but a positive trend in grade 4. The different trends between 8th and 4th graders in TIMSS-math are similar to those between PISA and PIRLS in reading performance. We do concern whether the trend of TIMSS-math of 8th grade



Fig. 7 Taiwanese students’ performance in PISA, PIRLS, and TIMSS

is going to evolve into a downward trend, like in PISA math. More information will be obtained to examine this concern after the results of TIMSS 2019 come out. The evolutionary trends of science between PISA and TIMSS are roughly stable. At a closer check, it reveals that the PISA results show a flat but slight downward trend and the TIMSS results show a flat but slight upward trend.

The inconsistency between the trend found in the math and science performance in PISA and TIMSS was understandable. PISA aims to assess 15-year-olds' mathematic and scientific literacy, while TIMSS relates more to the curriculum and instruction. The trend of the TIMSS results indicated that Taiwanese students had acquired a great deal of what has been taught in their math and science classes.

5 Concluding Remarks

From 2009 to 2018, the percentage of top performers of reading in Taiwan doubled from 5 to 11%. However, the disparity in reading literacy among Taiwanese students has also increased (standard deviations were 84 for 2006 and 102 for 2018) due to the much greater proportion of top performers and slightly greater proportion of low performers.

The gender gap in Taiwan's PISA scores was significantly smaller than the OECD average. This gender gap also narrowed significantly from PISA 2009 to PISA 2018 because of improved reading performance among male students and constant reading performance among female students. However, PISA 2018 results indicated a significantly increased gap between Taiwan's high and low performers in reading; this gap was higher than 260 points, which is equivalent to 6–7 school years. Greater effort should thus be made to reduce this gap in reading performance. For instance, reading instruction in rural and suburban schools should be strengthened by providing teachers with professional development programs and instructional resources. Assistance programs should also target the students whose reading literacy are below level 2 in PISA. Instruction should be tailored to the student's reading levels so that every student can be nurtured by the scaffolding appropriate for them.

Since 2018, PISA has adopted computer-based adaptive testing and employed diverse materials and reading elements to simulate the conditions of reading on the Internet. Such testing requires students to evaluate the quality and credibility of information as well as to detect and resolve conflicts between pieces of information. These have rarely been the focus of traditional teaching and paper-and-pencil assessments in Taiwan. To enhance Taiwanese students' reading performance, in addition to encouraging students to be more proactive toward reading, digital reading instruction and assessment should be included in schools. Teachers should also instruct students on how to read strategically as well as formulate and clarify reading goals.

Although mathematics and science were not the main assessment domains in PISA 2018, several facets of trends in PISA scores for mathematics and science still allow countries to track student performance. Taiwan's outstanding performance

in mathematics and science education has been the cornerstone of its competitiveness. The mathematics and science scores in 2018 did not change considerably from the scores in the previous surveys, and Taiwan has maintained its excellent performance, remaining in the top group globally and having significantly higher scores relative to the OECD average. The proportion of top performers in mathematics and science declined moderately between 2015 and 2018, and the proportion of low performers increased. The difference between high and low performers in PISA scores for mathematics and science has decreased slightly over time.

The global demand for highly skilled technical human resources has been rapidly growing and the competition for talent has thus intensified across the globe. Examining students' performance in reading, mathematics, and science helps countries evaluate their future talent pool. The percentage of Taiwanese students who were top performers in all three domains was 6.7%, which was twice the OECD average. However, 9.0% of Taiwanese students were low performers in all three domains. This figure merits attention despite being lower than the OECD average, where low-performing students will face difficulties in their careers and further study. Educators must continue to provide high-quality and differentiated instruction to support these students.

PISA attaches great importance to educational equality, and its results serve as a reference indicator that allows for comparison across countries, in addition to elucidating interschool variation, gender gaps, and urban–rural gaps, and the relationship between student SES and educational performance. The correlation between ESCS and reading performance among Taiwanese students was similar to the OECD average. This correlation decreased slightly from 2009 to 2018. The proportion of interschool variation decreased from 32% in 2009 to 29% in 2018. As for the gender gap, the male–female gap in reading PISA scores decreased from 37 to 22 points from 2009 to 2018, both in favour of female students. The urban–rural gap in Taiwan's reading PISA scores narrowed slightly from 2009 to 2018. Overall, Taiwan's education parity indicator indicated a slight improvement in educational equity from 2009 to 2018.

After PISA 2000, educational policymakers in many countries have referred to PISA results in their initiation of educational reforms. Although Taiwanese education policymakers do not undertake educational reforms based on PISA results alone, participation in PISA helps Taiwanese policymakers and educators to not only familiarize themselves with the concepts of literacy but also track the literacy performance of Taiwan's 15-year-old students using a globally held framework. Therefore, the *Curriculum Guidelines of 12-Year Basic Education* focus on core competencies, which are used as the basis of curricular development to ensure continuity between educational stages and integration between domains and subjects. The concept of core competencies encompasses the information, skills, and attitudes that a person ought to possess in their daily life and in the face of future challenges. The concept of core competencies underscores how learning should not be limited to the knowledge and skills taught in school and should instead engage real-life scenarios and emphasize holistic development through action and self-development. The new curriculum

in Taiwan's 12-year basic education policy is undoubtedly consistent with the notion of literacy measured in PISA.

Based on Taiwanese students' performance in PISA 2018, we strongly recommend the provision of more assistance to students who read below level 2 proficiency. We also recommend a focus on problem-solving skills and self-regulation in learning among students as well as the promotion of teachers' professional skills in literacy-based instruction and assessment. These recommendations are likely to be realized through the implementation of Taiwan's 12-year basic education policy, which will likely result in better performance in future PISA assessments by Taiwanese students.

References

- Chen, C., & Stevenson, H. W. (1989). Homework: A cross-cultural examination. *Child Development*, 60(3), 551–561.
- Consultation Committee of Education Reform. (1996). *The consultants' concluding report on education reform*. Retrieved from <https://www.naer.edu.tw/files/15-1000-7992,c1315-1.php?Lang=zh-tw>.
- Fejgin, N. (1995). Factors contributing to the academic excellence of American Jewish and Asian students. *Sociology of Education*, 68(1), 18–30. <https://doi.org/10.2307/2112761>.
- Hou, P. C., Tu, S. H., Liao, P. S., Hung, Y. T., & Chang, Y. H. (2008). The typology of townships in Taiwan: The analysis of sampling stratification of the 2005–2006 “Taiwan Social Change Survey”. *Survey Research—Method and Application*, 23, 7–32 <https://doi.org/10.7014/tcydffyy.200804.0007>.
- Lin, S. W., Hung, P. H., & Lin, F. L. (2013). *A remedial action based on Taiwanese students' results of TIMSS*. Paper presented in the 5th IEA international research conference, Nan-yang Technological University, Singapore.
- MOE. (2010). *Education report of the Republic of China*. Retrieved from <https://www.naer.edu.tw/files/15-1000-2976,c551-1.php>.
- MOE. (2013). *MOE talent white paper*. Retrieved from <https://ws.moe.edu.tw/001/Upload/3/RelFile/6315/6919/E6%95%99%E8%82%B2%E9%83%A8%E4%BA%BA%E6%89%8D%E5%9F%B9%E8%82%B2%E7%99%BD%E7%9A%AE%E6%9B%B8.pdf>.
- MOE. (2014). *The curriculum guidelines of 12-Year basic education: general guidelines*. Taipei, Taiwan: Department of Statistics of the Ministry of Education.
- MOE. (2019). *The 2019 edition of “Education Statistics of the Republic of China”*. Taipei, Taiwan: Department of Statistics of the Ministry of Education. Retrieved from http://stats.moe.gov.tw/files/ebook/Education_Statistics/108/108edu_ODF.htm.
- MOE. (n.d.). Educational reform. Retrieved from <http://history.moe.gov.tw/policy.asp?id=7>.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2015/international-results/>.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/pirls2016/international-results/>.
- OECD. (2007a). Database-PISA 2006. Retrieved from <https://www.oecd.org/pisa/data/database-pisa2006.htm>.
- OECD. (2007b). *PISA 2006: Science competencies for tomorrow's world*. Paris: OECD Publishing.
- OECD. (2010a). Database-PISA 2009. Retrieved from <https://www.oecd.org/pisa/pisaproducts/pisa2009database-downloadabledata.htm>.

- OECD. (2010b). *PISA 2009 results (Volume I): What students know and can do: Student performance in reading, Mathematics and Science*. OECD Publishing, Paris.
- OECD. (2013). Database-PISA 2012. Retrieved from <https://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm>.
- OECD. (2014). *PISA 2012 results (Volume I): What students know and can do: Student performance in Mathematics, reading and Science*. Paris: PISA, OECD Publishing.
- OECD. (2016a). PISA 2015 Database. Retrieved from <https://www.oecd.org/pisa/data/2015database/>.
- OECD. (2016b). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris: PISA, OECD Publishing.
- OECD. (2019a). *Chinese Taipei—Country Note—PISA 2018 Results*. Retrieved from http://www.oecd.org/pisa/publications/PISA2018_CN_TAP.pdf.
- OECD. (2019b). 2018 database—PISA. Retrieved from <https://www.oecd.org/pisa/data/2018database/>.
- OECD. (2019c). *PISA 2018 results (Volume I): What students know and can do*. Paris: PISA, OECD Publishing. Retrieved from <https://doi.org/10.1787/5f07c754-en>.
- OECD. (2019d). *PISA 2018 results (Volume II): Where all students can succeed*. Paris: PISA, OECD Publishing. Retrieved from <https://doi.org/10.1787/b5fd1b8f-en>.
- Sung, Y-T., Liu, P-Y., & Chien, H-Y. (2003). A report on the revision of the Motivations for Reading Questionnaire of the fifth- to seventh-grade students. *Psychological Testing*, 50(1), 47–71. <https://doi.org/10.7108/PT.200306.0047>.
- Stevenson, H. W., Chen, C., & Lee, S. Y. (1993). Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science*, 259, 53–58. <https://doi.org/10.1126/science.8418494>.
- Tan, C. (2015a). Teacher-directed and learner-engaged: Exploring a Confucian conception of education. *Ethics and Education*, 10(3), 302–312.
- Tan, C. (2015b). Education policy borrowing and cultural scripts for teaching in China. *Comparative Education*, 51(2), 196–211.
- Tan, C. (2017). Confucianism and education. *Oxford Research Encyclopedia of Education*. Retrieved from <https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-226>.
- Wei, M. H., & Eisenhart, C. (2011). Why do Taiwanese children excel at math? *The Phi Delta Kappan*, 93(1), 74–76.

Su-Wei Lin is a professor at the Department of Education (Measurement and Statistics, Master and Ph.D. Programs), National University of Tainan, Taiwan. She had been the co-PI in Taiwan PISA 2006, 2009, 2012, and 2018. She also has involved in many essential assessment projects in Taiwan, such as the Comprehensive Assessment Program for Junior High School Students, Project for implementation of Remedial Instruction-technology-based testing. Her research interests include the development and implication of assessments in mathematics education, primarily related to computerized adaptive testing, dynamic assessment, and large-scale assessment.

