



CROSS-NATIONAL ACHIEVEMENT SURVEYS FOR MONITORING EDUCATIONAL OUTCOMES

POLICIES, PRACTICES
AND POLITICAL
REFORMS WITHIN
THE EUROPEAN UNION

Louis Volante
Sylke V. Schnepf
Don A. Klinger
(editors)

Joint
Research
Centre

This publication is a book by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information

Name: Sylke V. Schnepf

Email: Sylke.Schnepf@ec.europa.eu

EU Science Hub

<https://ec.europa.eu/jrc>

JRC121884

EUR 30380 EN

Print	ISBN 978-92-76-22353-5	ISSN 1018-5593	doi:10.2760/59628	KJ-NA-30380-EN-C
PDF	ISBN 978-92-76-22352-8	ISSN 1831-9424	doi:10.2760/406165	KJ-NA-30380-EN-N

Luxembourg: Publications Office of the European Union, 2022

© European Union, 2022



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content: © European Union, 2022, except image on the cover page:

© kras99/stock.adobe.com

How to cite this book: Volante, L., Schnepf, S. V., Klinger, D. A., eds., *Cross-National Achievement Surveys for Monitoring Educational Outcomes: Policies, Practices and Political Reforms within the European Union*, EUR 30380 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-22352-8, doi:10.2760/406165, JRC121884

CROSS-NATIONAL ACHIEVEMENT SURVEYS FOR MONITORING EDUCATIONAL OUTCOMES

POLICIES, PRACTICES
AND POLITICAL
REFORMS WITHIN
THE EUROPEAN UNION

Louis Volante
Sylke V. Schnepf
Don A. Klinger
(editors)

Dedication

This volume was initiated before the emergence of COVID-19 and finalised during what many consider the fourth wave of this pandemic. Given the countless families, communities and societies that have suffered devastating losses of loved ones, we dedicate this book to their memory, with the sincerest hope that the necessity of such a dedication will be lessened in future generations.

Foreword by Stefaan Hermans

Directorate-General for Education, Youth, Sport and Culture

Since the emergence of cross-national achievement surveys in the 1960s, governments around the world have been keen to evaluate the knowledge and skills of their student populations against a common benchmark. This has been particularly the case among the Organisation for Economic Co-operation and Development nations and the European Union Member States, which share a single market, collaborate with one another to foster European education and are working together to create the European Education Area (EEA).

The EU does not run education systems. However, its Member States share a responsibility to ensure that their national education systems help build strong, competitive, fair and sustainable societies. They are deeply committed to their common responsibility to develop education systems that enable every student to reach his or her full potential, recognising that Europe cannot afford to waste the precious talents of its youngest generation.

The COVID-19 pandemic has placed unprecedented pressure on the education and training sector, creating new challenges while also aggravating pre-existing ones. This shows once more the importance of building resilient education and training systems that are informed by robust, comparable data. Measures such as the Programme for International Student Assessment (PISA) and Programme for the International Assessment of Adult Competencies (PIAAC) surveys carried out by the Organisation for Economic Co-operation and Development have become invaluable in both taking stock of a nation's international standing in the education domain and monitoring the impact of particular policies, practices and political reforms. They guide our policymaking and will be essential to track countries' progress towards the EEA targets by 2025 and beyond.

This volume addresses these timely issues by examining a select group of EU countries (Germany, Estonia, Ireland, France, Italy, the Netherlands, Poland, Slovakia, Finland and Sweden). These countries are committed to concerted action to eradicate the lowest levels of achievement and educational inequalities within the framework of EEA 2025. In this regard, at European level, cross-national surveys help us understand the diversity among European partners in terms of educational outcomes and inequalities, and provide insights into how the EEA can be further improved. The chapters in this volume have been written by notable scholars, following a structured approach that allows the reader to make seamless comparisons across national contexts. The introductory chapter provides an overview of educational achievement surveys and their role in monitoring cross-national educational outcomes. Each subsequent country chapter offers a brief overview of a country's education system and its participation in educational achievement surveys. It also analyses how these results are used to monitor and influence education policy development and reforms in the country. The final chapter provides a concise summary of how such survey data can affect education policymaking in Europe.

Monitoring and evaluating education policy is a complex process, as the present volume illustrates. It is impacted by the structure and organisation of education within a country, historical and cultural approaches to education and political processes. A key strength of this volume is that it highlights a diversity of reactions and policy directions across Member States with respect to educational achievement surveys. As an example, many of the authors describe the current challenges involved in using cross-national educational achievement surveys as a primary tool to inform education policy development. As a result of these challenges, many countries supplement cross-national survey data with their own national data. Other countries (such as Finland) stress the potential limitations of national testing for measuring children's achievement, thereby questioning whether test scores can capture the many important outcomes and personality features that schools aim to foster. The authors also explore various manifestations of 'PISA shock', from Germany, which instigated a huge number of education reforms following unexpectedly poor PISA results, to Ireland and Sweden, which experienced similar shocks in later years, to Estonia, which has experienced a positive PISA shock, with the results highlighting the successes of its educational policies and efforts.

The EU has long supported educational change in its Member States, through Structural Funds and programmes such as Horizon and Erasmus+, and by comparing and sharing what works in different systems to improve young people's educational outcomes. International surveys have strengthened this work and have improved the available evidence on which reforms can be based. Nevertheless, as highlighted throughout this volume, current data sources have limitations that prevent deeper analyses of ongoing policies and reforms. The discussion about the role and future of the surveys suggests that some changes would be in our common interest. Going forward, policy evaluation could improve considerably if a large-scale international survey followed students over time and included achievement measures at different points in time, using the same student cohort. There is also a need to collect student and school-level data that would enable exploration of the 'whole school' approach, which aims to foster resilience, belonging and well-being. This would be especially beneficial now, as education systems focus on and address the impact of the COVID-19 pandemic.

The authors of this volume are to be congratulated for highlighting not only the diversity of education systems within the EU, but also the shared goals that need to be in place to ensure that our education systems are able to provide the best possible support to our children. They have given us much to consider, including guidance on the direction in which we need to travel to meet our EEA goals.

Stefaan Hermans
Director for Policy Strategy and Evaluation
Directorate-General for Education, Youth, Sport and Culture
European Commission
Brussels, Belgium

Foreword by Delilah Al Khudhairy

Joint Research Centre

The editors of this volume provide the reader with an interesting and timely analysis of the relationship between international testing and educational policy formulation across 10 EU Member States. The resulting national profiles and findings come at a sensitive time in our collective history, when education systems across Europe, and around the world, are grappling with the significant challenges presented by COVID-19.

This volume begins with an introductory chapter that provides an overview of key cross-national data sources, such as the Programme for International Student Assessment (PISA), which have been instrumental policy levers in the field of education. The main part of the volume, the national profiles, offers compelling historical analyses of the importance of educational policy monitoring in relation to salient national, cultural, economic and political traditions. It is clear from the national profiles and the resulting summary analysis that the challenges and complexities that policymakers face in using cross-national achievement surveys to instigate policy reform are largely due to a complex combination of factors that influence their uptake and interpretation. The volume aptly illustrates how different countries can and do possess markedly different views on what the same data source can tell a nation about its education system. Cultural understanding of what education is, how it is produced and how it can be measured affects how the monitoring of educational outcomes influences national policy design processes.

The editors – Louis Volante, Sylke V. Schnepf and Don A. Klinger – also elucidate the limitations of current international surveys for monitoring educational outcomes and evaluating education policies. They offer suggestions on how to strengthen the robustness of testing measures and gain a better understanding of the policy implications stemming from their comparative results. In particular, they argue for more thoughtful and critical analyses of existing survey results, so that their impact on policy formulation is both measured and appropriate. Clearly, understanding what works in education is not dependent on simple national or ordinal rankings – a fact made clear by the survey limitations that the editors so astutely outline.

This volume makes an invaluable contribution to the existing literature and, more importantly, illustrates how countries' cultures and political processes have a pronounced influence on the interpretative processes used by national governments. It also offers a succinct discussion of various cross-national policy monitoring approaches in an interesting cross-section of Member States that have experienced a range of achievement outcomes in international surveys.

Overall, the contributions from more than 25 authors provide the European and international communities with a thoughtful discussion of the opportunities and

constraints inherent in the intersection of large-scale surveys and policy development and monitoring processes. I hope policymakers and academics are inspired by the insights offered in this volume and consider the main findings when contemplating national education reforms. Ultimately, policy development, monitoring and evaluation should be guided by an ethos that promotes evidence-based approaches, and this volume makes an invaluable contribution to enhancing our collective capacity to achieve this goal.

Delilah Al Khudhairy
Director for Competences
Joint Research Centre
European Commission
Brussels, Belgium

Acknowledgements

This volume would not have been possible without the gracious support of the European Commission Joint Research Centre Unit I1 on Modelling, Indicators and Impact Evaluation in Ispra, Italy.

Portions of this edited volume were also supported by the Social Sciences and Humanities Research Council of Canada.

CONTENTS

PART I: INTRODUCTION

Chapter 1

Cross-national achievement surveys and educational outcomes: a brief history..... 18

Louis Volante, Sylke V. Schnepf and Don A. Klinger

PART II: NATIONAL PROFILES

Chapter 2

Cross-national achievement surveys and educational monitoring in Germany 36

Nele McElvany

Chapter 3

Cross-national achievement surveys and educational monitoring in France..... 51

Romuald Normand and Bénédicte Gendron

Chapter 4

Cross-national achievement surveys and educational monitoring in Italy 69

Orazio Giancola and Luca Salmieri

Chapter 5

Cross-national achievement surveys and educational monitoring in the Netherlands..... 91

Anneke C. Timmermans, Harm M. Naayer, Martina R. M. Meelissen, Joyce Gubbels, Jaap Scheerens and Greetje van der Werf

Chapter 6

Cross-national achievement surveys and educational monitoring in Sweden..... 115

Linda Rönnerberg and Christina Wikström

Chapter 7

Cross-national achievement surveys and educational monitoring in Finland..... 136

Heidi Katarina Harju-Luukkainen, Sari Sulkunen and Minna Maunula

Chapter 8

Cross-national achievement surveys and educational monitoring in Ireland 153

Anne Looney, Michael O'Leary, Darina Scully and Gerry Shiel

Chapter 9

Cross-national achievement surveys and educational monitoring in Poland..... 177

Maciej Jakubowski and Tomasz Gajderowicz

Chapter 10

Cross-national achievement surveys and educational monitoring in Estonia 192

Gunda Tire and Einar Värä

Chapter 11

Cross-national achievement surveys and educational monitoring in Slovakia..... 208

Katarína Zamborová and Zlatica Jursová Zacharová

PART III: CONCLUSION

Chapter 12

Cross-national achievement surveys and educational outcomes: a future path..... 230

Don A. Klinger, Sylke V. Schnepf and Louis Volante

About the editors

Louis Volante (PhD) is a professor of education at Brock University, a professorial fellow at the United Nations University's Maastricht Economic and Social Research Institute on Innovation and Technology / Maastricht Graduate School of Governance and president-elect of the Canadian Society for the Study of Education. His research is primarily focused on global education policy, performance monitoring and transnational governance, inequality in educational opportunities and outcomes, and impact evaluation of policies and programmes. Louis' scholarship is widely cited in academic and policy communities and has received generous funding from international and national agencies such as the Social Sciences and Humanities Research Council of Canada. A frequent invited speaker around the world, Professor Volante also advises governments in North America, Europe and the United Kingdom on a range of issues related to large-scale reform.

Sylke V. Schnepf (PhD) is a senior researcher in the Monitoring, Indicators and Impact Evaluation Unit of the European Commission's Joint Research Centre (Ispra, Italy) and a research fellow at the Institute of Labour Economics (Bonn, Germany). Previously, she was an associate professor in social statistics at the University of Southampton. Her main research interests regard patterns of cross-national educational inequalities and counterfactual impact evaluation of policies aimed at mitigating these inequalities. She also conducts research in the fields of survey design, fairness and loneliness.

Don A. Klinger (PhD) is the pro vice-chancellor of Te Wānanga Toi Tangata Division of Education at the University of Waikato. Prior to his current role, Professor Klinger was at Queen's University in Canada. His research explores measurement theory and the use of assessments to support teaching, learning and policy across education contexts. Professor Klinger is especially interested in efforts to identify and reduce learning inequities exposed through large-scale measures of educational outcomes. Among his accomplishments, Professor Klinger co-chaired the writing of the *Classroom Assessment Standards for PreK-12 Teachers*, and he served as president of the Canadian Society for the Study of Education.

Contributors

Tomasz Gajderowicz (PhD) is a policy advisor in the field of education and the labour market. Tomasz specialises in data science methods for uncovering incentives that drive educational and professional decisions. He developed the framework for the implementation of modern microeconomic techniques in research on preferences and assessments. Tomasz holds a PhD in economic sciences, works as a European Commission expert and a World Bank and Education and Training Evaluation Committee consultant, and holds the positions of research director at the Evidence Institute and assistant professor at the University of Warsaw. Tomasz is experienced in many international projects worldwide on educational data analysis and policy. He is an author of several publications in the field of transition from education to the labour market.

Bénédicte Gendron (PhD) is the former vice president of Paul-Valéry University Montpellier 3 and is a professor of education and human development. She earned two PhDs – one in economics and another in neuropsychology – after a postdoctoral fellowship at the University of California, Berkeley. Specialised in human development, her research is focused on *bienvivance*, quality of life and enabling leadership through emotional capital development. A member of the international Consortium for Research on Emotional Intelligence in Organizations, her concepts and research have received three national awards and international distinctions. A frequent invited speaker and expert for the Organisation for Economic Co-operation and Development (OECD), the European Union and other international and national scientific organisations and public authorities, she has authored hundreds of articles and several books and serves on several editorial boards of scientific journals.

Orazio Giancola (PhD) is an associate professor of sociology of educational systems in the Department of Social and Economic Sciences at the Sapienza University of Rome. His research fields include comparative analysis of school and university systems and policies; equity in education; comparative analysis of performance and inequalities through international large-scale assessment; and governance processes at local, national and transnational levels. He is a member of network 28, ‘Sociologies of Education’, of the European Educational Research Association and of the Italian Association of Sociology (Education Section, AIS-EDU), and he was previously a member of the European Group for Research on Equity in Educational Systems.

Joyce Gubbels (PhD) is senior researcher at the Dutch Centre for Language Education in Nijmegen. She is the national project manager for the international large-scale assessment study of reading comprehension in primary education (Progress in International Reading Literacy Study (PIRLS)) and the co-national project manager, together with Martina Meelissen, for the international large-scale assessment programme in secondary education (Programme for International Student Assessment (PISA)). She has ample experience with policy-oriented studies in the field of education, and her research interests include language development, more specifically reading comprehension, and giftedness.

Heidi Harju-Luukkainen (PhD) is a professor at Nord University in Norway and a professor and vice director at the University of Jyväskylä, Kokkola University Consortium Chydenius, in Finland. She holds a PhD in education, a special education teacher qualification and a qualification in leadership and management. She has published more than 200 international books, journal articles and reports, and has worked on more than 25 projects globally. Heidi has worked at top-ranked universities in the United States, such as the University of California, Los Angeles, and the University of Southern California, as well as in many Nordic research universities. She has developed education programmes for universities, been a principal investigator of PISA subassessments in Finland and served as a board professional. Her research areas are early childhood education, justice in education and international student assessment. At the moment, she leads two international research groups in the field of education.

Maciej Jakubowski (PhD) is a policymaker, advisor and researcher in education and labour market policy. Dr Jakubowski holds a PhD degree in economics and a Master of Arts in sociology from the University of Warsaw, where he still works as an assistant professor. In 2014, he established the Evidence Institute to promote evidence-based policymaking and large-scale student assessments. Between 2008 and 2012, he worked in the OECD PISA team in Paris. Between 2012 and 2014, he served as an undersecretary of state (deputy minister) at the Polish Ministry of National Education. He is a consultant for the World Bank, the European Commission, the United Nations Development Programme, the United Nations Educational, Scientific and Cultural Organization and the OECD, and cooperates with think tanks and governments in Europe and the United States.