Huey-Ing Tzou is a professor at the Department of Education (Measurement and Statistics, Master and Ph.D. Programs), National University of Tainan, Taiwan. She had been the co-PI in Taiwan PISA 2009 and 2018. In addition to teaching and research, she has a lot of experience in involvement in national exams over the years. Her current research focuses on educational measurement and assessment, especially the application of cognitive diagnostic model on large-scale assessments.

I-Chung Lu is an associate professor at the Department of Education, National Pingtung University, Taiwan. She is the co-PI in Taiwan PISA 2018. Her research interests lie in children's

reading development and reading instruction. Her current research focuses on developing effective reading strategy instruction for elementary school children, especially the indigenous children, and establishing professional development programs in reading instruction for teachers.

Pi-Hsia Hung is a professor at the Department of Education (Measurement and Statistics, Master and Ph.D. Programs), National University of Tainan, Taiwan. She had been the PI in Taiwan PISA 2009, 2012, and 2018. She was also responsible for the development of many large-scale assessments in Taiwan, such as the Taiwan Assessment of Student Achievement, the project of After-School Alternative Program-technology-based testing, etc. Her current research focuses on educational measurement, educational assessment, computerized adaptive testing, and cognitive diagnosis.

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.



United States: The Uphill Schools' Struggle



Eric A. Hanushek

Abstract The United States has seen generally flat performance on both international and national tests. Moreover, the achievement gaps between disadvantaged and more advantaged students have been large and constant for a half century. The remarkable aspect of these outcomes is that federal and state programs have changed significantly—considerably greater resources, added school choice, test-based accountability, and school desegregation. Because of the importance of skills for the economy, it is important that the schools improve, but there is no indication of finding the set of policies that will do this.

1 Introduction

Some nations have reacted strongly to international achievement results, particularly after the introduction and expansion of PISA results that began in 2000.¹ The Germans were horrified with the initial results in 2000, while the Finns basked in the glory of high performance. The United States reaction was, however, at best subdued to the point of generally ignoring the results.

For those who have followed the PISA scores for the United States, there are few surprises. In terms of time trends across the subjects, the 2018 scores in mathematics and reading were not significantly changed over the entire period of PISA. The science scores were significantly better in 2018 than in 2006, but a substantial gap with the better performing nations remains.

¹PISA is the Programme for International Student Assessment, conducted by the OECD (<https://www.oecd.org/pisa/>).

²TIMSS is the Trends in International Mathematics and Science Study. It has been operational (with a changing group of countries) since the mid-1960s, and has been organized and run by the IEA (International Association for the Evaluation of Educational Achievement), which is an international cooperative (<https://www.iea.nl/>). See the summary of international tests in Hanushek and Woessmann (2011).

E. A. Hanushek (✉)
Hoover Institution, Stanford University, Stanford, USA
e-mail: hanushek@stanford.edu

The international scores on PISA and on the parallel TIMSS² testing program did not receive much attention until the Obama administration began publicizing the 2009 results. The fact that the international testing did not receive much attention does not mean, however, that there was no prior attention to student achievement. For almost 50 years, there has been consistent testing of U.S. students, and this permits tracking changes in performance over time. For the past two decades, it has also been possible to compare performance across U.S. states.

As described below, the different testing programs—PISA, TIMSS, and the longitudinal testing within the U.S., have given very similar pictures of the performance of U.S. students. Thus, there is no U.S. PISA shock, because the results from PISA can overall be seen in the other existing programs.

The picture is remarkable: First, with some nuances, overall U.S. performance has remained virtually constant for a half century; second, gaps in achievement across socio-economic groups have also remained constant for the past half century.

If attention to schooling and if programmatic elements of schooling were also constant, we could conclude this essay now. In other words, if a stagnant system produced constant results, there would not be much to say. But that is not the case. Schooling in the United States has changed in many ways. These ways have been focused on changing the performance picture, both in overall level and in the distribution of achievement. Therefore, it is useful to consider what policy changes have taken place along with the picture of constant results.

The overall story is simple. U.S. performance on international tests has never been good. There is a general notion in society that the schools should be doing better, and, toward that end, there have been large policy changes. Yet the changes that have been taken have not led to better outcomes. Even with a general appreciation for the economic importance of educational quality, the changes that have occurred have not been effective.

This chapter begins with an overview of the performance of U.S. students as seen from both international and national tests; this includes information on the level of achievement and the distribution of performance. It then turns to a discussion of the structure and organization of U.S. schools (Sect. 3) and of the major programs of the federal government (Sect. 4) and the state governments (Sect. 5). This discussion is followed by consideration of evidence about why this performance measures important things from the standpoint of the economy (Sect. 6) and why the U.S. has done better than would be expected based on the quality of its graduates (Sect. 7). It concludes with speculation about whether the good fortune of the U.S. economy will last if the schools do not improve.

2 Long Term Achievement Patterns

The international testing of achievement began in 1964 with the First International Mathematics Study. Of the 11 participating countries, the U.S. ranked tenth, beating

out Sweden.³ When the Second International Mathematics Study was conducted in 1980–82, the U.S. was in 13th place out of 17 participants—beating out Sweden, Luxemburg, Thailand, and Swaziland. Thus, it is not a surprise if a significant proportion of developed countries taking the tests outpace the U.S.

The overall trends in performance of U.S. students are easy to describe and are very consistent.

2.1 Pattern of PISA Scores

Since the beginning of PISA in 2000, the U.S. has been slightly above or slightly below the OECD average depending on the specific test. And it has stayed there. Figure 1 shows the performance on the separate reading, math, and science assessments of PISA. The dashed line in each panel shows the pattern of the OECD average. While some movement can be seen, the plots visually demonstrate the lack of significant movement.⁴

2.2 Pattern of National Assessment of Educational Progress (NAEP) Scores

The lack of surprise with PISA scores is easily explained by the pattern of scores on the U.S. National Assessment of Educational Progress (NAEP). This is an assessment given to a random sample of students using tests that can be linked over time. Figure 2 displays performance on the NAEP math and reading tests for students age 13 and age 17.⁵ The top two lines show reading and math scores of 17-year-olds, while the bottom two lines cover 13-year-olds. These patterns are best described as flat student performance over three to four decades—with one exception. The math performance of 13-year-olds rises significantly over the period. The puzzle, and the concern, is that higher middle school math performance does not readily translate into higher performance four years later in secondary schools. In any event, it is clear that the earlier performance improvements do not produce improved performance at the time that students are entering the labor force or further education.

³For a history of international testing along with scores on earlier tests, see Hanushek and Woessmann (2011).

⁴Note that the psychometric linking of the PISA scores occurred at different times for the separate subjects so that the reading series begins in 2000, math in 2003, and science in 2006. The U.S. does not have reading scores for 2006 because of a problem with the testing in that year.

⁵The National Assessment of Educational Progress has changed over time. The original test (Long Term NAEP) began in the 1970s and considered just a national sample. In 1990, an alternate test (Main NAEP) was introduced in order to provide state representative data. The Long Term NAEP collection was stopped in 2012. All data are cross-sectional for newly constructed representative samples.

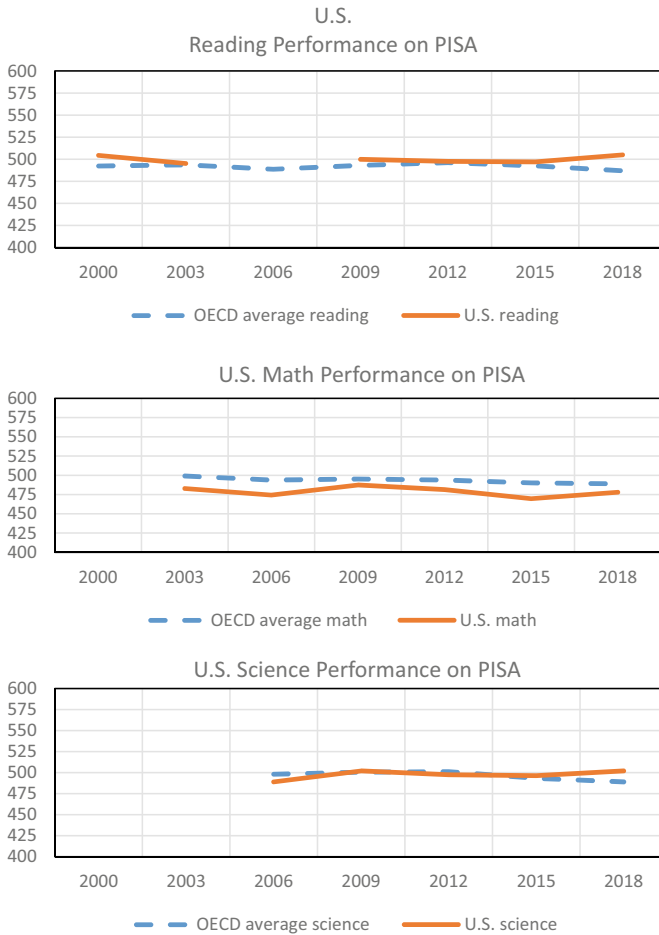


Fig. 1 United States PISA Scores, 2000–2018. *Notes* U.S. reading scores for 2006 unavailable because of a test administration problem. Aligned math tests begin in 2003, and aligned science tests begin in 2006. *Source* <https://nces.ed.gov/surveys/international/ide/>

2.3 Pattern of Achievement Gaps

Educational policy clearly has a variety of objectives, but the two recurring goals are higher overall achievement and equitable provision of education. What is a particularly important goal in most countries is using schooling and human capital investments to break the intergenerational transmission of poverty. When translated into achievement differences, this goal implies narrowing any gaps in student performance that are correlated with family socio-economic status (SES). Indeed, the U.S. has a wide range of programs (described below) that are aimed at improving the

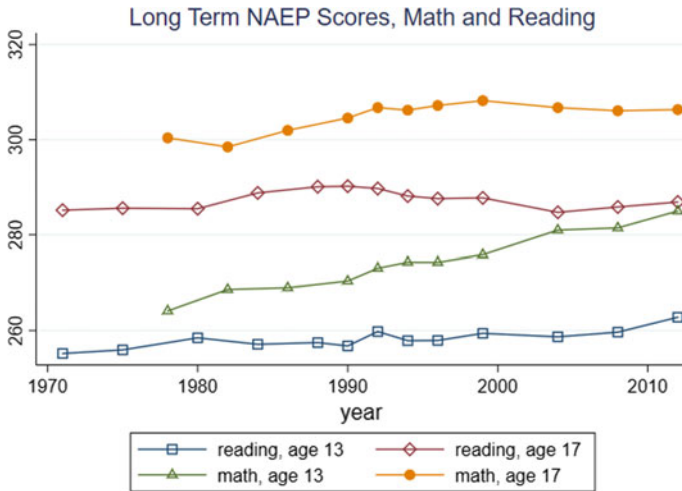


Fig. 2 Long Term NAEP Scores, Math and Reading. Source <https://nces.ed.gov/nationsreportcard/>

education and achievement of children from poor families. Here it is important to see what has happened to achievement gaps by SES.

Hanushek et al. (2020) combined test information from NAEP, TIMSS, and PISA with background information on the SES of each child. They then compared over time achievement of those in the top quarter of the SES distribution with those in the bottom quarter. Figure 3 shows the pattern of achievement gaps over the past half century. Achievement gaps have not changed!

After the 1954 desegregation of schools ordered in the U.S. Supreme Court decision of *Brown v. Board of Education*, the black-white achievement gaps narrowed until roughly 1990, but then progress stopped (Hanushek et al. 2020). The remaining gap is unacceptably large at roughly 0.9 standard deviations. This difference implies that the average black student is below the twentieth percentile of white students.

2.4 Conclusions on Achievement

The pattern of achievement—as seen by PISA or more broadly by NAEP—indicates little has changed over long periods of time. When broken down by SES of the family, the answer is the same—no movement over long periods of time.

To put this into perspective, it is important to see what changes in school programs and policies have occurred, because they will say something about what look to be good policies.

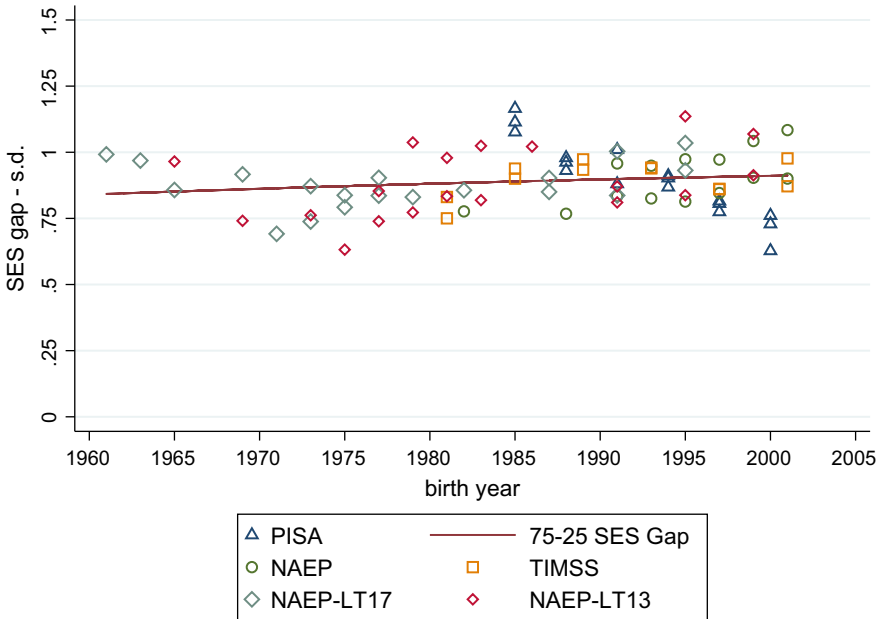


Fig. 3 Trend in the SES-Achievement Gap with Underlying Test Data, Birth Cohorts 1961–2001. *Notes* Achievement difference between the students in the top and bottom quartiles of the SES distribution (75–25 SES-achievement gap). The separate points are the 75–25 SES-achievement gap for the individual test administrations, and the line is the quadratic trend through the points. PISA is scores in math, reading, and science for 15-year-olds; NAEP is the Main-NAEP scores for eighth graders in math and reading; TIMSS is scores in math and science for eighth graders; NAEP-LT17 is the long-term trend NAEP scores in math and reading for 17-year-olds; and NAEP-LT13 is the long-term trend NAEP scores in math and reading for 13-year-olds. *Source* Hanushek et al. (2020)

3 Organization of U.S. Schools

The picture of U.S. schools is complicated from both a governance and a decision making viewpoint. By the U.S. Constitution, the individual states are the primary government body controlling schools, but this has interacted with the federal government in a variety of ways.

3.1 Governance

The U.S. education system is highly decentralized. At the beginning of the 20th Century, there was a federal Office of Education, which was not at the “cabinet rank.” Over the past century, there have been several attempts to enhance the federal role. In 1953, the Department of Health, Education, and Welfare was created at the

cabinet level, and the Office of Education was included along with federal health and welfare functions. In 1979, the Department of Education was created to give cabinet rank to federal education programs, although there have been periodic attempts to disband the department and to demote the status of education at the federal level. Notwithstanding the federal department, the states retain primary responsibility for education programs.

The states have always established separate programs that differ in terms of regulations, finance, local district autonomy, accountability, and ultimately performance. But, as discussed below, they have changed the operations and details of their systems considerably over time.

There was a dramatic consolidation of school districts following WWII. While there were 117,000 districts in 1940, this fell to 18,000 in 1970, and 13, 600 in 2016. There were 133,000 public and private schools in 2016.

Because each of the states is free within broad bounds to set its own policies, it is difficult to implement any common policies across the country. This also makes it difficult even to describe what actions and policies have been undertaken. There are, however, a few notable exceptions outlined below. But there are also common trends.

3.2 Resources and Expenditures

The first fact of U.S. schools is that expenditures have been rising very consistently—at least up to the time of the 2008 recession. Whenever discussions consider the pattern of achievement, they inevitably go to the resources available to the schools. Implicitly if not explicitly the argument inevitably turns to how resources are the answer to any improvements. Table 1 shows the pattern of resources over the past half century, both in terms of the components and of the overall spending per pupil. There were large decreases in pupil-teacher ratios with increases in teacher education. These changes added up to dramatic increases in real spending per pupil—over quadrupling between 1960 and 2016.

Table 1 Public school resources in the United States, 1960–2016

	1960	1980	2000	2016
Pupil-teacher ratio ^a	25.8	18.7	16.4	16.0
% teachers with master’s degree or more	23.5	49.6	56.8	56.4 ^a
median years teacher of experience	11	12	14	n.a
Real expenditure per pupil ^b (2017–18 \$’s)	\$2959	\$6675	\$10,131	\$12,330

n.a. Not available

^aData for 2012

^bData on expenditure per pupil are adjusted for inflation using the Consumer Price Index.

Sources U.S. Department of Education (2019)

It is difficult to argue from these data that the U.S. has overly tight with resources for the schools.

4 Federal Government Programs and Activities

The federal government has concentrated its attention on education of poor and disadvantaged students. These programs have been in place for a long time and have generally grown in size over time.

4.1 *The War on Poverty*

In 1965, President Lyndon Johnson declared a “war on poverty.” A major component of this was providing human capital to children from poor families so as to break the cycle of poverty. This compensatory education funding from the federal government, called Title 1 because of its legal foundations, led to a significant increase in funding, one that has grown over time.⁶

Soon after, the federal government initiated Head Start, a preschool program for 3- and 4-year olds from poor families. While never serving all poor children, this program also grew over time so that it served roughly 1 million 3- and 4-year olds, or roughly one-third of income-eligible students.

Finally, rounding out major programmatic support, the federal government legislated requirements for educating children with both physical and mental special needs in 1975. Support for this program has been split between the federal government and state governments. Over time, enrollment in special education has grown from 8.3% in 1976 to 13.7% in 2018. (On average, expenditures for special education students are roughly twice those for other children, although spending varies widely across different disabilities).

This set of federal programs underscores the fact that the U.S. federal government has programs chiefly driven by concerns about equity in education. Each of these programs is designed to support the education of disadvantaged students and is intended to reduce disparities in educational outcomes between children of poor families and children of better off families.

It is also important to note at the outset that funding for these programs was not tied to any specific use of the funds (other than general support for poor children). There are also no regular requirements that programs evaluate performance. As a result, periodic national evaluations of Title 1 compensatory funding and of Head Start⁷

⁶For a more complete history of Title 1 and of Head State, see Vinovskis (1999).

⁷The recent randomized evaluation of Head Start found that any positive effects disappeared by grade 3; Puma et al. (2010) and Puma et al. (2012). Note, however, that other studies have found

have not found them to be very successful in terms of increasing the achievement of the targeted students.

4.2 *Desegregation*

In 1954, the U.S. Supreme Court ruled in the court case of *Brown v. Board of Education* that *de jure* segregation of schools was unconstitutional. A number of southern states had previously had laws that separated students by race, and this led to an extended period of legal actions designed to desegregate schools.

Segregation of schools goes beyond the laws that were the subject of the *Brown* decision. Because housing tends to be segregated and because schools are based on local political jurisdictions, there is segregation of schools both because of local school attendance zones within large cities and because of differences in racial composition across school districts.

The *Brown* decision was followed by continuing legal and policy actions revolving around race and schooling. There was a significant rise in the chance of black students having white classmates through the late 1980s, but then the improvements lessened (Rivkin 2016). The main reason for the decline in exposure was the changing overall composition of U.S. students. White students went from 80% in 1968 to less than half today. The largest change has been the significant increase in Hispanic students who today make up over one-quarter of the public school population. The black student population has been quite constant since the 1960s at slightly over 15%.

There is no doubt that school desegregation led to better schools for black students. And, as was discussed above, this shows up in reduced achievement gaps between black and white students but a pattern of change that stopped a quarter century ago.⁸ The combination of changing demographics, policy changes, and legal decisions, led to progress that stagnated and imply that this area offers limited possibilities for improvement.

The changing composition of the overall student population does have potential impacts on the aggregate scores for U.S. students. If the immigrant population that makes up the majority of the increase in Hispanic students is also less prepared for school, demographics could influence the trends in achievement that are seen. Some simple calculations that use the changing demographic composition of the U.S. student population suggests, however, that this is not a very powerful force affecting the aggregate scores (Hanushek et al. 2020).

long term impacts of Head Start even if any achievement effects disappeared over time; Currie and Thomas (2000), Johnson and Jackson (2017).

⁸The improvement in outcomes related to desegregation is also found in the evaluation literature. See, for example, Angrist and Lang (2004), Hanushek, Kain, and Rivkin (2009).

4.3 Accountability

School accountability was universalized by the federal government when it passed the No Child Left Behind legislation (NCLB) in 2001. NCLB called for all states to institute annual student testing in grades 3–8 and once in high school, and it required regular reporting of achievement levels by major subgroup (race, poverty, and special education). In reality this was just an extension of the existing policy in a majority of the states.

NCLB set the target that all students had to be proficient (as defined by the individual states) by 2014. There were also intermediate goals that were to be met by each school between 2002 and 2014. If not met, there were various sanctions that were imposed as specified by the federal government: expanded student choice of schools, remedial programs, and ultimately elimination of failing schools.

Over time, it became apparent that few schools would actually meet the proficiency goals. Moreover, resistance to the entire program grew over time. As a result, the NCLB legislation was replaced in 2015 with the Every Student Succeeds Act (ESSA). While this federal law still required annual student testing, most parts of the design of the measurement system, its goals, and its remedial actions were returned to the individual states.

4.4 The Federal Government Role

In sum, the federal government in the U.S. has been particularly focused on equity goals and has introduced both funding and regulatory approaches to improving the achievement of students at the bottom of the poverty distribution. As Fig. 3 showed, these policies have not been successful in terms of narrowing achievement gaps.