Anne Looney (PhD) is the executive dean of Dublin City University's Institute of Education, Ireland's largest faculty of education. From 2001 to 2016, she was the chief executive officer of the National Council for Curriculum and Assessment in Ireland. Professor Looney's current research interests include assessment policy and practice, curriculum design, teacher identity and professional standards for teachers and teaching. She has also published material on religious, moral and civic education, and education policy. She has conducted reviews for the OECD on school quality and assessment systems, and acts as a technical expert for the European Commission on initial teacher education reform. She is the current president of the International Professional Development Association.

Minna Maunula (PhD) works at the University of Jyväskylä, Kokkola University Consortium Chydenius, in Finland, as a university teacher and a developer of academic adult education. Her research interests focus on social issues in education, education policy and pedagogy discourse, and changing education, teaching and assessment. Higher education, academic adult education and digitalisation in education are also areas in which she has expertise.

Nele McElvany (PhD) is the executive director of the Centre for Research on Education and School Development and vice president for research at TU Dortmund University. As a professor of empirical educational research, her focus is educational processes from psychological and pedagogical perspectives. Her work on the individual, social and institutional conditions of educational processes and outcomes is nationally and internationally recognised.

Martina Meelissen (PhD) is senior researcher in the Department of Cognition, Data and Education at the University of Twente and has been the project manager of the Dutch part in several international large-scale assessment studies since 2003. These studies include the Trends in International Mathematics and Science Study (TIMSS), TIMSS Advanced, the International Computer and Information Literacy Study and PISA (co-national project manager with Joyce Gubbels). Her research interests include gender differences, digital literacy, and science, technology, engineering and mathematics in education.

Harm Naayer is a researcher and lecturer at the Faculty of Behavioural and Social Sciences at the University of Groningen, the Netherlands. His interests include the transition between primary and secondary education, comparative educational science and citizenship education. In 2007–2011, Harm was part of the research team for the first Dutch participation in the international large-scale assessment study on civic and citizenship education (International Civic and Citizenship Education Study 2009) and is now working, together with colleagues from the University of Amsterdam, on the International Civic and Citizenship Education Study 2022 assessment cycle.

Romuald Normand (PhD) is a professor of sociology in the Faculty of Social Sciences at the University of Strasbourg in France (Societies, Actors and Government of Europe Research Unit at the National Centre for Scientific Research) and an associate visiting professor at the Beijing Normal University and Aarhus University in Copenhagen. He works on comparative education policies and politics, Europeanisation and lifelong learning, higher education and research, and educational management. Romuald is convener of network 28, ‘Sociologies of Education’, of the European Educational Research Association. He is also a member of the editorial board of the *British Journal of Sociology of Education* and an advisory board member of the Springer series *Educational Governance Research*.

Michael O’Leary (PhD) holds the Prometric chair in assessment at Dublin City University and is director of the Centre for Assessment Research Policy and Practice in Education (CARPE) at the Institute of Education at Dublin City University. Professor O’Leary has acted as Ireland’s representative on the OECD’s Network A (student outcomes), was a member of the Board of Participating Countries of PISA during its first cycle and has provided statistical analysis for national surveys of reading and mathematics achievement in Ireland. He now leads an extensive programme of research at CARPE focused on assessment at all levels of the educational system and in the workplace.

Linda Rönnerberg (PhD) is a professor at the Department of Applied Educational Science at Umeå University in Sweden and a senior research fellow at the Department of Education at the University of Turku in Finland. She is also a docent in political science. Her research focuses on education governance and politics, with a particular focus on evaluation, privatisation and internationalisation in compulsory and higher education. She is a co-convenor of network 23, 'Policy Studies and Politics of Education' of the European Educational Research Association, deputy representative for Sweden in the Nordic Educational Research Association, co-editor of *Education Inquiry* and executive editor of the *European Educational Research Journal*.

Luca Salmieri (PhD) is a professor of sociology at the Sapienza University of Rome and a member of the Scientific Committee of the Italian Association of Sociology of Education. His research activity covers educational inequalities, cross-national differences in educational achievement, digitisation of educational space, gender and education, and transitions to adult life. He has published widely on these topics, especially in the context of the Italian educational systems. Professor Salmieri is a member of the scientific board of the *International Review of Sociology* and of the journal *Scuola Democratica*.

Jaap Scheerens (PhD) is a professor emeritus at the University of Twente in the Netherlands. He coordinated a research programme on school effectiveness. During his career, he was scientific director of the national research school for postgraduate training in educational science (Interuniversity Centre for Educational Sciences) and director of the research institute OCTO of the Faculty of Education at the University of Twente. He was involved in many international research projects funded by international organisations. He has published 20 books and approximately 100 articles in scientific journals, mainly addressing educational effectiveness and educational evaluation.

Darina Scully (PhD) is an assistant professor in child/adolescent learning and development at Dublin City University's Institute of Education. Her research interests include child development, well-being and the assessment of complex competencies. Early in her career, Darina was involved in the implementation of international large-scale assessments such as PISA, TIMSS and PIRLS at Dublin's Educational Research Centre. She subsequently held a postdoctoral position at Dublin City University's CARPE, where her research contributed to assessment policy within Irish education and also to industry decisions for global testing services provider Prometric. She has several peer-reviewed publications in the fields of education and assessment.

Gerry Shiel (PhD) is a researcher at the Educational Research Centre in Dublin. Prior to retirement, he had worked on the PISA assessment since the first cycle in 2000, and served as national project manager and as Ireland's representative on the PISA Governing Board. As a literacy researcher, Gerry is especially interested in the transition to computer-based testing in national and international assessments, and in how this impacts on factors associated with reading performance, including student attitudinal and engagement factors. He is also interested in links between instruction and reading performance. Gerry is an adjunct professor at Dublin City University's CARPE.

Sari Sulkunen (PhD) is an associate professor at the Department of Language and Communication Studies at the University of Jyväskylä in Finland. She has worked on several international literacy studies, for example as principal researcher on PISA reading in 2009, national research coordinator for PIRLS 2011 and vice national project manager for the Programme for the International Assessment of Adult Competencies 2012. More recently, she was a member of the PISA 2018 Reading Expert Group. She was also a member of the European Commission's High Level Group on Literacy (2011–2012), is a member of the European Literacy Policy Network and has regularly served on national expert committees on literacy policy. Her current research interests include teaching disciplinary literacy, multiliteracy education and adult literacy. She is also interested in the sociocultural factors related to the literacy performance of adolescents and adults.

Anneke Timmermans (PhD) is an associate professor at the Faculty of Behavioural and Social Sciences at the University of Groningen in the Netherlands. Her research interests include teacher expectations, factors related to the transition from primary to secondary schools in tracked educational systems, educational equity, multilevel modelling, and the validity and reliability of school performance indicators. In 2016, Anneke received the Early Career Researcher Award from the *British Journal of Educational Psychology* for the paper 'Accurate, inaccurate, or biased teacher expectations: do Dutch teachers differ in their expectations at the end of primary education?'

Gunda Tire is the head of international assessments on the Education and Youth Board of Estonia. She has been PISA national project manager for Estonia since 2007. Her responsibilities include preparation, implementation and dissemination of PISA and the Teaching and Learning International Survey in Estonia. She has also been involved in other international education projects. Gunda has been the editor of four national PISA reports, written articles about aspects of the Estonian education system, and worked with schools and the 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 into the Estonian education system. Gunda holds a Master of Science degree in educational sciences from KU Leuven University in Belgium.

Greetje van der Werf (PhD) is a professor emeritus (honorary) at the University of Groningen (the Netherlands). Until her retirement in June 2019, she was full professor of learning and instruction at the Faculty of Behavioural and Social Sciences (Department of Educational Science). Since 2013, she had also been a member of the board and vice dean of this faculty. Her main research expertise includes educational effectiveness, civics and citizenship education, and the influence of psychological precursors of school success. She has more than 30 years of experience in conducting large-scale multilevel longitudinal research in primary and secondary education, aimed at establishing the effects of national educational innovation programmes on student outcomes.

Einar Värä (PhD) is a chief expert on the Estonian Education and Youth Board. His main field of activities is in the area of secondary education. He has 14 years' experience in the field of curriculum development in the Estonian Ministry of Education and Research and subordinate institutions, and 35 years' experience in teaching, school management and teacher training. Dr Värä has participated in various international education cooperation projects in Georgia, Kyrgyzstan, Ukraine, Uzbekistan, etc. She is an author, co-author and translator (English, Estonian, Finnish, Russian) of approximately 100 textbooks, research articles and other publications in education, modern history, civics, law, the EU and other subjects.

Christina Wikström (PhD) is an associate professor in educational measurement and head of research and doctoral studies in the Department of Applied Educational Science at Umeå University. Her research interests include fairness in assessment and grading, and validity issues in large-scale educational measurement, especially in transition to higher education. Christina is chair of the Swedish Scholastic Aptitude Test scientific advisory board, president of the Association for Educational Assessment (Europe) and vice coordinator of the School of Quantitative Research Methods in Education.

Zlatica Jursová Zacharová (PhD) is a researcher in the Department of Psychology and Pathopsychology at the Faculty of Education at Comenius University in Bratislava, Slovakia. She teaches general, developmental, pedagogical, school and counselling psychology. Her primary research interests focus on developmental psycholinguistics, first and second language acquisition, bilingual teaching and teacher training. Since 2017, she has been a member of the International Association for Early Years Education and Care, a member of the Research Group for Early Years Language Learning and an external expert in the Standardisation of Counselling and Prevention System Project for Inclusion and Success in the Labour Market.

Katarína Zamborová (PhD) is an assistant professor in the Department of English Language in the Faculty of Applied Languages at the University of Economics in Bratislava, Slovakia. She teaches business English to undergraduate students. She was a visiting scholar at Northern Illinois University in the United States for two terms in 2014/2015. Her field of interest includes learner-centred teaching, incorporating mobile apps into English instruction and issues in international assessments. She presents and writes for local conferences in Czechia and Slovakia but has been also a presenter at conferences in the United States. In 2019, she was awarded a Betty Azar Travel Grant for Practicing English as a Second Language / English as a Foreign Language Teachers from the organisation Teaching English to Speakers of Other Languages.

Part I

INTRODUCTION

CHAPTER 1

Cross-national achievement surveys and educational outcomes: a brief history

Louis Volante, Sylke V. Schnepf and Don A. Klinger

Abstract

This chapter provides an introduction to the topic of cross-national achievement surveys and their use in monitoring educational outcomes. After a brief review of the scope of prominent surveys administered by the Organisation for Economic Co-operation and Development and the International Association for the Evaluation of Educational Achievement, we examine methodological issues as well as the contested nature of policy implications stemming from comparative studies. We discuss how design issues rarely, if ever, factor into debates on strategic large-scale reform. At the same time, this chapter, as well as the ensuing national profiles, underlines the diversity of policy decisions that are often taken in response to achievement results. A key conclusion is therefore that reacting to cross-national achievement surveys is not a uniform process, as is often suggested, but influenced by a range of political, cultural and historical perspectives. A framework for the national profiles in the second part of this book is also provided at the end of this chapter.

Introduction

Cross-national achievement surveys administered by the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) have steadily grown in scope and global significance since their inception in the 1960s. Governments around the world anxiously await the results from these benchmark measures to evaluate the effectiveness of their education systems; to assess student outcomes such as reading, and scientific and mathematical literacy; and, ultimately, to judge the effectiveness of their national education policies. Critics of international testing – particularly of the Programme for International Student Assessment (PISA) – have argued that these measures promote a narrow view of education that often leads to short-term and ill-informed reforms to national education systems (Andrews et al., 2014; Goldstein, 2014; Rappleye et al., 2019; Steiner-Khamsi and Waldow, 2018). However, it is equally clear that reacting to cross-national achievement surveys is not a uniform process, as is often suggested, but influenced by a range of political, cultural and historical perspectives (Volante, in press).

This volume examines the range and salience of different international achievement surveys for monitoring educational outcomes and policy reform in 10 European countries: Germany, Estonia, Ireland, France, Italy, the Netherlands, Poland, Slovakia, Finland and Sweden. Collectively, the national profiles provide a critical analysis of the use, and misuse, of cross-national achievement surveys for monitoring educational outcomes and policy formation. It is worth noting that, although these countries represent a minority of European Union (EU) Member States, they were strategically selected on the basis of their relative results, their histories with international testing and their general geographical locations. A key aim of this volume is, therefore, to document the trends and trajectories of education reform associated with cross-national achievement surveys administered by the OECD and/or the IEA and to offer insights into the future of transnational governance within the EU.

Cross-national achievement surveys

As previously noted, the OECD and the IEA are the most prominent international organisations that design and administer cross-national achievement surveys. The ensuing discussion offers a brief summary of the various OECD and IEA measures that currently exist, so that the reader is aware of the nature, scope and level of international participation in these surveys. This section also briefly outlines some of the most notable changes and additions to these various measures since they were introduced. It is important to note at the outset that, although we outline the wide range of measures that currently exist, it is the OECD, and PISA in particular, that has traditionally received the vast majority of international attention in academic and policy communities. This general pattern is also highlighted in the national profiles in Part 2 of this volume.

Organisation for Economic Co-operation and Development surveys

The OECD's interest in education began in 1964, when it laid the groundwork for a new field of study known as the 'economics of education', and the closely related concept of human capital (Hanushek and Woessman, 2015; Ydesen, 2019). It was posited that economic growth may depend as much on increases in human capital (at that time measured by the number of years of education) as on the changes in physical capital (e.g. infrastructure, buildings, technology) (Svennilson et al., 1962). Over time, notions of human capital began to increasingly recognise the importance of the quality of learning, not just the number of years of schooling. In order to capture cross-national differences in educational quality, the OECD began developing PISA in the 1990s to measure student performance (OECD, 1997).

The initial administration of PISA in 2000 included 28 OECD countries and 4 partner countries. Since then, this triennial survey of 15-year-old students has steadily grown, and it included 80 nations in the 2018 administration ⁽¹⁾. An interesting feature of PISA is that each of three 'life skill' domains – reading literacy, mathematical literacy and scientific literacy – is assigned as a major testing domain on a rotating format and, as a result, is assessed in greater detail. Table 1.1 summarises the timelines for the major and minor literacy emphases since the inception of PISA in 2000.

Table 1.1: Chronology of major and minor literacy domains in PISA

Domain	2000	2003	2006	2009	2012	2015	2018
Reading literacy	Major	Minor	Minor	Major	Minor	Minor	Major
Mathematical literacy	Minor	Major	Minor	Minor	Major	Minor	Minor
Scientific literacy	Minor	Minor	Major	Minor	Minor	Major	Minor

NB: Owing to the COVID-19 pandemic, the OECD has postponed the PISA 2021 assessment to 2022 and the PISA 2024 assessment to 2025.

The OECD has continually added new measures to complement its focus on major and minor literacy domains. For example, measures of creative problem-solving and financial literacy were introduced in 2012, collaborative problem-solving in 2015 and global competencies in 2018. The OECD also plans to introduce a test of creative thinking in 2021 and digital learning in 2024 – with delays likely because of the COVID-19 pandemic. This scope, coupled with the 'PISA for Development' initiative, which facilitates PISA participation by low- and middle-income countries, make it easy to understand how the OECD's influence has dwarfed that of comparable, and older, surveys, such as those administered by the IEA (Volante, 2016). It is hardly surprising that PISA has been referred to as 'one of the largest non-experimental research exercises the world has ever seen' (Murphy, 2014, p. 898).

⁽¹⁾ See the OECD's website for a full list of participating nations/economies (<https://www.oecd.org/pisa/aboutpisa/pisa-2018-participants.htm>).

The OECD also has a long history of developing measures designed to assess adult literacy. For example, it initially introduced the International Adult Literacy Survey in 1994, which was replaced by the Adult Literacy and Life Skills Survey in 2003. The current Programme for the International Assessment of Adult Competencies (PIAAC) builds on the previous surveys by examining ‘foundational’ information-processing skills in three key areas: literacy, numeracy and problem-solving (Volante and Ritzen, 2016). According to the OECD, these core skills form the basis for the development of other higher-level skills that are considered essential for adults aged 16–65 years in home, school, work and community settings (OECD, 2015). In line with its home page tagline ‘Better policies for better lives’, the OECD asserts that PIAAC, like PISA, helps governments in assessing, monitoring and analysing the level and distribution of skills among their adult populations (OECD, 2013). Overall, when one considers also the 2018 launch of the International Early Learning and Child Well-being Study, nicknamed ‘Baby PISA’, which assesses 5-year-olds, it is easy to understand the enormous influence that the OECD has amassed in monitoring educational outcomes across the lifespan.

International Association for the Evaluation of Educational Achievements

The IEA appears to have been founded in Hamburg, Germany, in 1959 during a meeting of the United Nations Educational, Scientific and Cultural Organization Institute of Education (Purves, 1987). The first IEA study was a mathematics survey conducted in 12 countries from 1962 to 1965 (Husen, 1979). The Six Subject Survey soon followed, from 1966 to 1973, which assessed science, reading comprehension, literature, English as a foreign language, French as a foreign language and civic education in a larger sample of 21 countries (Walker, 1976). Collectively, these early IEA studies laid the groundwork for its two most prominent and cyclical testing programmes: the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS). Introduced in 1995, and taking place every 4 years, TIMSS measures trends in mathematics and science achievement at fourth- and eighth-grade levels. TIMSS Advanced was also introduced in 2015 as a new measure of advanced mathematics and physics for students in their final year of secondary school. PIRLS was introduced in 2001 and measures trends in reading comprehension at fourth-grade level every 5 years. The most recent administrations of TIMSS in 2019 included 64 countries and 8 benchmarking systems; in 2016, PIRLS included 50 countries and 11 benchmarking systems. Table 1.2 summarises the chronology of TIMSS and PIRLS.

Table 1.2: Chronology of TIMSS and PIRLS

Administration	TIMSS	TIMSS Advanced	PIRLS
First	1995	1995	2001
Second	1999	2008	2006
Third	2003	2015	2011
Fourth	2007		2016
Fifth	2011		2021 (*)
Sixth	2015		
Seventh	2019		

(*) At the time of writing, it was still not clear how the PIRLS 2021 survey would be run, given the difficulties involved in measuring pupils' achievement during the COVID-19 pandemic. PIRLS 2021 will offer the PIRLS assessment of literary and informational reading in a digital format.

The IEA also administers less well-known and less well-publicised assessments, such as the International Civic and Citizenship Education Study, initially administered in 2009, which examines the degree to which eighth- and ninth-grade students are prepared to undertake their roles as citizens in a world where contexts of democracy and civic participation continue to change. Similarly, the International Computer and Information Literacy Study was introduced in 2013 to measure eighth-grade students' computer and information literacy. According to the IEA, computer and information literacy refers to an individual's ability to use computers to investigate, create and communicate in order to participate effectively at home, at school, in the workplace and in the community. Finally, the IEA has also undertaken the Early Childhood Education Study, Language Education Study, Preprimary Project, Computers in Education Study, Reading Literacy Study, and Written Composition Studies involving smaller samples of participating nations at various points in time ⁽²⁾.

Although the OECD and the IEA produce the most internationally well-known educational achievement surveys, other regionally focused, but similarly designed, cross-country achievement surveys exist, such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality study; the Programme for the Analysis of Education Systems, covering countries in sub-Saharan Africa, the Indian Ocean, the Middle East and South-East Asia; and the Latin-American Laboratory for Assessment of the Quality of Education.

Policy outreach

Both the OECD and the IEA draw on extensive data sets from their respective international surveys to offer a range of reports and policy briefs that are freely available online. For example, the OECD's *PISA in Focus* and *Education Indicators in Focus* aim to identify features and characteristics of the best-performing education systems around the world that follow from their cross-national analyses of PISA and PIAAC data. The OECD also

⁽²⁾ See the IEA's website for further details (<https://www.iea.nl/studies/iea/other>).

launched a one-stop online tool called Education GPS, which provides users with an easy way to analyse results by country and to review education policies that are considered instrumental in promoting system effectiveness⁽³⁾. The IEA also offers similar tools and open access to policy briefs, such as its *IEA Compass: Briefs in Education Series*, which addresses issues of interest to a broad range of educational stakeholders, particularly those involved in influencing educational decision-making and policymaking⁽⁴⁾. Thus, it is clear that both international organisations take an active role in disseminating best practices and influencing education policy debates at international and national levels.

Creation of educational achievement data

Data provided by the international organisations' achievement surveys are huge in number and coverage of pupils' achievement measures, and comprise vast quantities of contextual information on parents, teachers and schools. PISA, TIMSS and PIRLS pupil sample sizes are often larger than those of other internationally organised surveys (e.g. country samples of the European Social Survey), given that, for example, PISA 2018 included samples of more than 3 000 pupils for most European countries. The organisers' provision of one single continuous achievement score scale, which is assumed to be the same for all participating countries, whether Finland, South Africa or Spain, generates the only possibility currently available to compare the outcomes of education systems cross-nationally. It is therefore not surprising that PISA, TIMSS and PIRLS have transformed educational research opportunities and are extensively used to provide evidence for policymaking purposes.

Although survey results generally attract substantial attention, considerably less effort is spent on reflection on what kind of culturally comparable education skills are actually measured and how the data are generated. This is not surprising, given that educational achievement surveys are highly complex and the accompanying large technical reports explaining their design are not easy to digest. In general, two key components of all achievement surveys determine their quality: reaching a valid measurement of educational achievement and realising a representative sample of pupils (Groves et al., 2009).

Measurement

Clearly, the most important aspect is how international organisations measure educational achievement cross-nationally. In general, the OECD and the IEA apply very similar approaches, which differ only in specific details. We focus now on the more well-known PISA, in which the OECD first defines proficiency levels for each subject tested. All participating countries need to agree to this definition framework. On the basis of these proficiency levels, a consortium proposes draft questions. Participating countries review these items and rate them in terms of appropriateness for measuring the subject within

⁽³⁾ See the OECD's website for further details (<http://gpseducation.oecd.org/>).

⁽⁴⁾ <https://www.iea.nl/index.php/publications/series-journals/iea-compass-briefs-education-series>.

their country. Country rankings are taken into account in selecting the final items. After their translation, they are assessed in field trials in all countries. An international expert group reviews which items are fit for purpose and perform similarly across countries, using statistical analysis. In a last step, the participating countries agree on the final set of items (OECD, 2019).

Students who have been sampled sit a test (a 2-hour test for PISA), completing randomly assigned booklets containing only a fraction of the final items. As a consequence, all students do not provide answers on many of the final items. Survey organisers fill this gap by using multiple imputation to create test scores for each pupil in each subject. Students' answers to many questions are summarised in an achievement score by applying complex item response models. For PISA, this involves estimating the difficulty of each question and its power to discriminate between individuals with low ability and individuals with high ability based on students' answers to the questions. These estimates are then used to estimate the final achievement score (see OECD, 2016, for a more detailed discussion of this process). It is the mean of these achievement values across all pupils in a country that constitutes the country's overall educational outcome. Those outcomes are regularly reported in 'league tables' that provide country rankings by pupils' mean literacy scores measured in the survey.

As described, the generation of the achievement scale proceeds through numerous steps from education definition to item generation, implementation and complex modelling. All these steps are described in general in the international organisations' reports. However, organisers' technical reports rarely discuss assumptions and choices made for each of the steps of the scale generation, or the robustness of the scale to possible choices available. The academic community cannot fill this void, as organisations' data for creating the scores, such as deleted items, field trials data and all items used to create achievement scores, are not provided. This lack of transparency is likely to foster a number of concerns regarding the validity of the cross-national continuous achievement scale provided by international organisations (for a summary, see Araujo et al., 2017; Schnepf, 2018). Most notable is criticism doubting the comparability of achievement scores between countries. Functional literacy as measured by PISA, or curriculum literacy focused on by TIMSS, is likely to differ between countries given diversity in culture, history and economic development. Measuring achievement based only on items that fit the model-based approach ignores important and country-specific dimensions of education skills that the one-dimensional PISA, TIMSS and PIRLS measures cannot capture (Goldstein, 2017; Hopmann et al., 2007).

Another argument touches on whether items indeed measure students' proficiency or, at least to some degree, their ability to understand survey-specific test structures (Meyerhoefer, 2007). With the move from pen and paper to computer surveys since PISA 2015, this problem might have become even more significant. Furthermore, although item response models are the key component for generating the educational achievement scale, their suitability for evaluating educational achievement between different groups has been doubted, given that the models can produce biased measures of student ability (Jacob, 2016; Jacob and Rothstein, 2016). In addition, TIMSS and PISA are based on different item response models. Using diverse achievement survey data

and item response models, Brown et al. (2007) found that the choice of model has a clear impact on educational outcomes, whereas Jerrim et al. (2018), using a smaller set of countries, did not find such an impact.

Representation

Besides measurement, representativeness is also a key element of a high-quality survey. Only if the sampled students are representative of all of a country's population of pupils (for PISA this is 15-year-olds) can the survey results be generalised to country level and be compared cross-nationally. Similarly to measurement, there are a number of criteria that need to be followed correctly to reach representativeness. In general, international organisations first select a representative sample of schools, within which they then sample students. The first problem regards exclusions: survey organisers allow countries to exclude newly immigrated and special needs students. Some countries exclude considerably more students than others, and can thereby influence their overall country results (Wuttke, 2007). Second, although survey organisers set response rate thresholds (the OECD requests countries to meet an 85 % school response and an 80 % student response), these thresholds are no guarantee that possible bias of results due to non-response is negligible (Micklewright et al., 2012). Without any information on cross-national patterns of non-response, it is impossible to say how much non-response might have affected a country's achievement results. Exploiting in-depth data for England, Micklewright et al. (2012) show that the bias could be two to three times bigger than the published standard error of a country's achievement score.

Clearly, the lack of transparency on measurement and representation issues for data generation is problematic. Because additional data for the production of the scores are not available, the opportunities for research communities to examine the quality of the educational achievement data are limited. Hence the examination of the quality of the educational achievement scale remains mainly a monopoly of the data producer. Given that the results of educational achievement surveys inform education policy design in many countries in the world, international organisations' full transparency on data creation and the involvement of the research community and stakeholders in the process of data generation could provide much more legitimacy for these data (Araujo et al., 2017).

Policy monitoring and evaluation in the European context

Educational achievement survey results are used for monitoring educational outcomes cross-nationally, without much consideration of the problems related to their creation. The country chapters of this volume will describe in greater detail country-specific aspects of data use in the policy context; this section describes the framework for educational monitoring in the EU, of which all countries covered in this volume are Member States. In the EU, education falls under the principle of subsidiarity, so that the primary responsibility for education and training policies lies with the Member States. This limits the EU's role in education policy support. Over the past decade, European policy monitoring has been

guided by the strategic framework for European cooperation in education and training (ET 2020) and the European semester. ET 2020 outlined the indicators and arrangements for educational monitoring at EU level. One of the seven indicators of ET 2020 was based on PISA data: 'The rate of 15-year-olds with insufficient abilities in reading, mathematics and science should be less than 15 %' (European Commission, 2020a).

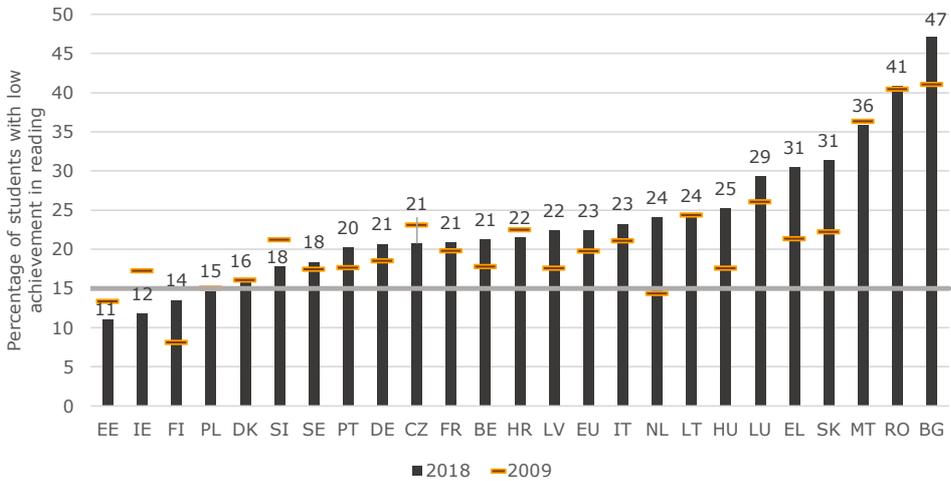
The European semester, an annual activity introduced in 2010, coordinates economic and social policies among Member States and addresses EU challenges. In addition, country education policies are closely monitored by providing country-specific information on educational outcomes. Furthermore, annual semester country-specific recommendations address education priorities and are vital for supporting Member States' education reforms. The results of the European semester's findings enter the European Commission's Technical Support Instrument, which, for example, had a budget of EUR 223 million for 2017–2020 (European Commission, 2020a). This instrument supports Member States with specific projects to tackle the challenges they face. In addition, Member States can also receive funding for education projects from the European Structural and Investment Funds.

The seven ET 2020 benchmarks, including the one on low-achieving pupils deriving from PISA data, have been valuable in encouraging national policy dialogues, ensuring momentum for reforms and building up a working framework for European cooperation (European Commission, 2020a). As a consequence, at the end of 2020, a general sentiment among European stakeholders in favour of more far-reaching coordination, probably also nourished by challenges faced in the area of digital learning during COVID-19, led to the formulation of the ambitious plan to set up a European Education Area (EEA) by 2025 (European Commission, 2020a). The communication on the EEA (European Commission, 2020b) outlines how cooperation between Member States can further strengthen the quality, inclusiveness, and green and digital dimensions of Member States' education systems. It is supported by, among other instruments, the digital education action plan (2021–2027), which aims to ensure high-quality, inclusive and accessible digital education in Europe, learning from the COVID-19 pandemic and making European education fit for the digital age (European Commission, 2020c). Like ET 2020, the EEA also envisages the use of PISA data for monitoring educational outcomes and inclusiveness across European countries.

Although the European education policy framework is clearly set out as described previously, and was recently adapted to take stock of current global challenges, such as the COVID-19 pandemic and the green and digital transition, monitoring components of the EEA is more difficult.

Figure 1.1 provides an example of European educational monitoring with PISA data. It gives the percentages of students who were low achievers in PISA reading in 2009 and 2018 by country. Only four European countries managed to reach the benchmark of fewer than 15 % of 15-year-olds low achieving in 2018, as set by ET 2020. For countries that do not perform well on benchmark indicators, this is picked up in country-specific European semester reports, which then also formulate recommendations for improvement.

Figure 1.1: Percentages of students with low PISA reading achievement in 2009 and 2018, by country and compared with the ET 2020 benchmark



NB: The grey line shows the ET 2020 benchmark.

Source: OECD PISA 2018 data (OECD, 2019).

Measuring, and hence monitoring, educational outcomes every 3 years in accordance with the timing of PISA data collection seems generally sensible, as it is unlikely that education systems can change quickly and have an immediate effect on students' learning. In the aftermath of an external shock, such as the COVID-19 pandemic, which greatly affected the learning opportunities of students during physical school closure, more timely mapping of educational outcomes would be highly valuable. The literature agrees that the pandemic has considerably augmented pre-existing educational inequalities for most European countries (Blaskó and Schnepf, 2020). The postponement of PISA data collection from 2021 to 2022 and the problem of collecting data on learning outcomes during physical school closure, however, currently present severe problems in terms of monitoring educational outcomes. Monitoring them is important for understanding the severity of changes in educational outcomes as a result of COVID-19. Currently, the most important information for educational monitoring derives from a small number of studies exploiting administrative data for single countries (Engzell et al., 2020; Maldonado and De Witte, 2020) and simulation studies (Azevedo et al., 2020; Di Pietro et al., 2020). Up to now, despite discussions at European level, there has been no prospect of ensuring, or vision of how to ensure, mapping of educational outcomes across Europe in a more timely manner.