5 State Programs and Policies

The main responsibility for schools in the U.S. resides with the individual states. The states in turn delegate considerable responsibility to individual school districts. (Only Hawaii has a single school district that coincides with the state).

5.1 School Finance Issues

The clearest way to see the state role is by observing the pattern of revenue raising over time. As Fig. 4 shows, a century ago almost all revenues were raised by individual localities. But this changed with the local share falling rather steadily. The largest

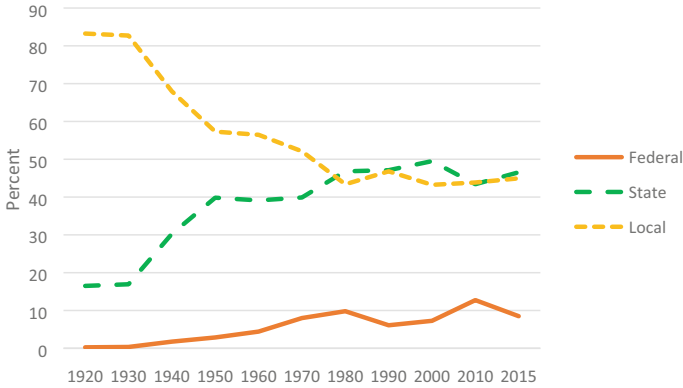


Fig. 4 Sources of U.S. School Revenue. *Note* Percentage shares of revenues for U.S. public schools. *Source* U.S. Department of Education (2018)

Table 2 Sources of state school revenue in 2015

	Federal	State	Local
Average	8.5%	46.5%	45.0%
Minimum	4.2	24.9	3.9
Maximum	14.9	90.1	66.8

Note This table reflects the range of revenue sources across states in 2015
Source U.S. Department of Education (2019)

changes in revenues came with two policy issues. First was the increase in federal spending that occurred with the War on Poverty in the 1960s leading to an increase of the federal government share to roughly 10%. The second was the beginning of court involvement in spending, starting around 1970 and continuing to today.

The court involvement started with law suits that argued that the funding of schools was not equitable across school districts. Since some districts found it easier to raise funds than others, a number of lawsuits were introduced individually across the states.⁹ Beginning with California in the late 1960s, almost all states have now faced law suits about the pattern of spending. The results of these suits, which sometimes require changes in funding and other times do not, has been a general increase in the state share of spending. The pattern of school revenues does, however, differ noticeably across states. As Table 2 shows, while two-thirds of revenues come from localities in Illinois, only four percent do in Vermont. Federal revenues also vary

⁹Local districts disproportionately raise revenues by property taxes. Since localities vary widely in the size of their tax base (which comes from the value of homes plus the value of commercial and industrial property in the district). States will general distribute funds to districts in ways that are inversely related to the local property tax base, but this seldom completely overcomes differences in tax bases. See Hanushek and Lindseth (2009) for a further discussion plus a history of the court involvement.

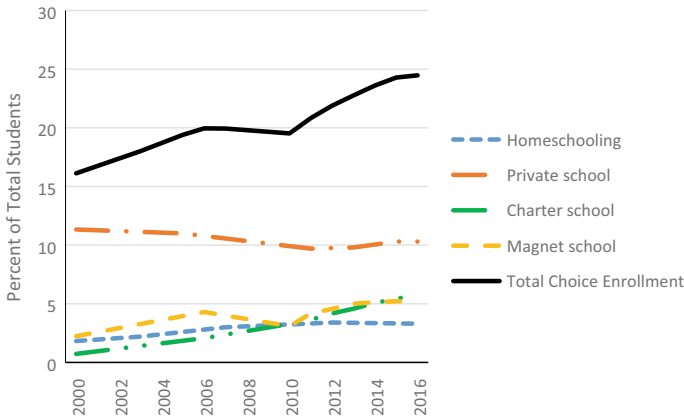


Fig. 5 Students Attending Schools of Choice. *Source* U.S. Department of Education (2018)

noticeably, depending on the overall level of spending in each state and on the proportion of students from poor families.

A different variety of lawsuits (“adequacy cases” instead of “Equity cases”) developed in the late 1980s. These put forward the general argument that, even if state funds were equitably distributed, the level of funding was not adequate to meet the achievement goals of the state. Again, these court cases pursued the general presumption that resources were the problem with the low achievement of students.

Importantly, the source of funds as well as the level of overall spending appears to have little to do with student performance differences across states (Hanushek 2003). Nor does the increase in spending levels relate to the increase in student performance (Hanushek et al. 2012).¹⁰

5.2 Choice: Private, Homeschool and Charter Schools

One thing that has been happening over time is substantial changes in the percentage of students actively choosing what kind of school they attend. As recently as 2000, 85% of students went to the traditional public school to which they were assigned (Fig. 5).¹¹ By 2016, one-quarter of students made choices of the sector of instruction.

¹⁰Some recent analyses, relying on the estimated impact of court decisions, have argued that extra spending has an impact. These are part of a continuing and unresolved debate. See Jackson, Johnson, and Persico (2016), Lafortune, Rothstein, and Schanzenbach (2018).

¹¹Note that these shares of students with choice do not include a number of districts that allow or require students to choose among the traditional public schools. Because all students stay within the traditional public schools, there is no pressure on the school district to try to keep the students. This feature differs from the other forms of choice with the exception of magnet schools. Magnet schools offer specialty curricula (academic, the arts, or other vocational focus), and they offer an alternative to the traditional schools.

Private schooling has been constant at roughly 10%, with the vast majority being religiously based. But charter schools—public schools that are not controlled by the local districts—have grown significantly (Baude et al. 2020). Perhaps most surprising has been a rising share of students who are home-schooled.

In sum, the U.S. has consistently moved toward more choice of schools. The micro evidence, however, does not show a clear impact of choice programs within the United States (CREDO 2013).

5.3 Common Core and Curriculum

One of the major education debates of the past decade has been whether to introduce a common curriculum across the nation. While the federal government cannot impose this, it did help to support the voluntary adoption of the “common core curriculum” across states. Initially over 40 states adopted the common core curriculum, but it became very controversial, and a number of states subsequently repealed it. The state alternatives to the common core, however, often had strong similarities. In the end, however, little evidence suggests superior results with adoption of the common core.

5.4 The State Government Role

The states are responsible for the quality of schools. For whatever reason, however, the policy choices have not led to improvements.

6 Why It Matters

Existing research shows a very strong and consistent relationship between scores on common standardized tests and economic outcomes. This linkage with future economic well-being motivates the attention to PISA and to alternative approaches to improving student performance. Surprisingly, most policy makers believe that education has important economic outcomes—and yet they often are unwilling to go very far to promote major changes.

6.1 Economic Growth of Nations

Economic growth determines the future economic wellbeing of nations, and virtually all empirical studies of the long-run growth of countries have highlighted a role for

human capital. The early economics literature overwhelmingly employed measures related to school attainment, or years of schooling, to test for the effects of human capital. But, average years of schooling is an incomplete and potentially misleading measure of education when comparing different countries. It implicitly assumes that a year of schooling delivers the same increase in knowledge and skills regardless of the education system. For example, a year of schooling in Peru is assumed to create the same increase in productive human capital as a year of schooling in Japan. It also neglects cross-country differences in the quality of schools and in the strength of family, health, and other influences is a major drawback in such research.

International achievement test scores can be thought of as measures of human capital differences across countries. Indeed, once long run growth rates across countries are related to international test scores, which in the aggregate we call “knowledge capital,” three-quarters of the cross-country variation in growth rates can be explained by differences in scores on international math and science tests. (See Fig. 6 and Hanushek and Woessmann 2015a). Moreover, there is reason to believe that this relationship is causal—i.e., if cognitive skills are raised, growth rates will increase (e.g., Hanushek and Woessmann 2012). These estimates indicate that just increasing school attainment without also increasing the amount of learning has no impact.

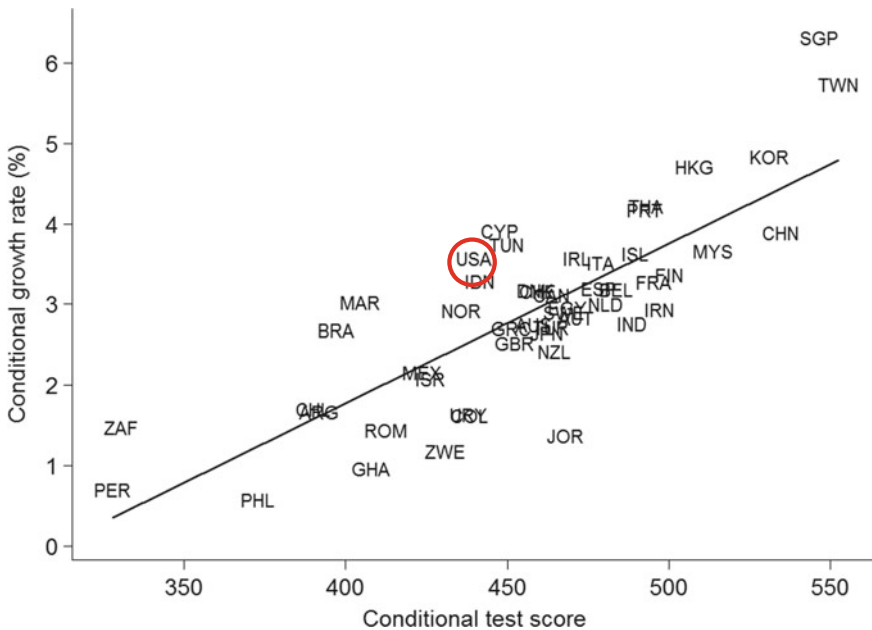


Fig. 6 Knowledge Capital and Long Run Economic Growth (1960–2000). Note: Added variable plot from regression of average annual growth rates in GDP per capita from 1960–2000 on average test scores of nations. The regression includes the level of GDP per capital in 1960 and average years of school attainment in 1960. Source Hanushek and Woessmann (2015a)

In other words, just getting students through more schooling without ensuring high levels of learning is not an effective policy.

The historical impact on economic growth of differences in test scores is large. One easy way to see the importance of cognitive skills is to project the economic value of school improvement (Hanushek et al. 2013; Hanushek and Woessmann 2015b). For example, consider the estimated impact of bringing just the bottom of the U.S. achievement distribution up to a basic skill level—i.e., a policy similar to the ideas behind U.S. accountability policies. Hanushek and Woessmann (2015b) estimate that, according to historical growth patterns, this would lead to average GDP levels that were 3.3% higher across the remainder of the century when compared to expected GDP levels with current skill levels. Such increases would be sufficient to deal with most of the fiscal problems suggested for the pension and medical systems.

While politicians may tend to underestimate the importance of education for economic growth, they by all public statements still think that education is extraordinarily important for the nation. Nonetheless, perhaps because it takes time to see the results of any improvements, they are unwilling to make difficult decisions in the short run.

6.2 *Economic Growth of States*

Given the high levels of mobility in the U.S., the work location of somebody might be very different from where the person grew up and went to school. As a result, states do not directly experience all of the results of their school systems. Therefore, while improving schools might be in the national interest, individual states might benefit less and thus might not have strong incentives to invest in better schools. The tension in America between centralized and decentralized education policy has been a pivotal policy issue for decades.