Requirements for evaluating education policies are always set higher than requirements for monitoring. Underperformance in monitoring is likely to lead to failure in evaluation. The evaluation of policies requires us to answer the question 'What policies work?' and, thus, requires the examination of a causal link between a specific policy design and the

outcome it produces. All cross-national educational achievement surveys are based on cross-sectional data and therefore collect new samples of students for every round. Cross-sectional survey data are less suitable to measure policy impact, as any results could be guided by differences in the student samples. Scholars (see, for example, Goldstein, 2017) therefore argue that educational achievement surveys do not possess the necessary data requirements for policy evaluation. Possible strategies that use counterfactual impact analysis to measure policy impact (Cordero et al., 2017) have shortcomings, such as specific additional data requirements and the need to make assumptions that are difficult to prove. As long as the existing educational achievement surveys do not employ a longitudinal design, that is, a design that follows the learning progress of students over time and hence includes the same students in every survey wave, education policy evaluation with educational achievement surveys, such as PISA, TIMSS and PIRLS, will remain limited at cross-national level.

The contestation of policies, practices and political reforms

Concerns regarding the policy impact of the OECD and IEA surveys are not a recent phenomenon. Indeed, scholars have long voiced concerns over the undue influence of both these international organisations well before their current survey iterations existed (Centeno, 2019; Husen, 1987). Nevertheless, it was not until the introduction of PISA that academics around the world began to extensively document the intended and unintended influence of this measure on national assessment policies, curricular standards and strategic planning across a wide range of international jurisdictions (Grek, 2009; Pons, 2017; Volante, 2018; Zhao, 2020). Indeed, cross-disciplinary international criticism of PISA is exemplified by the open letter that was published in *The Guardian* (British national daily) newspaper, entitled ‘OECD and PISA tests are damaging education worldwide – academics’, which was directed to Dr Andreas Schleicher, the current director of the OECD programme (Andrews et al., 2014). This open letter, which was signed by a group of more than 80 high-profile academics and education practitioners from around the world, underlined a litany of negative policy impacts associated with PISA. Interestingly, these points were reiterated in another open letter that was published in a special edition of the academic journal *Policy Futures in Education* (Vol. 12, No 7), which grew to include more than 130 signatories at the time of publication (Murphy, 2014).

It is fair to say that a healthy debate continues to exist on the relative role cross-national achievement surveys should play in monitoring national educational outcomes, as well as on the underlying belief that these measures are the most effective levers to spur national education reforms. Nevertheless, although one may be naturally inclined to assume that policymakers automatically use international and national large-scale test results to inform evidence-based policy decisions, this relationship is unclear based on a closer examination of the antecedents of national education reforms (Volante, 2015, in press). Indeed, comparative analyses suggest that policymakers across Europe do not necessarily react in a predictable manner when confronted with large-scale

achievement results (Breakspear, 2012; Volante and Ritzen, 2016). These differences have less to do with outright objection to cross-national achievement surveys or their methodological limitations than with unique contextual issues that influence national policy agendas.

Understanding how policy responses are shaped by various political, cultural and historical perspectives across the EU is a notable and key analytical objective of this volume. In doing so, we seek to explicate under what conditions particular achievement results are utilised for monitoring purposes, and their resulting impacts, both positive and negative, on national policy agendas. This volume also builds on and extends previous studies by providing a contemporary analysis of the level of reactivity that various EU jurisdictions exhibit in relation to cross-national achievement surveys (Baird et al., 2011; Breakspear, 2012; Eurydice, 2009; Volante, 2018). It is worth noting that some of the countries profiled in this volume (i.e. France, Poland and Slovakia) have traditionally received less attention in the international research literature than their European counterparts such as Germany and Finland. Ultimately, this volume attempts to evaluate the policies, practices, and political discourses and reforms that have accompanied the reception of cross-national achievement results across a diverse cross-section of EU Member States that have experienced various degrees of measurable education system ‘success’.

Structure and contents of this volume

This edited volume is divided into three parts. Part I, the current introductory chapter, provides a brief overview of the various cross-national surveys administered by the OECD and the IEA, along with a general discussion of measurement, evaluation and monitoring issues.

Part II provides a more nuanced analysis of policies, practices and political reforms that stem from cross-national achievement surveys in 10 EU nations: Germany, Estonia, Ireland, France, Italy, the Netherlands, Poland, Slovakia, Finland and Sweden. These source chapters are each divided into four distinct sections to allow readers to easily compare different national contexts.

The first section in each national profile provides a brief overview of the structure of the compulsory school system in the specified country. The general organisation and institutional features of the compulsory school system are outlined, along with the general governance structures used within the education system. The second section describes the level of participation and general performance patterns in relation to prominent international achievement surveys, such as PISA, TIMSS, PIRLS and PIAAC. The third section provides a description of the monitoring processes used in each country. It includes a description of how achievement results are used to inform education policy development and reform. On the basis of this discussion, the reader will have a clear indication of how salient particular international achievement surveys are for data-driven decision-making processes. The final section offers an evaluation of the

intended positive and unintended negative outcomes that derive from education quality monitoring that is associated with cross-national achievement surveys. How evidence from international survey results has been scrutinised and used to spur political reforms, as well as the impact of these reforms in relation to the available empirical evidence, is also discussed.

Part III (the conclusion) provides a summary of the trends and trajectories of education reform associated with cross-national achievement surveys. Of course, we freely acknowledge that the diversity of approaches represented in this volume would undoubtedly be magnified if we had analysed all 27 EU nations. Nevertheless, we are confident that this volume offers valuable insights into the prominent policy discourses that have shaped educational policy monitoring and transnational governance within the EU.

REFERENCES

- Andrews, P., Atkinson, L., Ball, S. J., Barber, M., Beckett, L., Berardi, J. et al. (2014), 'OECD and PISA tests are damaging education worldwide – academics', *The Guardian*, 6 May (<https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>).
- Araujo, L., Saltelli, A. and Schnepf, S. (2017), 'Do PISA data justify PISA-based education policy?', *International Journal of Comparative Education and Development*, Vol. 19, No 1, pp. 20–34.
- Azevedo, J. P., Hasan, A., Goldemberg, D., Iqbal, S. A. and Geven, K. (2020), 'Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes: a set of global estimates', *Policy Research Working Papers*, No 9284, World Bank Group.
- Baird, J., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T. and Daugherty, R. (2011), *Policy Effects of PISA*, Oxford University Centre for Educational Assessment, Oxford.
- Blaskó, Z. and Schnepf, S. V. (2020), 'Educational inequalities in Europe and physical school closures during COVID-19', *Fairness Policy Brief Series*, No 04/2020 (https://ec.europa.eu/jrc/sites/jrcsh/files/fairness_pb2020_wave04_covid_education_jrc_i1_19jun2020.pdf).
- 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*, No 71, OECD Publishing, Paris.
- Brown, G., Micklewright, J., Schnepf, S. V. and Waldmann, R. (2007), 'International surveys of educational achievement: how robust are the findings?', *Journal of the Royal Statistical Society: Series A*, Vol. 170, No 3, pp. 623–646.
- Centeno, V. G. (2019), 'The birth of the OECD's education policy area', in Ydesen, C. (ed.), *The OECD's Historical Rise in Education*, Palgrave Macmillan, London, pp. 63–82.
- Cordero, J., Cristobal, V. and Santin, D. (2017), 'Causal inference on education policies: a survey of empirical studies using PISA, TIMSS and PIRLS', *Journal of Economic Surveys*, Vol. 32, No 3, pp. 878–915.
- Di Pietro, G., Biagi, F., Costa, P., Karpinski, Z. and Mazza, J. (2020), *The Likely Impact of COVID-19 on Education: Reflections based on the existing literature and recent international datasets*, JRC Technical Report, Publications Office of the European Union, Luxembourg.
- Engzell, P., Frey, A. and Verhagen, M. D. (2020), 'Learning loss due to school closures during the COVID-19 pandemic' (<https://osf.io/preprints/socarxiv/ve4z7/>).
- European Commission (2020a), Commission staff working document accompanying the document Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions on achieving the European Education Area by 2025, SWD(2020) 212 (<https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12524-Achieving-the-European-Education-Area>).
- European Commission (2020b), Commission communication – Achieving the European Education Area by 2025, COM(2020) 625 final (https://ec.europa.eu/education/sites/default/files/document-library-docs/eea-communication-sept2020_en.pdf).

- European Commission (2020c), *Digital Education Action Plan, 2021–2027* (https://ec.europa.eu/education/sites/default/files/document-library-docs/deap-communication-sept2020_en.pdf).
- Eurydice (2009), *National Testing of Pupils in Europe: Objectives, organisation and use of results* (https://www.eurydice.si/publikacije/National-testing-of-pupils-in-Europe_Objectives,-organisation-and-use-of-results-EN-HI.pdf?t=1569928954).
- Goldstein, H. (2014), 'Responses to Andreas Schleicher's reply to open letter', *Policy Futures in Education*, Vol. 12, No 7, pp. 880–882.
- Goldstein, H. (2017), 'Measurement and evaluation issues with PISA', in Volante, L. (ed.), *The PISA Effect on Global Educational Governance*, Routledge, Abingdon-on-Thames, UK.
- Grek, S. (2009), 'Governing by numbers: the PISA "effect" in Europe', *Journal of Education Policy*, Vol. 24, No 1, pp. 23–37.
- Groves, R. M., Fowler, F., Couper, M., Lepkowski, J., Singer, E. and Tourangeau, R. (2009), *Survey Methodology – Second edition*, John Wiley & Sons, Inc., Hoboken, NJ.
- Hanushek, E. A. and Woessman, L. (2015), *The Knowledge Capital of Nations: Education and the economics of growth*, MIT Press, Cambridge, MA.
- Hopmann, S., Brinek, G. and Retzl, M. (eds) (2007), *PISA According to PISA*, University of Vienna Press, Vienna.
- Husen, T. (1979), 'An international research venture in retrospect: the IEA surveys', *Comparative Education Review*, Vol. 23, No 3, pp. 371–385.
- Husen, T. (1987), 'Policy impact of IEA research', *Comparative Education Review*, Vol. 31, No 1, pp. 29–46.
- Jacob, B. A. (2016), 'Student test scores: how the sausage is made and why you should care', *Economic Studies at Brookings Evidence Speaks Reports*, Vol. 1, No 25, Center on Children and Families at Brookings (https://www.brookings.edu/wp-content/uploads/2016/08/ccf_20160811_jacob_evidence_speaks1.pdf).
- Jacob, B. and Rothstein, J. (2016), 'The measurement of student ability in modern assessment systems', *Journal of Economic Perspectives*, Vol. 30, No 3, pp. 85–108.
- Jerrim, J., Parker, P., Choi, A., Chmielewski, A., Sälzer, C. and Shure, N. (2018), 'How robust are cross-country comparisons of PISA scores to the scaling model used?', *Educational Measurement: Issues and Practice*, Vol. 37, No 4, pp. 28–39.
- Maldonado, E. and De Witte, K. (2020), 'The effect of school closures on standardised student test outcomes', *KU Leuven Discussion Paper Series*, No DPS20.17 (<https://lirias.kuleuven.be/3189074?limo=0>).
- Meyerhoefer, W. (2007). 'Testfähigkeit – Was ist das?', in Hopmann, S., Brinek, G. and Retzl, M. (eds), *PISA According to PISA*, pp. 57–92, University of Vienna Press, Vienna.
- Micklewright, J., Schnepf, S. V. and Skinner, C. J. (2012), 'Non-response biases in surveys of school children: the case of the English PISA samples', *Journal of the Royal Statistical Society, Series A*, Vol. 175, No 4, pp. 915–938.

- Murphy, D. (2014), 'Issues with PISA's use of its data in the context of international education policy convergence', *Policy Futures in Education*, Vol. 12, No 7, pp. 893–916.
- OECD (Organisation for Economic Co-operation and Development) (1997), Resolution of the Council on a decentralised programme for producing indicators on student achievement on a regular basis ([http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=C\(97\)176/FINAL&docLanguage=En](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=C(97)176/FINAL&docLanguage=En)).
- OECD (2013), *OECD Skills Outlook 2013: First results from the Survey of Adult Skills*, OECD Publishing, Paris ([https://www.oecd.org/skills/piaac/Skills volume 1 \(eng\)-full v12-eBook \(04 11 2013\).pdf](https://www.oecd.org/skills/piaac/Skills volume 1 (eng)-full v12-eBook (04 11 2013).pdf)).
- OECD (2015), 'Adult literacy' (www.oecd.org/edu/innovation-education/adultliteracy.htm).
- OECD (2016), *PISA 2015 Technical Report*, OECD Publishing, Paris (<https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>).
- OECD (2019), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (<https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>).
- Pons, X. (2017), 'Fifteen years of research on PISA effects on education governance: a critical review', *European Journal of Education*, Vol. 52, No 2, pp. 131–144.
- Purves, A. C. (1987), 'The evolution of the IEA: a memoir', *Comparative Education Review*, Vol. 31, No 1, pp. 10–28.
- Rapplee, J., Komatsu, H., Uchiba, Y., Krys, K. and Markus, H. (2019), "Better policies for better lives"? Constructive critique of the OECD's (mis)measure of student well-being', *Journal of Education Policy*, Vol. 35, No 2, pp. 258–282.
- Schnepf, S. V. (2018), *Insights into survey error of large scale educational achievement surveys*, JRC Technical Report, Publications Office of the European Union, Luxembourg.
- Steiner-Khamsi, G. and Waldow, F. (2018), 'PISA for scandalisation, PISA for projection: the use of international large-scale assessments in education policymaking – an introduction', *Globalisation, Societies and Education*, Vol. 16, No 5, pp. 557–565.
- Svennilson, I., Harbison, F. H., Tinbergen, J., Bos, H. C., Poignant, R., Mober, S. and Lyons, R. F. (1962), *Policy conference on economic growth and investment in education*, OECD Publishing, Paris.
- Volante, L. (2015), 'The impact of PISA on education governance: some insights from highly reactive policy contexts', *International Studies in Educational Administration*, Vol. 43, No 2, pp. 103–117.
- Volante, L. (ed.) (2016), *The intersection of international achievement testing and educational policy: Global perspectives on large-scale reform*, Routledge, Abingdon-on-Thames, UK.
- Volante, L. (ed.) (2018), *The PISA Effect on Global Educational Governance*, Routledge, London.
- Volante, L. (in press), 'The PISA pendulum: political discourse and education reform in the age of global reference societies', in Mattei, P., Dumay, X., Mangez, E. and Behrend, J. (eds), *OUP Handbook on Education and Globalization*, Oxford University Press, Oxford.

- Volante, L. and Ritzen, J. (2016), 'The European Union, education governance, and international education surveys', *Policy Futures in Education*, Vol. 14, No 7, pp. 988–1004.
- Walker, D. A. (1976), *The IEA Six Subject Survey: An empirical study of education in twenty-one countries*, Wiley, New York.
- Wuttke, J. (2007), 'Uncertainties and bias in PISA', in Hopmann, S., Brinek, G. and Retzl, M. (eds), *PISA According to PISA*, University of Vienna Press, Vienna.
- Ydesen, C. (ed.) (2019), *The OECD's Historical Rise in Education: The formation of a global governing complex*, Palgrave Macmillan, London.
- Zhao, Y. (2020), 'Two decades of havoc: a synthesis of criticism against PISA', *Journal of Educational Change*, Vol. 21, pp. 245–266.

Part II

NATIONAL PROFILES

CHAPTER 2

Cross-national achievement surveys and educational monitoring in Germany

Nele McElvany

Abstract

This chapter provides an overview of the history and current state of and future perspectives on cross-national achievement surveys and educational monitoring in Germany. After an introduction that provides an overview of the German education system, the second section describes the history of Germany's participation in international achievement surveys and the general performance patterns of German students in international large-scale assessment studies. The Progress in International Reading Literacy Study is used as an in-depth example. The third section examines education policy monitoring in Germany and discusses the role of international achievement surveys in data-driven decision-making processes in federal and state education systems. The final section discusses the intended positive and unintended negative outcomes of cross-national achievement surveys for the national education systems and their practices, as well as for political reforms.

Introduction

Cross-national achievement surveys administered by the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) have attracted considerable attention and made a substantial impact on Germany ever since Programme for International Student Assessment (PISA) 2000. The following section provides a brief overview of the structure, organisation and institutional features of the compulsory school system in Germany.

In Germany, both school-based and university education is within the legislative and administrative mandate of the 16 federal states. Thus, there are 16 educational systems. The 16 states coordinate their work through the Standing Conference of the Ministers of Education and Cultural Affairs of the *Länder* in the Federal Republic of Germany (Kultusministerkonferenz (KMK)) and have started a process to implement shared *Bildungsstandards* (educational standards of the KMK) since 2003.

Overall, there are three core phases of schooling: *Grundschule* (primary education), *Sekundarstufe I* (secondary level I) and *Sekundarstufe II* (secondary level II; see also Table 2.1). Compulsory primary education typically starts at age 6 and lasts for 4 years (grades 1–4) in 14 of the 16 federal states. Consequently, students finish primary school, on average, at age 10. This 4-year primary school system dates back to 1920 and the Weimar Republic and was continued in West Germany after the Second World War, whereas East Germany introduced an 8-year comprehensive school system with the option of an additional 4 years, which was in place until 1990. Two states (Berlin, Brandenburg) have 6-year primary schools. There are currently 15 431 primary schools in Germany (school year 2019/2020; Statistisches Bundesamt, 2020). The core subjects in primary schools are German, maths, science and a first foreign language (mostly English). Typically, primary schooling is organised into age-homogeneous grades, with students grouped into classes of 21 students on average (Statistisches Bundesamt, 2018). Some states allow for cross-grade grouping within classes. Many primary schools have introduced full-day schooling, mostly on a voluntary basis for students. Students are split after fourth or sixth grade based on their academic achievement and predicted academic potential into several different, ostensibly academically homogeneous school tracks.

Sekundarstufe I typically covers grades 5–10, with deviations ranging from starting in grade 7 to ending after grade 9. A great variety of school tracks exist within and across the 16 federal states. Historically, after many decades of a traditional three-tiered system (*Hauptschule* as the lowest track ending in 9th or 10th grade; *Realschule* ending in grade 10; *Gymnasium* ending in grade 12 or 13), many federal states first added a comprehensive school track starting in the 1970s. Comprehensive schools themselves were designed in multiple versions: integrative (all students together with internal tracking for specific subjects) or cooperative (multiple separate tracks), as well as comprehensive schools with or without the option for students to acquire the *Abitur*, the highest form of school-leaving certificate that allows students to directly begin university studies without any further qualification measures.

In recent years, a tendency across the state-level systems, with states now offering mainly two types of school tracks – one leading to the *Abitur* and one not leading to the *Abitur*. Generally speaking, the wide variety of school tracks in Germany can be bifurcated into school types offering only one form of diploma and school types offering multiple forms of diplomas. *Sekundarstufe I* ends with diplomas that can be broadly classified into two categories: *Hauptschulabschluss*, which usually leads to students leaving school and starting an apprenticeship for an occupation, or *Realschulabschluss*, which allows students to either transition to apprenticeships for an occupation or continue their schooling and acquire an *Abitur*.

Table 2.1: Schools in Germany – types and numbers

Primary schools	15 431
<i>Gymnasium</i>	3 141
Special needs schools	2 819
Integrated comprehensive schools	2 130
<i>Hauptschulen</i> (lowest track schools)	1 915
Schools with multiple tracks	1 867
<i>Realschulen</i> (middle track schools)	1 781
Pre-classes, school-based kindergarten	1 199
Orientation schools with all track levels	1 063
Other or unidentified type	986

Source: Statistisches Bundesamt (2020).

Sekundarstufe II caters to students who want to attain a school-leaving certificate allowing them to begin university studies directly after finishing school (*Abitur*). *Sekundarstufe II* ends in some states after grade 12 and in other states after grade 13. The end point sometimes even depends on the kind of school within a state that students attend (e.g. in the state of Berlin, *Gymnasium* students receive their *Abitur* after 12th grade, whereas students at an *Integrierte Sekundarschule*, a kind of comprehensive school, finish after grade 13. Moreover, multiple types of schools exist as part of or are related to the German vocational education system.

Independently of the school track embarked on, 40.2 % of the population age cohort of students who left or had already left school in 2019 received an *Abitur* (Statista, 2020).

Parallel to its regular school systems, Germany maintains a set of schools catering to special needs students, which have been the subject of intense debate in recent years, with strong calls and attempts to increase the integration of special needs students into regular school classes. These attempts have yielded mixed results, with a higher number of students being integrated into regular schools, yet simultaneously a substantial increase in the number of students diagnosed as having special needs and thus stable high numbers of students in special needs schools. All in all, 91 % of students in Germany in the school year 2016/2017 attended public schools (Statistisches Bundesamt, 2018).

In summary, the major characteristics of the educational system in Germany are as follows: (i) the federal structure of 16 states maintaining 16 educational systems with a variety of different school types, (ii) the short duration of primary education compared with other countries, and (iii) the separate set of schools for children and young people with special needs.

International achievement surveys

Germany's participation in international achievement studies

Cross-national achievement surveys have been widely used to monitor educational achievement in Germany in the past two decades. Previously, Germany had participated in international student assessments by the IEA with only a limited scope: in 1964, two German federal states took part in the First International Mathematics Study and joined parts of the Six Subject Survey (English; civics); and in 1971, students from 10 federal states took part in the First International Science Study. In 1990/1991, 9- and 14-year-old Germans from federal states in both West Germany and East Germany were included in the International Study of Reading Literacy. During the same period, 10- and 13-year-olds in nine federal states were tested as part of the 1989 and 1992 Computers in Education Study (followed later by the Second Information Technology in Education Study). Subsequently, Germany participated in the Trends in International Mathematics and Science Study (TIMSS) from 1994 to 1996 with representative samples of 13-year-olds and young adults. At the end of the millennium, Germany joined the Civic Education Study, which included representative samples of 14-year-olds in 28 countries.

Starting in 2000, Germany has participated every 3 years in the PISA study focusing on reading, mathematics and science skills among 15-year-olds (see Table 2.2 for an overview of current participation). At primary school level, Germany began participating in 2001 in the Progress in International Reading Literacy Study (PIRLS), which takes place every 5 years. Germany has also continued to participate in the regular waves of TIMSS (only fourth graders) and began participating in the IEA's International Computer and Information Literacy Study (ICILS) in 2013. Furthermore, Germany has also joined (i) the OECD large-scale assessment Programme for the International Assessment of Adult Competencies (PIAAC), which focuses on adult competencies in literacy, numeracy and problem-solving in technology-rich environments (since 2011); (ii) the Teacher Education and Development Study in Mathematics (2008; financed by a grant from the German Research Foundation to Humboldt University of Berlin), which focuses on future math teachers; and (iii) the Teaching and Learning International Survey (2018; financed by a grant from the Leibniz Society).

Table 2.2: Germany's participation in international comparison studies

Study	Abbr- viation	International research coordinator	National research coordinator	Interval	Last assessment (and previous assessments)	Next assessment	Main focus of assessment	Main target group	Typical number of participants
Progress in International Reading Literacy Study	PIRLS	IEA	TU Dortmund University	Every 5 years	2021 (2001, 2006, 2011, 2016)	2026	Reading	Fourth graders	~ 4 000 students in ~ 200 classes
Trends in Mathematics and Science Study	TIMSS	IEA	University of Hamburg	Every 4 years	2019 (1995, 2007, 2011, 2015)	2022	Mathematics, science	Fourth graders	~ 4 000 students in ~ 200 classes
Programme for International Student Assessment	PISA	OECD	Centre for International Student Assessment	Every 3 years	2018 (2000, 2003, 2006, 2009, 2012, 2015)	2022	Reading, mathematics, science	15-year- olds	~ 7 500 (+ 9 000) students in ~ 250 schools
International Computer and Information Literacy Study	ICILS	IEA	Paderborn University	Every 5 years	2018 (2013)	2023	Computer and information literacy	Eighth graders	~ 3 000 students in ~ 150 schools
Programme for the International Assessment of Adult Competencies	PIAAC	OECD	GESIS Leibniz Institute for the Social Sciences	Every 10 years	2011/2012	2022/2023	Literacy, numeracy, problem- solving in technology-rich environments	16- to 65-year- olds	~ 5 000 people

Source: Adapted from McElvany and Stang (2020).

Germany's performance pattern in international achievement studies: focus on the Progress in International Reading Literacy Study

Reading literacy is significant for success in academic, workplace and everyday situations (Becker et al., 2010). It includes the ability to extract relevant information from texts and to understand, use and reflect on written texts (Mullis et al., 2015). The international large-scale comparison study PIRLS has monitored trends in fourth-graders' reading literacy every 5 years since 2001. During the assessment, each student reads two texts – one literary text and one informational text – and completes 12–15 comprehension questions. In addition, students answer questions about their motivation to read and attitudes towards reading and about perceived instructional quality. Furthermore, students' parents, teachers and school principals fill out questionnaires, and provide information about students' reading comprehension as well as their school and family background (McElvany and Stang, 2020).

The general performance pattern stemming from international comparisons for PIRLS 2016 indicated that German fourth graders had a mean reading literacy score of 537 points (Bos et al., 2017; Bremerich-Vos et al., 2017). At the same time, the dispersion of performance in Germany was very high. The difference between the 5th and 95th percentiles was 257 points. In Germany, 11 % of children achieve proficiency level V, whereas just under 6 % of children achieve proficiency level I. Just under 19 % of children in Germany fell below proficiency level III, and thus were highly likely to have trouble succeeding in their future education.

When examining the factors affecting Germany's educational results, two issues have drawn particular attention in recent years, both referring to structural family characteristics: the socioeconomic and educational backgrounds of students' families and whether or not students have an immigrant background. The PIRLS 2016 results indicated that, in Germany, children from families with more than 100 books at home score 54 points higher than children from families with a maximum of 100 books at home (Hußmann et al., 2017). Even greater social disparities exist between children with parents belonging to different occupational groups. There are also very large differences between children from families in which at least one parent has a tertiary education and children from families with no experience of tertiary education. Children who always or almost always speak German at home have a mean score of 549 points, and thus score 40 points above children who never or almost never speak German at home (Wendt and Schwippert, 2017). Children with two parents born in Germany score 24 points higher than children with one parent born abroad and 49 points higher than children with two parents born abroad.

Turning to individual determinants, both gender and – especially for reading literacy – motivation have attracted attention with respect to the large-scale assessment results: in Germany, girls score 543 points on average, 11 points higher than boys (McElvany et al., 2017). Institutional determinants have also been assessed but have generated a somewhat lower level of attention outside the research community. One reason may be that the instructional variables used in cross-national achievement studies are rather broad and sometimes lack in-depth specificity regarding concrete classroom instruction

in the domain under investigation. However, one major institutional characteristic has been intensely discussed, particularly following the results of PISA 2000: school tracks. It has become empirically clear that (i) substantial average performance differences exist between the different types of schools that students can attend after primary school, while at the same time (ii) major overlaps in students' performance levels occur, (iii) students from certain groups (e.g. students from low socioeconomic status families and/or with an immigrant background) are over-represented in lower-track schools, and (iv) upward mobility between school tracks is low. This discussion has sparked changes in many German states, often with the intention of consolidating the tracked system and giving students more opportunities to decide later in their educational careers which school-leaving certificate is most suitable for them.

Germany's performance in international comparison: focus on the Progress in International Reading Literacy Study

The results of PIRLS 2016 placed Germany in the middle of the pack, with 537 points (Bos et al., 2017; Bremerich-Vos et al., 2017). Both participating European Union Member States and OECD countries had average scores of 540 points, which was not statistically significantly different from Germany's average performance. The difference of about 40 points between Germany and top-scoring countries such as Russia (581 points) and Singapore (576 points) was considerable and corresponded to almost 1 year of learning. Germany's performance dispersion was among the highest in the study. With 11 % of children at proficiency level V, Germany was roughly in line with the EU and OECD average. However, many European countries achieve percentages approaching 20 %. Hardly any country in Europe had a proportion of children at proficiency level I higher than 6 % and thus greater than that in Germany. At 19 %, the proportion of children in Germany who did not reach proficiency level III was also very high by international standards.

Examining differences among various subgroups of students, the performance gap favouring girls in Germany was about the same as that in the EU Member States and OECD countries as a whole (McElvany et al., 2017). All three indicators used internationally to measure social disparities (number of books at home; parents' occupational status; parents' educational level) showed that Germany was one of the countries where social disparities were most pronounced (Hußmann et al., 2017). In almost all participating countries, children who did not speak the test language at home scored lower. However, the magnitude of the difference varied widely across countries, and migration-related disadvantages were larger in Germany than in most other countries (Wendt and Schwippert, 2017).

Germany's performance patterns since initial administration of international achievement surveys: focus on the Progress in International Reading Literacy Study

Examining the performance patterns of German fourth graders in PIRLS since its initial administration in 2001 yields a clear result (Bos et al., 2017; Bremerich-Vos et al., 2017). In 2001, Germany had a mean value of 539 points. The difference of 2 points between 2001 and 2016 was not statistically significant. Thus, the average reading literacy of primary school students in Germany has not increased over the past 15 years. There

was an 'intermediate high' in 2006 (548 points), but this was temporary, and in 2011 the average score fell back down to 541 points. In contrast, other countries, including some EU Member States, have been able to improve their scores from 2001 to 2016. Thus, Germany's relative position has deteriorated considerably.

Performance dispersion increased in Germany from 2001 to 2016. The proportion of children at proficiency level V increased from 9 % to 11 % from 2001 to 2016, and the proportion of children at proficiency level I increased from 3 % to 6 % from 2001 to 2016. The percentage of children not achieving proficiency level III increased slightly, from 17 % in 2001 to 19 % in 2016. Germany is one of the few countries in which score dispersion worsened from 2001 to 2016.

Turning to specific subgroups, gender disparities in reading literacy in Germany in favour of girls remained constant from 2001 to 2016 (McElvany et al., 2017). In contrast, social disparities in Germany have widened significantly since 2001 (Hußmann et al., 2017). The country has not been able to achieve equality of opportunity or even to narrow the gap. However, the level of social disparities differs between and within the federal states. In the same vein, Germany was not able to reduce migration-related disparities from 2001 to 2016 (Wendt et al., 2017).

Germany's participation and performance in the Trends in International Mathematics and Science Study and the Programme for International Student Assessment

In the assessment of fourth graders' mathematical literacy in TIMSS, Germany had a mean score of 521 points in 2019, with about one quarter of children not reaching proficiency level III (Schwippert et al., 2019; Selter et al., 2019). Germany was statistically significantly above the international average but significantly below the EU and OECD averages. Although in 2007 and 2011 Germany was still in the top third, it had fallen to the middle of the pack by 2015 and 2019. Performance dispersion was rather low compared with other countries but increased slightly over time, as did the percentage of children who did not reach proficiency level III. Overall, mathematical literacy in 2019 was not statistically significantly different from that in 2015 and 2007 but was significantly lower than the 2011 results (Kasper et al., 2019).

Turning to the natural science assessment in TIMSS, Germany had a mean score of 518 points in 2019, with about one quarter of children not reaching proficiency level III (Steffensky et al., 2019). Therefore, German fourth graders, on average, scored statistically significantly above the international mean but significantly below the OECD average; in addition, their score was not significantly different from the EU average. Although mean scores in most countries were stable from TIMSS 2015 to TIMSS 2019, Germany was among the small set of countries in which performance deteriorated over this period. Science proficiency remained steady at 528 points from 2007 to 2015. In 2019, it fell to just 519 points (Kasper et al., 2019). The percentage of children who did not reach proficiency level III increased slightly. Germany's performance in 2019 was slightly less homogeneous than in previous cycles, whereas the dispersion in most participating countries has remained relatively stable.

To describe performance patterns in international achievement surveys for older students, we turn to PISA. Germany's 15-year-olds had a mean score of 498 points in 2018, with 20.7 % at the lowest level of competence (Weis et al., 2019). This average score was statistically significantly higher than the OECD average of 487 points. Compared with other OECD countries, Germany had a relatively high proportion of young people with particularly strong reading skills. At the same time, the dispersion of performance in Germany was particularly high in international comparison and greater in 2019 than in 2009 and 2015. Reading performance in 2018 was not significantly higher than reading performance in 2009 and 2015, but was significantly higher than reading performance in 2000, when Germany's low performance attracted considerable national attention.

Similarly, in 2018, Germany's mathematics score among 15-year-olds of 500 points was statistically significantly above the OECD average of 489 points (Reinhold et al., 2019). Still, 21.1 % of young people were at the lowest level of competence, and performance dispersion in Germany was significantly above the OECD average. Students' mathematics skills declined slightly in 2018 compared with 2012 (the last cycle with mathematics as the main domain).