How schools affect state-level measures of economic output is a high priority concern for policy makers (and researchers). In a series of studies, Hanushek et al. (2016, 2017a, b) show that economic growth of individual states, just like nations, is dependent on the quality of the labor force as measured by standardized tests, i.e., the knowledge capital of states. Moreover, the relationship between worker skills and growth at the state level is virtually identical to that found internationally.

Because a majority of students educated in a given state remain in the state when entering the labor force, even with migration, it pays for each state to invest in improved school quality. But since the labor force in each state is comprised of both locally educated workers and workers educated in other states, the largest gains come when all states improve their school quality, as opposed to a single state.

6.3 *Individual Incomes*

The previous sections focused on the effects of improved school quality on aggregate economic gains at the state and national level. Considerably more research has focused on the relationship between education and individual earnings. Innumerable economic studies show that school attainment affects earnings and income. These studies, pioneered by Mincer (1970, 1974), showed that economic success depends heavily on schooling. Nonetheless, they suffer from many of the same problems described in the previous aggregate studies. In particular, they ignore quality differences in schools, and they ignore sources of skills outside of schools. As demonstrated by the landmark “Equality of Educational Opportunity” report, commonly known as “the Coleman Report,” families are very important, as are peers in schools, neighborhood influences, and more (Coleman et al. 1966). An extensive body of research documents the multiplicity of inputs in educational production (e.g., Hanushek 2002).

The alternative, as with the aggregate studies, is to use measured skill from standardized tests to capture the totality of individual skills from families, schools, and other influences. This approach also relates the research more directly to educational policy. It has not been pursued extensively in the past, largely because few data sources combine information on both skills and individual earnings.

Recent international data provide the ability to estimate the economic value to individuals of higher educational achievement. The OECD surveyed random samples of adults age 15–65 across 32 countries in the Program for International Assessment of Adult Competencies (PIAAC). This survey contained information on backgrounds of individuals and their labor market experiences along with giving them a series of standardized tests (see Hanushek et al. 2015, 2017).

Hanushek et al. (2015, 2017) estimate the economic returns to greater individual skills. The U.S. has high returns, exceeding those found in almost all of the developed countries that are observed (see Fig. 7). These returns imply that an individual in the U.S. who has skills as defined and measured on international comparative assessments that are one standard deviation above the mean will on average see 28% higher earnings across the lifetime compared to the median person. But these high returns also imply that somebody one standard deviation below the mean can expect 28% lower earnings across a lifetime. In other words, the U.S. provides high rewards to acquired skills as measured by standardized tests, but it also severely punishes those with low skills. These estimates are consistent with research about the growing importance of basic cognitive skills from a quarter of a century ago (Murnane et al. 1995).

In sum, a wide range of evidence shows the substantial economic value of improved cognitive skills. This in turn suggests that student test scores merit policy attention. Yet this does not consistently show up in actions.

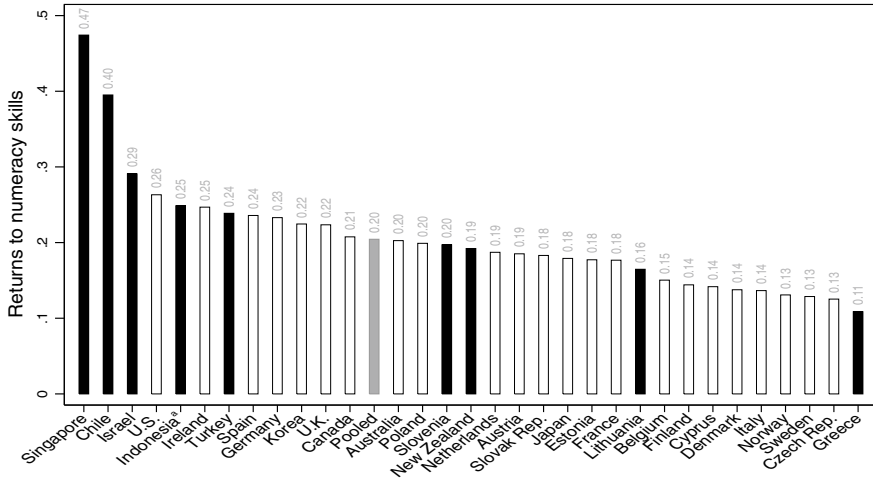


Fig. 7 Estimated Return to Numeracy by Country. Note: Estimates from the Program of International Assessment of Adult Competencies (PIAAC) of the returns to skills across PIAAC countries. Coefficient estimates on numeracy score (standardized to std. dev. 1 within each country) in a regression of log gross hourly wage on numeracy, gender, and a quadratic polynomial in age, sample of full-time employees aged 35–54. Regressions weighted by sampling weights. Hollow bars indicate first-round countries, Black bars indicate second-round countries. *Jakarta only. *Source* Hanushek et al. (2017)

7 Why Has the U.S. Done so Well?

One might ask ‘how has the U.S. done so well over the past century when achievement levels are so low?’ As seen by the growth chart (Fig. 6), the U.S. has done better than would be expected by its test scores.

Perhaps the most important factor is the favorable economic institutions that support productive use of resources and growth. The United States has generally less governmental intrusion into the operation of economic markets including lower tax rates and less regulation of labor and capital markets. There are strong property rights, and there is quite free movement of labor and capital within the U.S. All of these institutional factors are thought to promote more efficiency and growth.

Of course, these favorable growth institutions may have other implications such as a wider distribution of income or less certain provision of health care. But these are trade-offs made with the implication that growth is stronger than in other countries that choose different kinds of economic and political structures.

Additionally, at least historically the U.S. has had a larger quantity of schooling than other countries in the world, allowing it to substitute quantity for quality. This trade-off includes moving toward high levels of compulsory schooling before most other nations.

Moreover, by most evaluations, the U.S. has higher quality colleges and universities than are found elsewhere. This university quality has supported an active research and development system and has led to a high level of innovation.

Finally, in terms of factors supporting U.S. success, the United States has been able to attract a highly skilled group of immigrants, thus borrowing from the educational systems elsewhere. For example, of all of the Ph.D's in STEM fields in the U.S., over half are foreign born (Hanson and Slaughter 2019; Hanson et al. 2018).

8 Will Good Fortune Last?

The full story developed here is rather straightforward.

First, the U.S. has not done well as measured by international tests. PISA results for 2018 are just the most recent evidence of the mediocre performance of U.S. schools. The overall U.S. performance is around or below the average for the OECD. And, there is no evidence that equity in terms of educational achievement is improving.

Second, this long stasis is not the result of a constant, unchanging schooling system. While decision making in the U.S. is complicated because the 50 states are primary in schooling issues, there have been substantial changes aimed at improving the schools. Funding has increased dramatically. There has been clear school accountability. Parents have more options to choose schools that meet their demands. Many programs and policies are aimed at improving equity in the outcomes of schools including compensatory funding from the federal government, expansion of preschool access and usage, considerable desegregation of schools over the past half century, targeted funding for special education, and added state funding for disadvantaged students. For whatever reasons, these policies have not led to improved school outcomes in the United States.

There is at the same time considerable complacency. After all, with the current schools, the U.S. remains a rich nation with growth that exceeds that in much of the developed world. Isn't it possible simply to continue and to expect good fortune?

Much depends on whether the offsetting forces described above remain effective. Unfortunately, that might not be the case—making it important for the U.S. to depend more fully on its own knowledge capital. The potential for a negative change in fortune appears large enough that the U.S. should work harder at finding ways to improve its schools.

References

- Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94(5), 1613–1634.
- Baude, P. L., Casey, M., Hanushek, E. A., Phelan, G. R., & Rivkin, S. G. (2020). The evolution of charter school quality. *Economica*, 87(345), 158–189.

- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.
- CREDO. (2013). *National charter school study 2013*. Stanford, CA: Center for Research on Education Outcomes, Stanford University.
- Currie, J., & Thomas, D. (2000). School quality and the longer-term effects of head start. *Journal of Human Resources*, 35(4), 755–774.
- Hanson, G. H., Kerr, W. R., & Turner, S. (Eds.). (2018). *High-skilled migration to the United States and its economic consequences*. Chicago: University of Chicago Press.
- Hanson, G. H., & Slaughter, M. J. (2019). High-skilled immigration and the rise of stem occupations in US employment. In C.R. Hulten, & V.A. Ramey (Eds.), *Education, skills, and technical change: Implications for future US GDP Growth* (pp. 465–494). Chicago: University of Chicago Press.
- Hanushek, E. A. (2002). Publicly provided education. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of public economics* (Vol. 4, pp. 2045–2141). Amsterdam: North Holland.
- Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(485), F64–F98.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2009). New Evidence about *Brown V. Board of Education*: The complex effects of school racial composition on achievement. *Journal of Labor Economics* 27(3), 349–383.
- Hanushek, E. A., & Lindseth, A. A. (2009). *Schoolhouses, courthouses, and statehouses: Solving the funding-achievement puzzle in America's public schools*. Princeton, NJ: Princeton University Press.
- Hanushek, E. A., Peterson, P. E., Talpey, L. M., & Woessmann, L. (2020, February). Long-run trends in the U.S. Ses-achievement Gap. NBER Working Paper No. 26764. Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2012). Is the United States catching up? International and state trends in student achievement. *Education Next*, 12(4), 24–32.
- Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2013). *Endangering prosperity: A global view of the American school*. Washington, DC: Brookings Institution Press.
- Hanushek, E. A., Ruhose, J., & Woessmann, L. (2016). It pays to improve school quality: States that boost student achievement could reap large economic gains. *Education Next*, 16(3), 16–24.
- Hanushek, E. A., Ruhose, J., & Woessmann, L. (2017a). Economic gains from educational reform by US states. *Journal of Human Capital* 11(4), 447–486.
- Hanushek, E. A., Ruhose, J., Woessmann, L. (2017b). Knowledge capital and aggregate income differences: Development accounting for U.S. states. *American Economic Journal: Macroeconomics* 9(4), 184–224.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the World: Evidence from PIAAC. *European Economic Review*, 73, 103–130.
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2017). Coping with change: International differences in the returns to skills. *Economic Letters* 153, 15–19.
- Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E.A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 89–200), Amsterdam: North Holland.
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267–321.
- Hanushek, E. A., & Woessmann, L. (2015a). *The knowledge capital of nations: Education and the economics of growth*. Cambridge, MA: MIT Press.
- Hanushek, E. A., & Woessmann, L. (2015b). *Universal basic skills: What countries stand to gain*. Paris: Organisation for Economic Co-operation and Development.
- Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The Effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *Quarterly Journal of Economics*, 131(1), 157–218.