Finally, turning to natural science, Germany had a mean score of 503 points in 2018, which was statistically significantly above the OECD average (Schiepe-Tiska et al., 2019). Nevertheless, 19.6 % of young people were at the lowest level of competence, and performance dispersion in Germany was significantly above the OECD average. Moreover, students' science skills significantly decreased in 2018 compared with 2006 (the first cycle with science as the main domain).

Education policy monitoring

Germany's merely average performance in TIMSS in the 1990s had already raised awareness of the need for both empirical assessments of educational outcomes and improved teaching of maths and science in German schools. This resulted in a comprehensive programme to improve students' maths and science skills. Nevertheless, the real breakthrough came following the tremendous public outcry after the publication of the PISA 2000 results, in which German 15-year-olds' reading literacy performance was well below expectations. This result, which was surprising for most stakeholders and the public, considerably changed Germany's relationship with international student comparison studies. Since PISA 2000, such results continue to be broadly reported and discussed, even though the extreme attention they drew at the beginning of the millennium has dissipated. However, the original perception that the substantial relationship between family background characteristics and performance level is a source of concern and should be actively counteracted by educational policy measures has somewhat weakened, and this relationship is now accepted in some corners as an unfortunate yet common result of such studies. An important focus of attention in recent years, even before the COVID-19 pandemic resulted in school closures, has been the digitalisation of the German education system. The 2013 and 2018 waves of ICILS pointed to a substantial need for improvement in Germany (Eickelmann et al., 2019).

Germany's current educational monitoring processes are based largely on three core documents that define the current German perspective and were put into place starting with the 1997 *Konstanzer Beschluss* by the KMK of all federal states, in which it was decided to develop and assess measures to evaluate and improve the quality of the educational system. Educational policymakers declared that they would use empirical data from educational research to identify strengths and weaknesses of the educational systems in the 16 federal states, focusing on secondary schools (grades 9 and 10) and competencies in (German) language, arts, mathematics, science and foreign languages, as well as personal and social skills. In 2006, a general strategy for monitoring the German educational system was agreed on (known as the *Plöner Beschluss*), which included regular participation in national and international large-scale comparison studies on student achievement. Core studies and instruments were identified as a shared basis for evidence-based educational governance focused on the results (outputs) of educational processes. In 2015, the overall strategy was updated, emphasising the need for explanatory in addition to descriptive knowledge and identifying core areas of interest for further evidence to be used in educational policy and practice.

The four core pillars agreed on by all 16 German federal states are as follows (see McElvany and Stang, 2020):

- (1) participation in international large-scale assessments (PIRLS, PISA, TIMSS (primary school));
- (2) evaluation and implementation of educational standards (*Bildungsstandards*) – national assessments that allow for comparisons across the federal states and evaluate whether students meet the educational standards defined for specific subjects in specific grades, focusing on end of primary, secondary and continued secondary education, with centralised tests at the end of the first two phases and the selection of exam exercises from a central pool for the highest form of school-leaving certificate that if successfully passed allows students access to university studies;
- (3) quality assurance at school level – state-specific and cross-state assessments to compare the performance of individual schools and classes with the purpose of supporting instruction and school development (e.g. *Vergleichsarbeiten*);
- (4) publication of a comprehensive national report by the Federal Ministry of Education and Research and all German states on the status of the educational system every 2 years.

The three aforementioned international achievement studies, PIRLS and TIMSS (primary school) and PISA (secondary school), are implemented by consortia under the scientific directorship of the Center for Research on Education and School Development at TU Dortmund University (PIRLS), the Research Group for Evaluation of Educational Systems at the University of Hamburg (TIMSS) and the Centre for International Student Assessment (PISA), a collaboration between three institutes.

For national educational monitoring, in 2004, the federal states jointly founded the Institute for Quality Development in Education in Berlin, a scientific institution that has since been entrusted with operationalising the reviewing of the educational standards

decided on by the KMK as well as with coordinating standards-oriented task development and evaluation studies, known as *Bildungstrend* (KMK, 2016). The educational standards refer to general educational objectives and specify which competencies, with regard to key content areas, students should have acquired by a certain grade. The first evaluation phase took place from 2009 to 2012, the second phase took place from 2012 to 2015, and the third cycle of the institute's *Bildungstrend* studies began in 2021, examining fourth graders in the 16 states, and will continue in 2022 with ninth grade students. Apart from these shared efforts, the 16 states also have individual measures for internal and external school monitoring, including school inspection visits in many states. In addition, an interdisciplinary consortium for implementing a National Educational Panel Study was founded under the leadership of the Leibniz Institute for Educational Trajectories and originally initiated and funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) ⁽⁵⁾, with the goals of gathering extensive empirical evidence about how education is acquired, understanding how it affects individual biographies, and describing and analysing major educational processes and trajectories across the lifespan. A large starting cohort of about 20 000 students in grade 5, along with their parents, teachers and school principals, will be followed, with yearly assessments beginning in 2022.

To summarise, in Germany, international achievement surveys play a major role in educational monitoring and are closely linked to national measures informing education policy development and reform. International comparison studies are often complemented by national components, increasing the number of content areas covered and often also the number of students involved. A current example is PIRLS 2021. Due to the transition to digital reading assessment, the sample was substantially increased (with subsamples being assessed on digital reading, paper-based reading and reading in the online environment). National content additions included both additional skills assessments (e.g. vocabulary) and questionnaires (e.g. on students' well-being, instructional quality, digitalisation). Furthermore, an additional sample for PIRLS 2021 was drawn from schools that had participated in PIRLS 2016. This will allow for the first-ever analysis of developments at school level over a period of 5 years and will provide important insights into changes over time and their antecedents.

Analysis of policies, practices and political reforms

Germany's results in international assessments have varied in recent years depending on the age group and domain under investigation. The country's extensive participation in international comparison studies and public reception and discussion of the country's outcomes in the past two decades have led to various intended positive and some unintended negative consequences. As previously mentioned, the empirical evidence pointing to less than desirable performance outcomes and unfortunate correlations between family background variables and educational success in Germany has led to a

⁽⁵⁾ <https://www.neps-data.de/Mainpage>

variety of policies, practices and political reforms. These include, but are not limited to, the following.

- At a political level, the 16 federal states had intense discussions and decided on how to design a common form of educational quality monitoring in Germany. In the end, all federal states agreed on a shared overall strategy for educational monitoring with international, national and state assessments, as well as continuous reporting. These efforts have recently been further advanced by the establishment of an advisory board of education researchers.
- Reforms have been implemented addressing the types of secondary schools existing in multiple states and the option of including centrally provided exercises in final exams for school-leaving certificates.
- There have been impacts on educational practice through, for example, focusing attention on students' language skills as regards the classroom content, as well as assessment of language skills and research on effective support measures (e.g. the programme *Bildung durch Sprache und Schrift* ('Education through language and writing')).
- Generous funding has been provided for empirical educational research, for example by the Federal Ministry of Education and Research through its *Rahmenprogramm empirische Bildungsforschung* (framework programme for empirical educational research), which now emphasises (i) increasing educational justice by identifying and developing all individuals' potential, (ii) dealing with diversity and strengthening societal cohesion, (iii) supporting quality in the educational system, and (iv) designing and using technological developments in education (BMBF, 2018).
- There was a structural change in universities establishing numerous professorships in the area of empirical educational research, along with a substantial shift in the discipline towards empirical, and largely quantitative, research. This development also led to the founding of an interdisciplinary Society for Empirical Education Research in 2012 as well as a new interdisciplinary open access online journal (*Journal for Educational Research Online*) in 2009.

At the same time, critical issues must also be taken into consideration. Regarding areas of concern, McElvany and Stang (2020) pointed out the focus on a few core domains, and potential negative consequences for other subjects, such as the arts and history, as well as for other important outcomes of education, such as children's social skills or emotional well-being, when it comes to appreciation, attention, effort, further development and funding. This goes hand in hand with the more general question of the extent to which international comparison studies push a unilaterally utilitarian view of education in contrast to education primarily for the purpose of human development. In terms of instruction, as in many other countries, worries exist that teaching to the test is becoming a strategy among schools and teachers. In the area of research, there are valid concerns that educational research itself has been mainstreamed into a discipline serving educational administration, with research funding and positions awarded only to researchers and research closely linked to political interests regarding desired knowledge and governing priorities (for a critical discussion on the relationship between

trust and a knowledge-based educational system, see Bormann and John, 2014). Last but not least, the already substantial costs related to international assessments limit educational policymakers' willingness to additionally fund research accompanying the studies when the data sets are available after the presentation of the results.

Turning to the current situation and the question of impact, overall the perception prevails that a lot has happened and been implemented but that important goals such as (i) increasing the average results, (ii) enlarging the group of high performers, (iii) reducing the group of low performers, (iv) reducing the correlation between family socioeconomic background and performance, and (v) more effectively supporting students from immigrant backgrounds have not yet been achieved to a satisfactory level (McElvany and Stang, 2020). Recently, the unsatisfactory state of digitalisation within the German education system – both in terms of technologies and in terms of content and pedagogical concepts and implementation – has been increasingly discussed as a challenge for German educational practice, administration, politics and research. In addition, the need for empirical evidence on effective means to improve educational quality is becoming a greater area of focus.

In summary, core consequences of Germany's participation in international large-scale assessment studies are the shift from an input to an output orientation and the subsequent continuous evaluation of the educational system and outcomes based on quantitative empirical evidence. Furthermore, evidence-based thinking has changed teacher education, instructional approaches and school development in many ways (Sälzer and Prenzel, 2018). The excessive attention paid to the outcomes of international assessment studies has somewhat dissipated in recent years, but a key conclusion is that they can be expected to continue to be an important pillar of the educational quality monitoring system in Germany for the foreseeable future.

REFERENCES

- Becker, M., McElvany, N. and Kortenbruck, M. (2010), 'Intrinsic and extrinsic reading motivation as predictors of reading literacy: a longitudinal study', *Journal of Educational Psychology*, Vol. 102, pp. 773–785.
- BMBF (Federal Ministry of Education and Research) (2018), *Framework Programme for Empirical Educational Research*, BMBF, Bonn (https://www.empirische-bildungsforschung-bmbf.de/media/content/BMBF_287_RP_Bildungsforschung_engl_barrierefrei.pdf).
- Bormann, I. and John, R. (2014), 'Trust in the education system – thoughts on a fragile bridge into the future', *European Journal of Futures Research*, Vol. 2, No 35.
- Bos, W., Valtin, R., Hußmann, A., Wendt, H. and Goy, M. (2017), 'IGLU 2016: Wichtige Ergebnisse im Überblick', in Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (eds), *IGLU 2016 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 13–28.
- Bremerich-Vos, A., Wendt, H. and Bos, W. (2017), 'Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse', in Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (eds), *IGLU 2016 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 79–142.
- Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K. et al. (eds) (2019), *ICILS 2018 #Deutschland – Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking*, Waxmann, Münster.
- Hußmann, A., Stubbe, T. C. and Kasper, D. (2017), 'Soziale Herkunft und Lesekompetenzen von Schülerinnen und Schülern', in Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (eds), *IGLU 2016 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 195–217.
- Kasper, D., Wendt, H., Schwippert, K. and Köller, O. (2019), 'Trends in Schülerzusammensetzungen und mathematischen und naturwissenschaftlichen Kompetenzen', in Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. and Wendt, H. (eds), *TIMSS 2019 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 331–349.
- KMK (Kultusministerkonferenz) (2016), *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*, Wolters Kluwer, Berlin/Bonn (https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf).
- McElvany, N. and Stang, J. (2020), 'Educational assessment in Germany', in Harju-Luukkainen, H., McElvany, N. and Stang, J. (eds), *Monitoring of Student Achievement in the 21st Century – European policy perspectives and assessment strategies*, Springer, Cham, Switzerland, pp. 143–156.
- McElvany, N., Kessels, U., Schwabe, F. and Kasper, D. (2017), 'Geschlecht und Lesekompetenz', in Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (eds), *IGLU 2016 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 177–194.

- Mullis, I. V. S., Martin, M. O. and Sainsbury, M. (2015), 'PIRLS 2016 reading framework', in Mullis, I. V. S. and Martin, M. O. (eds), *PIRLS 2016 Assessment Framework – 2nd edition*, TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill, MA, pp. 11–29.
- Reinhold, F., Reiss, K., Diedrich, J., Hofer, S. and Heinze, A. (2019), 'Mathematische Kompetenzen in PISA 2018 – aktueller Stand und Entwicklung', in Reiss, K., Weis, M., Klieme, E. and Köller, O. (eds), *PISA 2018 – Grundbildung im internationalen Vergleich*, Waxmann, Münster, pp. 187–209.
- Sälzer, C. and Prenzel, M. (2018), 'Policy implications of PISA in Germany: the case of teacher education', in Volante, L. (ed.), *The PISA Effect on Global Educational Governance*, Routledge, London, pp. 109–126.
- Schiepe-Tiska, A., Rönnebeck, S. and Neumann, K. (2019), 'Naturwissenschaftliche Kompetenzen in PISA 2018 – aktueller Stand, Veränderungen und Implikationen für die naturwissenschaftliche Bildung in Deutschland', in Reiss, K., Weis, M., Klieme, E. and Köller, O. (eds), *PISA 2018 – Grundbildung im internationalen Vergleich*, Waxmann, Münster, pp. 211–240.
- Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. and Wendt, H. (2019), 'TIMSS 2019: Wichtige Ergebnisse im Überblick', in Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. and Wendt, H. (eds), *TIMSS 2019 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 13–24.
- Selter, C., Walter, D., Heinze, A., Brandt, J. and Jentsch, A. (2019), 'Mathematische Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse', in Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. and Wendt, H. (eds), *TIMSS 2019 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 57–113.
- Statista (2020), *Anteil der Schulabsolventen/-innen mit allg – Hochschulreife in Deutschland bis 2019* (<https://de.statista.com/statistik/daten/studie/440781/umfrage/anteil-der-schulabsolventen-innen-mit-allgemeiner-hochschulreife-in-deutschland/>).
- Statistisches Bundesamt (2018), *Schulen auf einen Blick* (https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00035140/Schulen_auf_einen_Blick_2018_Web_bf.pdf).
- Statistisches Bundesamt (2020), 'Allgemeinbildende und berufliche Schulen' (https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Schulen/_inhalt.html).
- Steffensky, M., Scholz, L. A., Kasper, D. and Köller, O. (2019), 'Naturwissenschaftliche Kompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse', in Schwippert, K., Kasper, D., Köller, O., McElvany, N., Selter, C., Steffensky, M. and Wendt, H. (eds), *TIMSS 2019 – Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 115–168.
- Weis, M., Doroganova, A., Hahnel, C., Becker-Mrotzek, M., Lindauer, T., Artelt, C. and Reiss, K. (2019), 'Lesekompetenz in PISA 2018 – Ergebnisse in einer digitalen Welt', in Reiss, K., Weis, M., Klieme, E. and Köller, O. (eds), *PISA 2018 – Grundbildung im internationalen Vergleich*, Waxmann, Münster, pp. 47–80.
- Wendt, H. and Schwippert, K. (2017), 'Lesekompetenzen von Schülerinnen und Schülern mit und ohne Migrationshintergrund', in Hußmann, A., Wendt, H., Bos, W., Bremerich-Vos, A., Kasper, D., Lankes, E.-M. et al. (eds), *IGLU 2016 – Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Waxmann, Münster, pp. 219–234.

CHAPTER 3

Cross-national achievement surveys and educational monitoring in France

Romuald Normand and Bénédicte Gendron

Abstract

In line with their republican vision, French national education policies aim to reduce inequalities between students. Based on international survey results, this objective seems quite difficult to achieve within the school system. Despite successive and multiple reforms, an inequality gap remains and few improvements are noticeable at local level. In this chapter, educational monitoring in France and the use of international/national data to implement reforms are analysed. Some concepts borrowed from policy studies help to characterise the main changes but also illustrate unintended and sometimes adverse consequences. The chapter shows that the French Ministry of National Education has used different policy instruments to support a basic skills and soft accountability policy. However, the most recent large-scale reforms face limitations in their scope and extent due to implementation gaps that are explained from an international research perspective.

Introduction

As its name suggests, the Ministry of National Education, Youth and Sports (MNE) in France oversees a highly centralised public education system. Since the 2000s, France has been strongly committed to a focus on basic skills in education while using ‘soft’ accountability policies to monitor the education system. Simultaneously, international survey data have become increasingly valued by policymakers, and these international data have shown that French students’ scores in international surveys have steadily declined over the years. This chapter presents the French education system and its governance and compares some key scores attained by French students in international surveys reported by official international and national reports. It provides an overview of policy instruments implemented to support a basic skills and soft accountability policy and describes some unintended and adverse impacts on the past decade of large-scale reforms, acts and policies and strategic governance in the light of international research findings on school change and improvement.

The governance of the French education system

The French education system has become highly bureaucratic and centralised, as responsibilities and powers, including for the design of curricula, national assessments, initial and continuous teacher training, and the selection and recruitment of teachers, school principals and inspectors, have been concentrated in the current MNE (Normand, 2020a). Its technostructure is composed of different ministerial directorates, including the Directorate of the School System (Direction Générale de l’Enseignement Scolaire), which develops targeted programmes in compulsory education, and the Directorate of Evaluation, Forecast and Performance (Direction de l’Evaluation, de la Prospective et de la Performance (DEPP)), created in 1986, which compiles statistics and indicators to help the MNE monitor the education system and also implement national assessments of student skills at grades 1, 2 and 6.

Through its multiyear programming policies, the MNE has established general objectives for the education system heavily based on official decrees and instructions, which are often published weekly in the *Bulletin Officiel de l’Education Nationale* (the official journal of the MNE). Education is governed by the Civil Service Code, which outlines the responsibilities of civil servants, including educators, as well as by the Education Code. Combined, these bring together the relevant laws and regulations enacted by legislative bodies that oversee education in France.

A centralised and hierarchical educational system

Although decentralising principles have been integrated into the French constitution, the hierarchical centralisation system within the country explains why local independent educational authorities have little authority in the governance of education. They tend to be limited to the maintenance of school buildings and equipment, along with youth professional training systems and policies. Although the French education system is highly bureaucratic, it is also corporatist, with powerful trade unions and interest groups; however, superintendents (*recteurs*) have more power in implementing national reforms. Teachers, as a professional body defended by their unions, have a certain degree of autonomy. As a result, the monitoring developed by the MNE's technostructure entails a significant degree of inertia and slowness.

In the French tradition of the Enlightenment, the transmission of knowledge and curricular content is an important political issue that is frequently at the forefront of public debate. The Minister for National Education often delegates matters relating to the national curriculum to the High Council of Curricula (Conseil supérieur des programmes) (Heurdiere and Clément, 2016) and appoints a chairman and council members composed of general inspectors, experts and educational specialists, and eminent members of society. Successive ministers have also created a National Council for Education Evaluation, which has led to the implementation of a soft accountability system based on national assessments and school value-added indicators, and to greater recognition of international surveys.

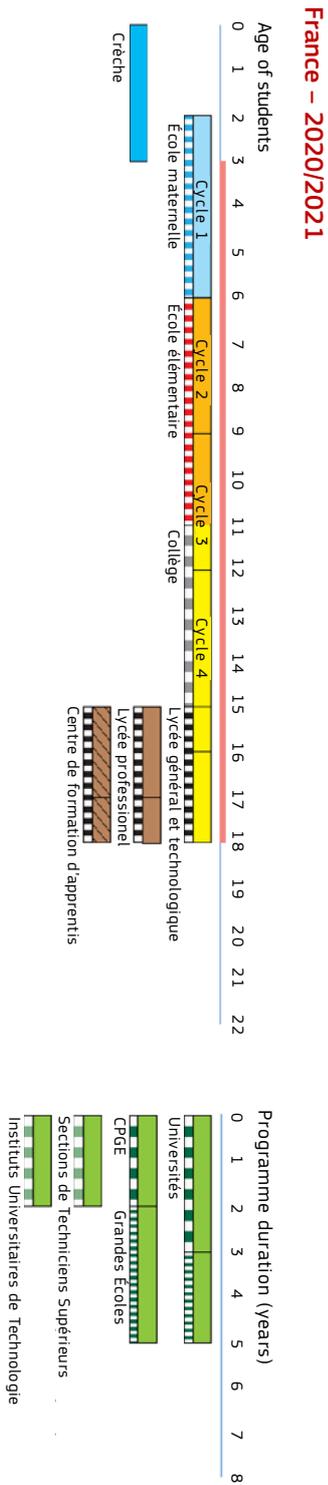
Limited decentralisation and school market

The French education system remains far from market mechanisms and school choice, unlike the system in some other countries. Education is first considered a national public service, secular, free and compulsory from 3 to 16 years old. The state is opposed to competition between schools, and there is little incentive for parents to move away from a given schooling area, despite increasing use of school choice strategies by middle- and upper-class families (van Zanten, 2015). The limited private sector is mainly driven by religious Catholic schools. The 1905 Act maintains a divide between the church and the government, as it affirms secularist principles.

A dual system of educational governance

French educational governance strongly separates primary and secondary education. Since the end of the 19th century, primary education including kindergarten and elementary schools has been governed locally by a *recteur*, representing the minister at regional level, who is in charge of primary school inspectors. These inspectors are responsible for supervising schools and inspecting teachers in their classrooms (Pons, 2015), since there is no elementary school principal (primary schools have a head teacher). Figure 3.1 provides a summary of the general structure of the French system.

Figure 3.1: The structure of the French education system



Note: ISCED 4 education covers less than 2 % of the total number of students (all levels). Since September 2020, training has become compulsory for students aged between 16 and 18. Young people will be able to fulfil this compulsory training by several means: schooling, apprenticeship, training courses, civic service, and support system or social and professional integration measures.

- Early childhood education and care (for which the MNE is not responsible)
- Early childhood education and care (for which the MNE is responsible)
- Primary education
- Secondary general education
- ISCED 0
- ISCED 1
- ISCED 2
- ISCED 3
- ISCED 5
- ISCED 6
- ISCED 7
- Secondary vocational education
- Tertiary education (full-time)
- Additional year
- ➔ Study abroad
- Combined school and workplace courses
- ➔ Programme being phased out during (year)
- Compulsory full-time education/training
- Compulsory part-time education/training
- Compulsory work experience and its duration

NB: CPE, Classes préparatoires aux Grandes Écoles; ISCED, International Standard Classification of Education.

Source: Eurydice (2021).

Secondary education, consisting of junior and senior high schools, is supervised by secondary education inspectors specialised in a particular academic discipline or in school auditing. Each secondary school has a principal, and larger schools have a deputy principal. All inspectors are also placed under the authority of the superintendent. The secondary territorial inspectors are supervised by the General Inspectorate, which is composed of high-ranking officials who assist the minister in supervising education policies, auditing and writing reports.

Strong professional segmentation and societal challenges

The French system is characterised by strong professional and hierarchical segmentation, which increases the bureaucratic burden of its overall governance, particularly within decentralised state authorities responsible for implementing policies ⁽⁶⁾. The coordination of these middle-ranking executives is mainly determined by student guidance policy and mechanisms focused on school learning cycles and guidance pathways. Indeed, most of those inspectors are mainly concerned with student mastery of basic skills at the end of compulsory education.

The national curriculum is supported by regular national assessments but without any high-stakes consequences for teachers. Several school councils (pedagogy councils, guidance councils, those that facilitate regular meetings between primary and lower secondary schools) have been designed to share best practices and facilitate pedagogical interactions between teachers. Schools are increasingly attentive to violence and the school climate, which are of special concern to the ‘school life service’ (*service de vie scolaire*) in secondary schools, a French peculiarity in which specific professionals are responsible for discipline and absenteeism, as well as citizenship education.

France and international achievement surveys: the decline in student results

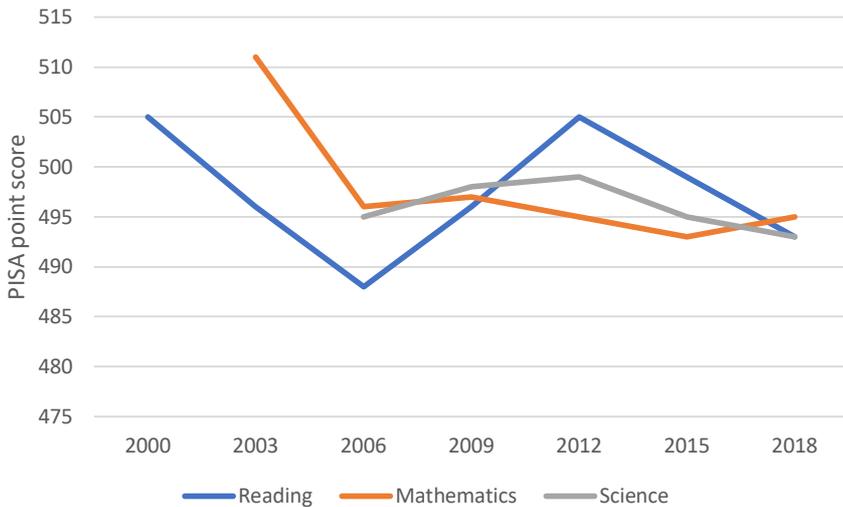
France has participated in international surveys since the end of the 1980s. Survey results are covered in the International Education Indicators Project and the first 1992 Organisation for Economic Co-operation and Development (OECD) publication *Education at a Glance* (Pons, 2011). The French MNE’s DEPP also published its first *State of the School System (L’ état de l’École)* review in 1992, which included 30 indicators describing students’ educational outcomes. Over the years, this publication has commonly questioned the French rankings reported in international surveys (even though there was no ‘PISA shock’ comparable with that in Germany) (Dobbins and Martens, 2012). The directorate also administers the Programme for International Student Assessment (PISA) and international surveys and ensures the MNE’s official publication of student results, which, with OECD reports have provided a general overview of France’s scores in reading, mathematics and science over the past decade.

⁽⁶⁾ State local authorities are referred to as *rectorats* (rectorships) and some deconcentrated administrative entities of the state are referred to as *directions académiques des services de l’éducation nationale* (directorates of national education services) to differentiate them from local authorities (municipalities, general councils, regional councils), which are independent of the state – another complicated patchwork of French policy and politics.

Programme for International Student Assessment scores and the very average French ranking

In reading – the major domain assessed in PISA 2018 – the average score for French students (493 points) placed France slightly above the OECD average (487 points) (DEPP, 2016a). Similar results were obtained in mathematics and science. As observed in previous PISA tests, France is one of the OECD countries in which socioeconomic status and performance are highly correlated, with a difference of 107 points between students from advantaged backgrounds and students from disadvantaged backgrounds. This gap is significantly higher than the average difference of 89 points in the OECD countries. Figure 3.2 shows French score trends in PISA’s three literacy domains.

Figure 3.2: Trends in French scores in PISA’s three literacy domains (reading, mathematics and science)



Source: OECD (2019a,b).

Reading: recurring difficulties

France ranks between 20th and 26th in reading literacy among PISA countries and between 15th and 21st among OECD countries, comparable with Belgium, Czechia, Germany, Portugal and Slovenia (DEPP, 2019a). Compared with the average of countries participating in PISA 2018 (Schleicher, 2018), French students appear to be slightly more comfortable with items asking them to ‘locate information’ in a text (496 points on average) than with items asking them to ‘understand’ (490 points on average) or ‘analyse and reflect on’ a text (491 points on average). French students can understand long texts, deal with abstract or counter-intuitive concepts, and make distinctions between facts and opinions based on implicit clues about the content or source of the information.

Mathematics: a French shame

In mathematics, France ranks between 15th and 24th among OECD countries, with an average performance. Only 1.8 % of French students have reached level 6, the highest level in PISA (OECD average: 2.4 %), which worries public authorities, even though mathematics remains a prestigious pathway and academic discipline in France, as exemplified by the Fields Medals won by French academics. French students scored 495 points in mathematics in the 2018 PISA test, slightly above the OECD average of 489 points (DEPP, 2019b). In total, 79 % of students reached at least level 2 on the mathematics test (OECD average: 76 %). At a minimum, these students can interpret and recognise, without direct instruction, how a (simple) situation can be represented mathematically (e.g. comparing the total distance between two alternative routes or converting prices into another currency). About 11 % of students performed very well in mathematics (at least level 5 on the mathematics test), a percentage very similar to the OECD average of 10.9 % but far behind the percentages of the combined Chinese provinces and municipalities of Beijing, Shanghai, Jiangsu and Zhejiang (slightly more than 44 %), Singapore (nearly 37 %), Hong Kong (China) (29 %) and Macao (China) (nearly 28 %) at the top of the rankings. French students face difficulties in modelling complex situations and in choosing, comparing and evaluating appropriate problem-solving strategies to deal with them.

Science: average performance

In scientific literacy, France ranks between 16th and 23rd among OECD countries (DEPP, 2016a). French students scored 493 points on PISA 2018, slightly higher than the OECD average (489 points). About four out of five students achieved level 2 or higher in science (78 %, compared with an OECD average of 79 %). At a minimum, these students can recognise the correct explanation of familiar scientific phenomena and use this knowledge to identify, in simple cases, whether a conclusion is valid on the basis of the data provided. Approximately 6.6 % of students are high achievers in science, which means that they have reached level 5 or 6 (OECD average: 6.8 %). Based on these results, French students therefore have difficulties in creatively and independently applying their scientific knowledge to a wide variety of situations, including those that are unknown to them. They lack creativity and thoughtfulness in relation to the experimental approach to science.

Average performance in reading in France has not changed significantly since the first edition of PISA in 2000 (DEPP, 2016b; DEPP, 2019a,b). Similarly, average performance in science did not change significantly between 2006 and 2018. Average scores in mathematics declined between 2003 and 2018. In reading comprehension, the apparent stability in performance between 2000 and 2018 period masks strong differences between students. Although the level of the best students tends to increase over time, the level of the weakest students has, in contrast, decreased. This widening gap is not observed in mathematics, in which the drop in performance over the period is observed for both the best and the weakest students, or in science. In France, students from an advantaged socioeconomic background (in the top quartile of the PISA index of economic, social and cultural status) scored 107 points higher than disadvantaged students (in the bottom quartile of the PISA index of economic, social and cultural status) in reading. This

is one of the largest socioeconomic achievement gaps among OECD countries (average gap: 89 points). The achievement gap is also widening for students with an immigrant background.

Trends in International Mathematics and Science Study results: a low ranking and French political authorities worrying about trends

In Trends in International Mathematics and Science Study (TIMSS) 2015, France had an average score of 483 points in mathematics, and was in last position and below the international average of 511 points across participating OECD countries and European Union Member States. French political authorities were particularly worried about this international ranking remaining unchanged for 4 years (DEPP, 2020). Indeed 12 % of French students scored below 400, illustrating that the basic knowledge level was not reached by a substantial portion of students. Only 2 % of students reached the advanced level, compared with 50 % in Singapore, South Korea and Taiwan and 11 % on average in the EU. Among mathematical topics, statistics and probability (496 points) and geometry (493 points) were the most likely to be mastered by French students, compared with the overall score of 483 points. French students showed a relatively low level of self-confidence and motivation in relation to learning mathematics compared with the average for participating OECD countries and EU Member States.

Progress in International Reading Literacy Study results: a widening gap in reading

In the 2016 Progress in International Reading Literacy Study (PIRLS) survey, French students' average score was 511 (DEPP, 2017). In total, 6 % of French students did not reach the most basic level of PIRLS, compared with 4 % across the EU. On average, 12 % of students in the EU reached the advanced level, compared with 4 % in France. France's overall average score decreased from 525 in 2001 to 511 in 2016 despite French students spending a high number of hours in language classes: 288 hours per year according to the 2008 survey, compared with 236 average hours in other EU Member States. However, fewer French teachers than their European counterparts reported providing their students with weekly activities that were likely to develop their strategies and skills in reading. When asked about their participation in professional development in reading over the past 2 years (courses, workshops, seminars, etc.), French teachers reported limited professional development compared with other countries.