- Johnson, R. C., & Kirabo Jackson, C. (2017, June). *Reducing inequality through dynamic complementarity: Evidence from head start and public school spending*. NBER Working Paper No. 23489. Cambridge, MA: National Bureau of Economic Research.
- Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School finance reform and the distribution of student achievement. *American Economic Journal: Applied Economics*, 10(2), 1–26.
- Mincer, J. (1970). The distribution of labor incomes: A survey with special reference to the human capital approach. *Journal of Economic Literature*, 8(1), 1–26.
- Mincer, J. (1974). *Schooling, experience, and earnings*. New York: NBER.
- Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics*, 77(2), 251–266.
- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head start impact study: Final report*. Washington, DC: Administration for Children and Families (January).
- Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the Head Start impact study final report*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Rivkin, S. G. (2016). Desegregation since the Coleman Report: Racial composition of schools and student learning. *Education Next*, 16(2), 29–37.
- U.S. Department of Education. (2018). *Digest of education statistics, 2017*. Washington, DC: National Center for Education Statistics.
- U.S. Department of Education. (2019). *Digest of education statistics 2018*. Washington, DC: National Center for Education Statistics.
- Vinovskis, M. A. (1999). Do federal compensatory education programs really work? A brief historical analysis of Title I and Head Start. *American Journal of Education*, 107(3), 187–209.

Eric Hanushek is the Paul and Jean Hanna Senior Fellow at the Hoover Institution of Stanford University. He is a recognized leader in the economic analysis of education issues, and his research has had broad influence on education policy in both developed and developing countries. He is the author of numerous widely-cited studies on the effects of class size reduction, school accountability, teacher effectiveness, and other topics. He was the first to research teacher effectiveness by measuring students' learning gains, which formed the conceptual basis for using value-added measures to evaluate teachers and schools, now a widely adopted practice. His recent book, *The Knowledge Capital of Nations: Education and the Economics of Growth* summarizes his research establishing the close links between countries' long-term rates of economic growth and the skill levels of their populations. His current research analyzes why some countries' school systems consistently perform better than others. He has authored or edited twenty-four books along with over 250 articles. He is a Distinguished Graduate of the United States Air Force Academy and completed his Ph.D. in economics at the Massachusetts Institute of Technology.

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.



Assessment Background: What PISA Measures and How



Luisa Araújo, Patrícia Costa, and Nuno Crato

Abstract This chapter provides a short description of what the Programme for International Student Assessment (PISA) measures and how it measures it. First, it details the concepts associated with the measurement of student performance and the concepts associated with capturing student and school characteristics and explains how they compare with some other International Large-Scale Assessments (ILSA). Second, it provides information on the assessment of reading, the main domain in PISA 2018. Third, it provides information on the technical aspects of the measurements in PISA. Lastly, it offers specific examples of PISA 2018 cognitive items, corresponding domains (mathematics, science, and reading), and related performance levels.

1 Introduction

PISA seeks to capture a common dimension of cognitive skills across countries. These skills are thought to be a good indication of the knowledge and skills that are essential for full participation in contemporary societies (OECD 2019a), and the attained level of these cognitive skills is viewed as an important determinant of economic growth (Heckman and Jacobs 2009). More specifically, PISA reinforces the idea that "...direct measures of cognitive skills offer a superior approach to understanding how

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

L. Araújo (✉)
Instituto Superior de Educação e Ciências, ISEC, Lisbon, Portugal
e-mail: luisa.araujo@iseclisboa.pt

P. Costa
European Commission Joint Research Centre, Ispra, Italy
e-mail: patricia.costa@lese-conference.org

N. Crato
Cemapre/REM, ISEG, University of Lisbon, Lisbon, Portugal
e-mail: ncrato@iseg.ulisboa.pt

human capital affects the economic fortunes of nations”, as expressed by Hanushek and Woessmann (2015, p.28). That is, as it is nowadays widely recognized, the quality of one’s education is a better indicator of life outcomes than the quantity of education, as measured in years of schooling or similar indicators (Heckman and Jacobs 2009).

PISA results are complemented by other ILSA studies, and it is reassuring that high correlations across studies have been found. In particular, consider the Third International Mathematics and Science Study (TIMSS), a curriculum-sensitive ILSA conducted by the International Association for the Evaluation of Educational Achievement (IEA). PISA and TIMSS assess similar mathematics and science knowledge and skills at approximately the same time during schooling and a comparison between the two reveals that “... the correlation between the TIMSS 2003 tests of 8th graders and the PISA 2003 tests of 15-year-olds across the 19 countries participating in both is as high as 0.87 in mathematics and 0.97 in science. It is also 0.86 in both mathematics and science across the 21 countries participating both in the TIMSS 1999 tests and the PISA 2000–02 tests” (OECD 2010, p. 38).

A corresponding comparison of PISA with IEA’s Program for International Reading Literacy Study (PIRLS) is not possible since this ILSA is designed to assess the reading skills of 4th graders, when most students are between 9 and 10 years of age. Still, a close look at both the PIRLS 2016 and the PISA 2018 assessment frameworks shows a very similar definition of reading. In PIRLS 2016 “Reading literacy is the ability to understand and use those written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment (Mullis et al. 2015, p.12). In PISA 2018, “reading literacy is understanding, using, evaluating, reflecting on and engaging with texts in order to achieve one’s goals, to develop one’s knowledge and potential and to participate in society” (OECD 2019c, p.28).

PISA, as the other ILSA such as PIRLS and TIMSS, also collects contextual information on students’ socio-demographic and dispositional characteristics, students’ home environment and teaching and schools’ learning contexts (Lenkeit et al. 2015). This is done through the application of several questionnaires.

PISA results attract public attention mainly because of the country rankings they present in a comparative perspective and of the results’ policy implications suggested by the OECD (Araújo et al. 2017). Educational implications can be drawn from statistical associations between cognitive performance and the information collected in the various questionnaires. In PISA 2018, such associations between cognitive performance and learning variables are discussed at length through several OECD volumes; main findings appear in the *Combined Executive Summaries* (OECD 2019b). For example, two findings with clear educational implications are: (1) students who perceived greater support from teachers scored higher in reading and (2) students whose parents discuss their progress on the initiative of the teacher had higher achievement in reading.

2 How Cognitive Skills Are Measured

All the ILSA here discussed use multistage sampling, unequal sampling probabilities, and stratification, but there are some differences.

PISA adopts a two-stage stratified sample design in which the primary sampling unit consists of at least 150 schools having 15-year-old students. Schools are sampled systematically from the school sampling frame, with probabilities proportional to a measure of the school size, which is a function of the estimated number of PISA-eligible 15-year-old students enrolled in the school. The second sampling unit includes *students* (around 5000 students) within the sampled schools.

TIMSS and PIRLS also employ a two-stage random sample design. In the first stage a sample of schools is drawn, but in the second stage one or more *complete classes of students* are selected from each of the sampled schools.

In PISA, TIMSS, and PIRLS, students' test scores are computed according to Item Response Theory (IRT) and standardised with a mean of around 500 and standard deviation of around 100. Even though the methodology is quite similar, the scores in these three ILSA are not directly comparable.

From the students' score points, proficiency levels are identified based on the PISA main domain scales. In this sense, PISA results can also be reported in terms of percentages of the student population at each of the predefined level. To define the proficiency levels and their cut off scores, IRT techniques are used to estimate simultaneously the difficulty and the ability of all students participating in PISA. Higher proficiency levels characterize the knowledge, skills, and capabilities needed to perform tasks of increasing complexity.

In PISA, TIMSS, and PIRLS, each student completes one booklet containing a subset of all the material. The booklets are created by combining different blocks of items in order to match to the framework characteristics. For the cognitive assessment of PISA 2018, the total testing time was 2 h and for TIMSS 2015 (8th grade), 1.5 h. PISA reading questions include a variety of items, including the conventional multiple-choice format and a complex multiple-choice format. TIMSS cognitive assessments primarily use multiple choice and constructed response items.

In all these surveys, national estimates are generated from the sample with different weights. To increase accuracy, these ILSA use plausible values (multiple imputations) drawn from a posteriori distribution which is constructed by combining the IRT scaling of the test items with a latent regression model with information from the student context questionnaire within a population model. For each student, 10 plausible values are computed in PISA (since 2015) and 5 plausible values are computed in all cycles of TIMSS and PIRLS.

All these ILSA studies allow for cross-country comparisons and for trend monitoring over time. In order to guarantee the comparability across countries, along years and delivery modes (paper and computer), linking procedures are used by considering a large number of common items in which the parameters are fixed to the same values. These items serve as *anchors* of the reporting scales and support the validity of cross-country and trend comparisons (OECD 2019c).

3 The Measurement of Student Performance in PISA

In PISA 2018, reading was the major domain of assessment, as it was in 2000 and 2009. The texts and items were selected based on a conceptual framework (OECD 2019a), which included five subscales. Three of the PISA 2018 assessment subscales have already been used in 2000 and 2009: “locating information”, “understanding” and “evaluating and reflecting”, (OECD, 2009). Two assessment subscales were newly created to describe students’ literacy with single-source and with multiple source texts. Additionally, PISA 2018 included for the first time a measure of reading fluency in order to assess the reading skills of students in the lower proficiency levels. Reading fluency is defined as “the ease and efficiency with which one can read and understand a piece of text” (OECD 2019c, p. 270).

This was an important addition. As recognized in the PISA assessment framework, research shows that many students have difficulties with reading comprehension because they have not developed effortless decoding or the automaticity in word recognition that enables readers to focus on comprehension processes (OECD 2019a). Numerous research studies on reading processes have confirmed this (Adams 1990, 2009; Perfetti et al. 2005). Although comprehension can be developed throughout schooling and reading comprehension skills can be improved (Catts 2009; Elbro and Buch-Iverson 2013), it is fundamental that students acquire the basic reading skills that will allow them to read fluently, which implies reading words and text fast and accurately (Perfetti et al. 2005).

In order to simplify the interpretation of results, PISA scale is categorized into six ordinal proficiency levels. Each proficiency level requires a certain set of competencies, knowledge, and understanding items to be successfully completed. 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. Table 1 reproduces the score limits for reading for PISA 2018.

Table 1 PISA 2018 reading scores levels of proficiency

Level 6	Above 698.32 score points
Level 5	From 625.61 to less than 698.32 score points
Level 4	From 552.89 to less than 625.61 score points
Level 3	From 480.18 to less than 552.89 score points
Level 2	From 407.47 to less than 480.18 score points
Level 1a	From 334.75 to less than 407.47 score points
Level 1b	From 262.04 to less than 334.75 score points
Level 1c	From 189.33 to less than 262.04 score points
Below level 1c	Less than 189.33 score points

Students scoring below level 2 are considered low-performers
 Students scoring above level 4 are considered high-performers
Source OECD, PISA 2018 Database, Table I.B1.4; Figure I.4.1

Students scoring below level 2 are considered *low-performers* and those scoring above level 4 are considered *high-performers*. In 2015, recognizing the worrisome number of low performers and the need to better discriminate those students, PISA has subdivided level 1 in 1a and 1b. In 2018, PISA introduced an additional lower level, 1c.

Reading comprehension in PISA is assessed by asking students to locate information in a text, to retrieve literal information, to generate inferences and to evaluate and reflect on the content and form of texts. Evaluating a text is a more complex skill than simply identifying the requested information, and the six difficulty levels that PISA establishes are related to the tasks students need to perform. Locating explicit information in a text is a very basic reading task typical of level 1, whereas reflecting on the content of a text is a complex skill that characterizes questions at level 6. The difficulty level of the test items correspond to what the OECD refers to as *aspect* and reflect the cognitive processes involved in the task: “the *access and retrieve* aspect assessing the lowest benchmark proficiency levels (1 & 2), followed by the *Integrate and interpret* level (3 & 4) and with the *Reflect and evaluate* levels at the highest text processing level (5 & 6)” (OECD 2019a).

Level 2 marks the point at which students have acquired the basic skills to read and can use reading for learning. “At a minimum, these students [scoring at least level 2] are able to identify the main idea in a text of moderate length, find information based on explicit criteria, and reflect on the purpose and form of texts when explicitly directed to do so.” Low performers are not able to attain this basic level.

Students who attained the highest proficiency levels 5 or 6 in reading, “are able to comprehend lengthy texts, deal with concepts that are abstract or counterintuitive, and establish distinctions between fact and opinion, based on implicit cues pertaining to the content or source of the information”. (OECD 2019c).

The test items used to assess these text processing abilities are a mixture of multiple-choice questions and questions requiring students to construct their own responses. Such question and formats appear for a wide range of texts types; narrative, expository, descriptive and argumentative texts. Text types are presented as both continuous texts, organized in paragraphs and non-continuous, matrix-like formats, or with the appearance of a list. Since the purpose of assessing reading performance in PISA is to obtain a measure of reading comprehension, even the questions that require the students to construct a written response do not ask for extensive responses (OECD 2019a).

4 Questionnaire Data

PISA includes compulsory questionnaires and optional questionnaires. Compulsory questionnaires are the student background questionnaire (distributed to all participating students) and the school questionnaire (distributed to the principals of all participating schools). The student questionnaire, which takes about 35 minutes to complete, includes socio-demographic information about the students, such as

age, gender, type of educational program the student is completing, immigrant background and parental occupation, a proxy for socio-economic status <https://www.oecd.org/pisa/pisaproducts/PISA-2018-INTEGRATED-DESIGN.pdf>. The school questionnaire that principals complete covers school learning experiences, school management, assessment, and school climate. For example, student truancy and bullying, cooperation among teachers and among students, and teacher enthusiasm and encouragement of reading are measures of school climate, a construct that includes social and academic dimensions believed to predict academic achievement and social skills (Costa and Araújo 2018; Chirkina and Khavenson 2018).

In 2018, the optional PISA questionnaires included three questionnaires for students (the educational career questionnaire, the ICT familiarity questionnaire, and the well-being questionnaire); one questionnaire for parents; one questionnaire for teachers (both for reading teachers and for all other subjects teachers); and one financial literacy questionnaire for students in countries that participated in the financial literacy assessment.

PIRLS and TIMSS usually include the following questionnaires: student, home (for 4th grade students and distributed to the parents of the students participating in the survey), teachers, schools, and curricular background data.

Teacher questionnaires in PISA are answered by the teachers of the sampled schools, while the PIRLS and TIMSS questionnaires are answered by the teachers of the assessed classes.

5 Examples of Cognitive Items in PISA 2018 and Other ILSA—What Questions Look Like

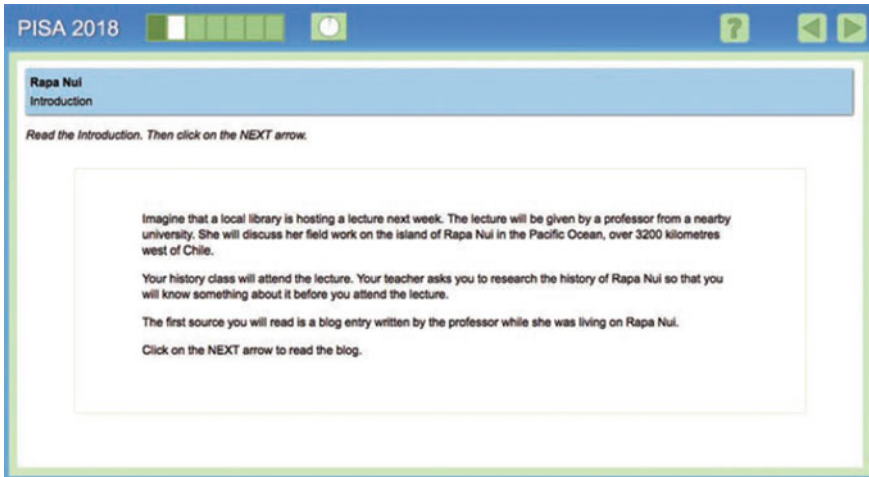
In the next pages we show examples of PISA reading items, followed by examples of some science and mathematics items, both from PISA and from TIMSS. Firstly, we will focus on the Rapa Nui Unit,¹ which is a scenario-based example. In this kind of unit, the student is given both a context and a purpose that helps to shape the way he/she searches for, comprehends, and integrates information. Rapa Nui refers to an island; the student is preparing to attend a lecture about a professor's field work, which was conducted on this island. This unit begins with a fictional scenario and is a multiple-source unit. It consists of three texts: a webpage from the professor's blog, a book review, and a news article from an online science magazine. The blog post is multiple-source text given that the comments section represents different authors. Both the book review and the news article are classified as single text, static, continuous, and argumentative. The Rapa Nui scenario prompts the student to integrate information in questions that are related to one text and then to demonstrate the ability to handle information from multiple texts. This design allows students

¹Example of a PISA 2018 reading scenario. "Released items from the PISA 2018 computer-based reading assessment", in OECD (2019c).

with varying levels of ability to demonstrate proficiency on at least some questions of the unit. Specifically, this unit is intended to be of moderate to high difficulty.

5.1 Example 1: Rapa Nui—Scenario

1. Introduction



Item #1 is a single source item and the student must find the correct information within the blog post. The cognitive process required to engage in this task is that of assessing and retrieving information within a piece of text and its difficulty level is 4.

Item #2 is an open response (human coded) item² where the student must understand the second mystery mentioned in the Blog Post. It involves the cognitive process of representing literal meaning and its difficulty level is 3.

Item #6 asks students to integrate information across the texts with respect to the differing theories put forward by several scientists. This item involves integrating and generating inferences across multiple sources and is a complex multiple-choice item with a complexity level of 5.

²More information and the coding guide used can be found at “Released items from the PISA 2018 computer-based reading assessment”, in OECD (2019c).

2. Released Item #1. The Professor's Blog - (Item number CR551Q01)

PISA 2018

Rapa Nui
Question 1 / 7

Refer to the Professor's Blog on the right. Click on a choice to answer the question.

According to the blog, when did the professor start her field work?

- During the 1990s.
- Nine months ago.
- One year ago.
- At the beginning of May.

Blog
www.theprofessorblog.com/fieldwork/RapaNui

The Professor's Blog

Posted May 23, 11:22 a.m.

As I look out of my window this morning, I see the landscape I have learned to love here on Rapa Nui, which is known in some places by the name Easter Island. The grasses and shrubs are green, the sky is blue, and the old, now extinct volcanoes rise up in the background.

I am a bit sad knowing that this is my last week on the island. I have finished my field work and will be returning home. Later today, I will take a walk through the hills and say good-bye to the moai that I have been studying for the past nine months. Here is a picture of some of these massive statues.



If you have been following my blog this year, then you know that the people of Rapa Nui carved these moai hundreds of years ago. These impressive moai had been carved in a single quarry on the eastern part of the island. Some of them weighed thousands of kilos, yet the people of Rapa Nui were able to move them to locations far away from the quarry without cranes or any heavy equipment.

For years, archeologists did not know how these massive statues were moved. It remained a mystery until the 1990s, when a team of archeologists and residents of Rapa Nui demonstrated that the moai could have been transported and raised using ropes made from plants and wooden rollers and tracks made from large trees that had once thrived on the island. The mystery of the moai was solved.

Another mystery remained, however. What happened to those plants and large trees that had been used to move the moai? As I said, when I look out of my window, I see grasses and shrubs and a small tree or two, but nothing that could have been used to move these huge statues. It is a fascinating puzzle, one that I will explore in future posts and lectures. Until then, you may wish to investigate the mystery yourself. I suggest you begin with a book called *Cokapea* by Jared Diamond. [This review of Cokapea is a good place to start.](#)

Traveler_14 May 24, 4:31 p.m.
Hi Professor! I love following your work on Easter Island. I can't wait to check out *Cokapea*!

KB_Island May 25, 9:07 a.m.
I also love reading about your experiences on Easter Island, however, I think there is another theory that should be considered. Check out this article: www.sciencenews.com/Polyesian_rabs_Rapa_Nui

3. Released Item #2. The Professor's Blog (Item number CR551Q05)

PISA 2018

Rapa Nui
Question 2 / 7

Refer to the Professor's Blog on the right. Type your answer to the question.

In the last paragraph of the blog, the professor writes: "Another mystery remained..."

To what mystery does she refer?

Blog


www.theprofessorblog.com/fieldwork/RapaNui

The Professor's Blog

Posted May 23, 11:22 a.m.

As I look out of my window this morning, I see the landscape I have learned to love here on Rapa Nui, which is known in some places by the name Easter Island. The grasses and shrubs are green, the sky is blue, and the old, now extinct volcanoes rise up in the background.

I am a bit sad knowing that this is my last week on the island. I have finished my field work and will be returning home. Later today, I will take a walk through the hills and say good-bye to the moai that I have been studying for the past nine months. Here is a picture of some of these massive statues.



4. Released item # 6. Science News (Item number CR551Q10)

PISA 2018

Rapa Nui
Question 6 / 7

Refer to all three sources on the right by clicking on each of the tabs.

Drag and drop the causes, and the effect they have in common, into the correct places in the table about the theories.

The Theories

Cause	Effect	Supporters of the Theory
		Jared Diamond
		Carl Lipo and Terry Hunt
The moai were carved in the same quarry.	Polynesian rats ate tree seeds and as a result no new trees could grow.	Settlers used coconuts to bring Polynesian rats to Rapa Nui.
The large trees disappeared from Rapa Nui.	Rapa Nui residents needed natural resources to move the moai.	Humans cut down trees to clear land for agriculture and other reasons.

Blog Book Review Science News

www.sciencenews.com/Polynesian_rats_Rapa_Nui

SCIENCE NEWS

Did Polynesian Rats Destroy Rapa Nui's Trees?
By Michael Kimball, Science Reporter

In 2005, Jared Diamond published *Collapse*. In the book, he described the human settlement of Rapa Nui (also called Easter Island).

The book caused a huge controversy soon after its publication. Many scientists questioned Diamond's theory of what happened on Rapa Nui. They agreed that the huge trees had disappeared by the time Europeans first arrived on the island in the 18th century, but they did not agree with Jared Diamond's theory about the cause of the disappearance.

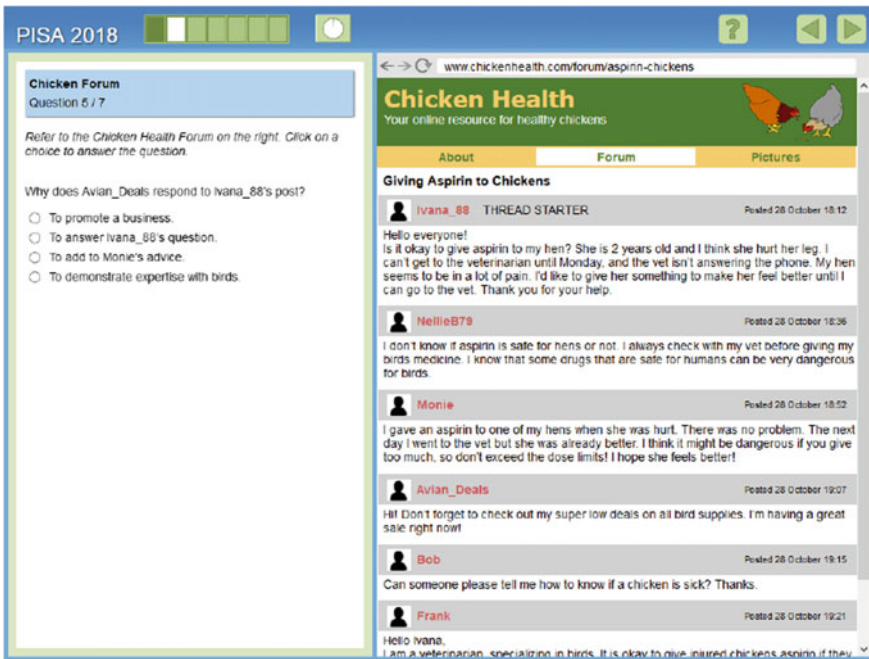
Now, two scientists, Carl Lipo and Terry Hunt, have published a new theory. They believe that the Polynesian rat ate the seeds of the trees, preventing new ones from growing. The rat, they believe, was brought over either accidentally or purposefully on the canoes that the first human settlers used to land on Rapa Nui.

Studies have shown that a population of rats can double every 47 days. That's a lot of rats to feed. To support their theory, Lipo and Hunt point to the remains of palm nuts that show the gnaw marks made by rats. Of course, they acknowledge that humans did play a role in the destruction of the forests of Rapa Nui. But they believe that the Polynesian rat was an even greater culprit among a series of factors.

Next, we present an example of a reading proficiency level 1 task in PISA 2018. The item is part of the Chicken Forum Scenario³ and describes a person who is seeking information about how to help an injured chicken. In this particular item it is expected that the student makes an inference from the information provided in a post. The item is classified as a single multiple choice one and it involves integrating and generating inferences as a cognitive process.

5.2 Example 2: Chicken Forum (Item Number CR548Q05)⁴

1. Released Item #5



Example 3 presents Science items from PISA and from TIMSS (8th grade). The PISA item is a multiple choice item classified as level 4 and it is an item “that requires students to be able to relate the rotation of the earth on its axis to the phenomenon of day and night and to distinguish this from the phenomenon of the seasons, which

³The units Chicken Forum was administered in the PISA 2018 Field Trial but was not selected for the Main Survey.

⁴More information can be found <https://www.oecd.org/pisa/test/> and in the document (OECD, 2019f).

arises from the tilt of the axis of the earth as it revolves around the sun. All four alternatives given are scientifically correct” (OECD 2004, p. 289).

5.3 Example 3: Science Items—PISA and TIMSS

1. PISA 2003 item: DAYLIGHT⁵

Read the following information and answer the questions that follow.

DAYLIGHT ON 22 JUNE 2002

Today, as the Northern Hemisphere celebrates its longest day, Australians will experience their shortest.	rise at 5:55 am and set at 8:42 pm, giving 14 hours and 47 minutes of daylight.
In Melbourne*, Australia, the Sun will rise at 7:36 am and set at 5:08 pm, giving nine hours and 32 minutes of daylight.	The President of the Astronomical Society, Mr Perry Vlahos, said the existence of changing seasons in the Northern and Southern Hemispheres was linked to the Earth’s 23-degree tilt.
Compare today to the year’s longest day in the Southern Hemisphere, expected on 22 December, when the Sun will	

*Melbourne is a city in Australia at a latitude of about 38 degrees South of the equator.

Question 1: DAYLIGHT S129Q01

Which statement explains why daylight and darkness occur on Earth?

- A The Earth rotates on its axis.
- B The Sun rotates on its axis.
- C The Earth’s axis is tilted.
- D The Earth revolves around the Sun.

2. TIMSS 2011 item: Recognizes the major cause of tides⁶

⁵We cannot help noticing the scientifically incorrect statement of the third paragraph: There is no such thing as the longest day in the Southern Hemisphere with the sun rising and setting at specific times; the length of the day and the specific times depend on the latitude.

⁶SOURCE: TIMSS 2011 Assessment. Copyright © 2013 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands.

Which of the following is the major cause of tides?


- A. heating of the oceans by the Sun
- B. gravitational pull of the Moon
- C. earthquakes on the ocean floor
- D. changes in wind direction

Example 4 shows Mathematics items from PISA and from TIMSS (8th grade). Both items are open-ended items.

5.4 Example 4: Mathematics—PISA and TIMSS

1. PISA 2012 item: DRIP RATE⁷

Infusions (or intravenous drips) are used to deliver fluids and drugs to patients.



Nurses need to calculate the drip rate, D , in drops per minute for infusions.

They use the formula $D = \frac{dv}{60n}$ where

- d is the drop factor measured in drops per millilitre (mL)
- v is the volume in mL of the infusion
- n is the number of hours the infusion is required to run.

Question 1: DRIP RATE PM903Q01 - 0 1 2 9

A nurse wants to double the time an infusion runs for.

Describe precisely how D changes if n is **doubled** but d and v do not change.

.....

.....

.....

⁷More information can be found at <https://www.oecd.org/pisa/test/> - PISA 2012, Mathematics items.

2. TIMSS 2011 item: Ann and Jenny divide 560 zeds⁸

<p>Ann and Jenny divide 560 zeds between them. If Jenny gets $\frac{3}{8}$ of the money, how many zeds will Ann get?</p> <p>Answer: _____</p>
--

6 Conclusion

This chapter offers a short description of what PISA measures and how it measures it. As such, it provides basic information about PISA’s assessment framework and technical specifications related to sampling and statistical procedures and analyses. For more detailed information, readers can access OECD documents, namely the PISA assessment framework reports and the technical reports published by OECD for every assessment cycle. The PISA questionnaires can be accessed through the OECD/PISA database webpage (<https://www.oecd.org/pisa/data/2018database/>). More examples of released items can be found in https://www.oecd.org/pisa/test/PISA2018_Released_REA_Items_12112019.pdf. In order to have a good insight about PISA student results it is important to get acquainted with a few testing items. We hope this concluding assessment background chapter provides information to better understand PISA analyses.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Adams, M. J. (2009). The challenge of advanced texts: The interdependence of reading and learning. In E. Hiebert (Ed.), *Reading more, reading better* (pp. 163–189). New York, NY: Guilford Press.
- Araújo, L., Saltelli, A., & Schnepf, S.V. (2017). Do PISA data justify PISA-based education policy? *International Journal of Comparative Education and Development*, 19(1), 20–34. <https://doi.org/10.1108/IJCED-12-2016-0023>.
- Catts, H. W. (2009). The narrow view of reading promotes a broad view of comprehension. *Language, Speech, and Hearing Services in the Schools*, 40, 178–183.

⁸SOURCE: TIMSS 2011 Assessment. Copyright © 2013 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands.

- Chirkina, T. A., & Khavensho, T. E. (2018). School climate: A history of the concept and approaches to defining and measuring it on PISA questionnaires. *Russian Education & Society*, 60(2), 133–160. <https://doi.org/10.1080/10609393.2018.1451189>.
- Costa, P., & Araújo, L. (2018). Skilled students and effective schools: Reading achievement in Denmark, Sweden, and France. *Scandinavian Journal of Educational Research*, 62, 850–864. <https://doi.org/10.1080/00313831.2017.1307274>.
- Elbro, C., & Buch-Iverson, I. (2013). Activation of prior knowledge for inferences making: Effects on reading comprehension. *Scientific Studies of Reading* 17, 435–452.
- Hanushek, E., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. Massachusetts: Massachusetts Institute of Technology. MIT Press.
- Heckman, J., & Jacobs, B. (2009). *Policies to create and destroy human capital in Europe*. IZA DP No. 4680.
- Lenkeit, J., Chan, J., Hopfenbeck, T. N., & Baird, J. (2015). A review of the representation of PIRLS related research in scientific journals. *Educational Research Review*, 16, 102–115. <https://doi.org/10.1016/j.edurev.2015.10.002> [Crossref], [Web of Science ®], [Google Scholar].
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *PIRLS 2016 assessment framework* (2nd ed.). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/pirls2016/framework.html>.
- OECD. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- OECD. (2009). *PISA 2009 assessment framework. Key competencies in reading, mathematics and science*. <https://www.oecd.org/dataoecd/11/40/44455820.pdf>.
- OECD. (2010). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes, PISA*. OECD Publishing. <https://doi.org/10.1787/9789264077485-en>.
- OECD. (2019a). *PISA 2018 assessment and analytical framework*. Paris: PISA, OECD Publishing. <https://doi.org/10.1787/b25efab8-en>.
- OECD. (2019b). *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. (2019c). *PISA 2018 Results (Volume I): What students know and can do, PISA*. Paris: OECD Publishing. 10.1787/5f07c754-en.
- OECD. (2019d). Released items from the PISA 2018 computer-based reading assessment. In *PISA 2018 results (Volume I): what students know and can do*. Paris: OECD Publishing. <https://doi.org/10.1787/098bab1a-en>.
- OECD. (2019e). *PISA 2018 technical report*. <https://www.oecd.org/pisa/data/pisa2018technicalreport/>. Accessed June 10, 2020.
- OECD. (2019f). *PISA 2018 released field trial and main survey new reading items. Version: October 2019*. https://www.oecd.org/pisa/test/PISA2018_Released_REA_Items_12112019.pdf. Accessed June 10, 2020.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skills. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–274). Oxford, UK: Blackwell.

Lúisa Araújo holds a Ph.D. in Curriculum and Instruction from the University of Delaware with a concentration in literacy studies and bilingual education. Her research interests include literacy and early reading development and second/foreign language learning. Her teaching activities include teaching courses on educational evaluation, research methods, and reading development. She has served as a director of several bachelor and master degrees in education and worked as a researcher for the European Commission in the area of Education and Training at the Joint Research Center. She is currently full professor in the School of Education in the Institute of Higher Education and Science (ISEC) in Lisbon, Portugal.

Patrícia Costa has worked at the Joint Research Centre of the European Commission since June 2011. She holds a PhD in the area of Pshycometrics/Statistics and received a master and bachelor's degrees in mathematics. Patrícia has been working as a statistician and psychometrician and contributing to research and policy support in the areas of Educational Assessment and Comparative Education Statistics. Her recent research is focused in secondary data analysis of international large-scale surveys and validation of instruments.

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.