Programme for the International Assessment of Adult Competencies results: an overall poor French performance

In the 2012 Programme for the International Assessment of Adult Competencies (PIAAC) survey, 7.7 % of French adults (aged 16–65) reached the highest levels (levels 4 and 5) and 34 % reached level 3, compared with 11.8 % and 38.2 %, respectively, in the participating OECD countries (Jonas, 2013). In terms of numeracy proficiency, only 8.3 % of French adults (aged 16–65) reached the highest levels (levels 4 and 5) and 29 % reached level 3, which were well below the averages of other OECD countries – 12.4 % and 34.4 %, respectively. The proportions of French adults with low literacy and numeracy

scores (level 1 or below) were some of the highest among participating OECD countries, at 21.6 % and 28.0 %, respectively, compared with 15.5 % and 19.0 % across the OECD countries. In addition, France is characterised by greater generational differences than other countries, with France's poor performance largely stemming from the results of 45- to 65-year-old adults, whereas 16- to 44-year-old adults scored closer to the average (though still below it), with parents' education level significantly influencing the results.

French basic skills and accountability reforms: various policy instruments to increase student performance and to reduce the inequality gap

The implementation of French education acts and policies in response to international survey results will be analysed through Lascoumes and Le Galès' theoretical framework (2007). Public policy is determined by objectives (equality, equity, etc.) and also by instruments (standards, indicators, etc.) to target a certain population (schools and students). Beyond increasing performance and reducing the inequality gap for disadvantaged French students, French policymakers have attempted to solve different problems, such as illiteracy and social exclusion, early school leaving and dropout, and school ghettoisation. We observe that some cognitive and instrumental components of French education policies impact on the resources and instruments allocated to different education programmes. Instruments can be defined, according to Lascoumes and Le Galès, as devices that are both technical and social, and that organise specific social relationships between public authorities and a diversity of actors.

A policy agenda focused on educational inequalities, basic skills and soft accountability

Education policies are often aimed at addressing shortcomings related to educational outcomes. As noted above, some of these shortcomings have been identified directly through international surveys, including (i) difficulties faced by French students in mastering basic literacy and numeracy skills, particularly for students with immigrant backgrounds; (ii) increasing inequalities in school achievement between the most socially advantaged students and the most socially disadvantaged students; (iii) violence and discipline problems in schools, which reflect a poor school climate and student well-being; (iv) a lack of support for teachers in relation to learning issues; and (v) challenges in developing in-depth skills and knowledge (problem-solving, analysis of texts, etc.). Other problems can be deduced from national reports and surveys from the DEPP (the MNE's directorate) and the General Inspectorate, including (i) grade repetition; (ii) absenteeism and dropout rates, which are detrimental to disadvantaged students; (iii) high failure rates in the early school stages; (iv) a lack of guidance for deprived students; and (v) social segregation between secondary schools.

The DEPP's missions and focus on international activities have led to an increasing reliance on international surveys for educational monitoring at national level (Chatel, 2013; Pons, 2011). Increased possibilities for quantifying educational problems, the development of

the national statistical information system, and strengthened links with national and international research have made it possible to conduct a fair and exhaustive diagnosis. The Centre national d'évaluation du système scolaire (National Centre for the Evaluation of the School System (CNESCO)), created by the former minister Vincent Peillon, has also produced, through multiple reports and consensus conferences, a compromise on reform paths, notably the development of a national strategy for literacy and numeracy, school dropout and guidance issues, school inclusion, and teacher training (Mons, 2013, 2016). The Conseil d'évaluation de l'École (School Evaluation Council) created by the current minister, Jean-Michel Blanquer, is currently working with experts, the DEPP and the General Inspectorate to improve the current accountability system and to promote school self-evaluation and evidence-based education.

Beyond the governmental and media coverage, the setting of a policy agenda that takes into account international surveys has been supported by important legislative and regulatory work through school programming acts that have strengthened curricular requirements in basic skills. Similarly, the implementation of an institutional law regulating finance laws – *loi organique relative aux lois de finances* (?) (LOLF) – has established new public management principles and accountability mechanisms for the whole public administration placed under the control of the parliament (see below).

The challenging mediation between the bureaucratic technostructure and interest groups

Various resources and processes have led to the systemic integration of data from international surveys to monitor the French education system. Given the French education trajectory, a consultation process was initiated, which was strongly guided by the MNE, its technostructure and its related organisations (high councils, General Inspectorate, parliamentary commissions, expert groups). The MNE has also been confronted by internal interest groups (professional bodies, unions, specialist associations) as well as external ones (parents, political parties, businesses) that have influenced successive ministers, depending on their ongoing ideology, to favour one group or another. Although implementing a basic skills policy has reached a certain level of consensus, reforms to redesign the national curriculum, implement national assessments, arbitrate between school choice and school mix, and revise teacher training and its institutions have been subjected to strong ideological disputes. As a consequence, the current political landscape is still overshadowed by disagreements about institutional positioning that are influenced by general elections and political options. The actions of the ministerial technostructure and hierarchical pipelines, from the MNE to schools, continue to be determining factors in implementing top-down reforms. Competition between ministerial

(?) The finance laws are governed by the LOLF, a text that sets out content, determines presentation and governs examination and voting in parliament. The rules thus defined by the LOLF, voted for in 2001 and applied from the 2006 Finance Act onwards, have replaced those of Ordinance No 59-2 of 1959 on the Organic Law on Finance Acts. A true financial constitution of the state, the LOLF has profoundly reformed the construction and monitoring of the state budget. The LOLF requires the creation of a readable budget, presented by public policies (security, culture, education, justice, etc.), called 'missions', instead of types of expenditure (staff, operation, investment, intervention, etc.). These missions are themselves broken down into programmes.

directorates and their departments, which have different but normally complementary implementing programmes, reveals tensions and contradictions across educational local territories and among deconcentrated administrations. Finally, legal and regulatory constraints that codify decision-making slow down the policy-setting agenda; however, a certain scepticism exists among practitioners who are quite aware of the gap between ministerial intentions and ministers' concrete daily actions.

The implementation of various policy instruments despite resistance

As Hood and Margetts (2007) argue, public action involves a varied typology of instruments, some of which mobilise data and their management in a context of greater digitalisation. From this perspective, the French MNE, in its performance and accountability policy, uses at least four types of instruments based on performance data. It uses legal and regulatory instruments, such as the LOLF, that, broken down into programmes and operational objectives, stipulate the MNE's expectations of its school administration. The LOLF requires the MNE to address an annual performance programme, tracing objectives and associated indicators, and expected outcomes. This typology of indicators, compiled at national level by the DEPP, measures governance effectiveness and efficiency within the education system and at local level (e.g. the management of staff replacements), and the quality of the service provided (users' assessment of services: frequency of courses and quality of timetables, for example). Within this framework, *recteurs* are responsible for using operational budgets to achieve the objectives defined by the MNE. The local decentralised state authorities are clearly the managing structure, but this new public management has had little impact until now on the daily work of teaching teams and principals.

Other types of instruments are more incentive based. For example, teachers and inspectors have discretion to use the results of national assessments, which are mandatory, for pedagogical purposes. Similarly, international survey data are presented in training sessions organised by the national training institute for school principals and inspectors, often in conjunction with other information related to accountability and management. Here, the empowerment is indirect, because executives are invited to raise teachers' awareness to develop a 'culture of evaluation', but they do not impose any accountability mechanisms apart from the administration of regular annual national assessments at grades 1, 2 and 6.

Owing to their informative and communicative relevance, reports from the General Inspectorate or those produced by CNESCO or expert groups are used to explain the value of and interest in international surveys as a kind of barometer for national education policies. These reports have a strong institutional and bureaucratic dimension, as they are intended to be applied to the whole education system under ministerial control. They produce a certain representation of what counts in terms of educational performance, as well as the values underlying administrative, scientific or political intentions, particularly in terms of equality of opportunities. Nevertheless, these policy instruments also have unintended effects and consequences, as we will see.

French political reforms and their unintended effects and consequences during the past decade

The French education policy trajectory over the past decade can be characterised as an attempt to make up for the delay in implementing an accountability and basic skills policy framework, while also incorporating the European Open Method of Coordination's indicators and benchmarks and international survey data. Several policy measures have been adopted to support these decisions under successive ministers: Vincent Peillon (2012–2014), Benoît Hamon (2014–2016), Najat Vallaud-Belkacem (2013–2017) and, since 2017, Minister Jean-Michel Blanquer. These policy measures are more or less coherent with international trends and European recommendations.

The Peillon reforms and their adverse impact

The Peillon administration continued the focus on a basic skills framework within schools with the adoption of a 'More teachers than classrooms' programme in priority education areas under the motto 'refoundation of the school system'. His efforts further targeted the reform of timetables in elementary schools, a national plan on school dropout, a reform of teachers' initial training and the creation of a national educative programme focused on the school climate.

In 2013, the High Council of Curricula, responsible for proposing curricular reforms to the minister, and reflecting on the modalities of student skill assessments as well as the initial training of teachers, was created. The content of school curricula was reconsidered in relation to its coherence and complementarity with the idea of reinforcing the basic skills framework and making it more interdisciplinary. However, these new national curricula had little impact on teacher training and teaching in schools, or even on the development of a 'culture of evaluation'.

The 'More teachers than classrooms' programme aimed to assign an additional teacher to challenging schools through the implementation of a new pedagogical organisation and a variety of learning situations better adapted to meet students' needs and enhance basic skills. Resources and support had to be implemented to address these new responsibilities and tasks among teachers. The priorities were learning to read, writing and mastering oral language, mathematics, and student work methodology. However, as a result of uneven implementation by local decentralised state authorities, due to factors such as the selection of schools, the conditions for training and support, the constitution of student groups, and the modalities of external intervention or co-intervention in schools, the national monitoring committee recommended in 2017 that it had to be more visible to be more effective. Despite an effective and more appropriate framework coordinated at various levels, pooling, harmonising and disseminating tools, and teacher training that was better supported by educational research and co-intervention and co-teaching, this programme was later abandoned by Minister Blanquer.

The decree of 24 January 2013 on the organisation of schools' schedules modified the timetables in kindergartens and elementary schools to set up complementary educational activities in small groups and to support students with learning difficulties.

The programme created a contractual commitment between local educational authorities, local public services and other partners to better articulate different timetables for children. The aim was to give new coherence to the children's day and to create positive conditions for school achievement and personal development. With resistance from and limited involvement of teachers and unions, this poorly prepared reform, which strongly modified the daily work of teachers without any financial compensation, resulted in significant additional costs for municipalities, led to strong opposition from local elected politicians and was also later abandoned by Minister Blanquer.

Following the OECD recommendations, the reform of) was intended to transform teacher training by better articulating theory and practice, developing observation and training on the job, and implementing induction and mentoring programmes. Future teachers would also learn to develop innovative teaching methods based on strengthened links with research and openness to international best practices. Here again, the ways in which the reform was implemented varied greatly from one institution to another, with little actual effect on or changes in the training model. This model remained highly academic, with a perceived misfit with the skills requirements to be met when recruiting new teachers. Initial teacher education in France is competitive, and admission procedures value academic skills, whereas the teaching profession increasingly requires pedagogical and socioemotional competences (Gendron, 2008). The mechanisms and processes designed to assess readiness or competence to take up a teaching post and to recruit are not questioned but structured by a recurrent divide between the training institute linked to the university and the employer, the Ministry of Education. Similarly, links with research remained limited, as did the transfer of scientific knowledge to teaching practices.

The Blanquer reforms: an international alignment with related uncertainties

In the presence of stop-and-go educational policy trends in France, and confusing professional communities on the ground (Normand, 2013), the Blanquer reforms introduced further change, with several new measures pursuing the basic skills and accountability policy, responding to problems raised by international surveys.

- The MNE's resources were rerouted to the first years of the elementary school system to enhance skills in numeracy and literacy, which was referred to by the minister as the 'reading, writing and counting' policy, getting back to international basics and mainly concentrating ministerial efforts on the 3Rs, 'reading, writing, arithmetic' (Gendron, 2017, p. 46).
- National assessments designed by the National Scientific Council chaired by Stanislas Dehaene, an eminent neuroscientist, were implemented and used to a greater extent, while the council promoted teaching practices inspired by neuroscience and evidence-based research methods. A greater emphasis was put on student memorisation and metacognition, as well as feedback from teachers in the classroom.
- A reduction in class size in elementary schools for priority education areas emphasised the development of basic skills in numeracy and literacy.

- A mandatory 3-year-old age enrolment in kindergarten was established so that children could begin their formal learning earlier.
- A 'homework done' measure allowed junior school students to do their homework in schools while being assisted by volunteer teachers and pedagogical assistants.

Two years after these reforms, the DEPP data underlined a greater stability in student results, with three quarters of students entering grade 1 reading at the expected level. Nevertheless, gaps remained significant between students in priority education areas and students in other areas. The comparison of French students' results at grade 1 between the 2018–2019 school year and the 2019–2020 school year showed very little improvement. At this level, results in both French language and mathematics were slightly lower, particularly in areas studied at the end of kindergarten. For example, the percentage of students with satisfactory mastery of 'Knowledge of the names of letters and the sounds they produce' dropped from 80.1 % to 77.8 %. Similarly, the percentage of students with satisfactory mastery of the skill 'Manipulating syllables' (sound discrimination) dropped from 81.3 % to 79.3 %. Lastly, the percentage of students with satisfactory mastery of the skill 'Understanding words read by the teacher' dropped slightly from 70.3 % to 69.1 %. In mathematics, the percentage of students with satisfactory mastery of the skill 'Solving problems' dropped from 66.1 % to 64.4 %. In grade 2, decreases were even more noticeable.

In French language, although performance was up in five of the six comparable domains in 2019, there were declines in seven out of eight domains in 2020. Thus, the proportion of students with satisfactory mastery of 'Reading text' dropped by 4.8 percentage points, the proportion with satisfactory mastery of 'Reading words' dropped by 4.3 percentage points, and the proportion with satisfactory mastery of 'Writing words' dropped by 4.5 percentage points. Across the six comparable domains across 2018, 2019 and 2020, the increase in performance averaged 2.2 percentage points between 2018 and 2019, whereas the decrease between 2019 and 2020 averaged 2.8 percentage points.

Decreases were observed between 2019 and 2020 at grade 2 level, especially in specific areas of reading and writing, regardless of students' schooling area. Most notably, the proportion of students with satisfactory mastery of 'Reading words aloud' dropped from 72.6 % in 2019 to 68.3 % in 2020, and the proportion of students with satisfactory mastery of 'Writing words' dropped from 77.1 % in 2019 to 72.6 % in 2020. In contrast, performance was comparable on average between 2019 and 2020 in mathematics, except for the most at-risk students. Only one area, 'Representing whole numbers', showed a drop of more than 1 percentage point (– 1.2 percentage points).

Conclusion: implementation gaps and the long road to reducing inequalities between students

According to international surveys and the DEPP national surveys, France has not been particularly successful in improving its position in international rankings or in increasing students' scores. Several explanations can be given regarding experiences

in other countries such as Finland or Asian countries, although their educational policy drivers are not necessarily easily transferable to the French context. Moreover, those societies might be considered in relation to other student characteristics, such as the rate of migration, or other negative consequences due to an excessive focus on high test scores, which creates student stress, anxiety and dropout. Regardless, the complexity of the French governance system does not help lead to the implementation of effective reforms. The lengthy transmission of official rules or instructions, with a multilevel hierarchical and command chain, leads to a loss of information that is detrimental to a clear understanding and school improvement (Normand and Derouet, 2016). Middle managers are locked into formal frameworks and procedural rules that inhibit programmes' effectiveness and efficiency instead of developing leadership practices.

In addition, despite the importance of continuous professional development as an essential driver of student success (Cordingley et al., 2005), France lacks effective designs capable of providing schools with sustainable support and positive conditions that would facilitate cooperation and peer learning, and a mechanism for exchange of experiences and best practices. Links between teachers and educational research are weak and lack dynamics in pedagogical innovation. Furthermore, principals do not have many skills in relation to these issues – and they prefer to maintain their bureaucratic roles and responsibilities (Normand, 2016, 2020b). The lack of local autonomy and difficulties in adapting the curriculum to students' needs, along with the challenges faced by teachers in diversifying assessment practices beyond marking, can explain the ongoing failure to improve students' basic skills. Although policy devices have been designed to personalise learning, including class size reductions and teaching methods centred on students' metacognition and teachers' formative feedback (see, for example, Hattie, 2012), they have had a limited impact. The rigid and bureaucratic structure of timetables and teacher services makes such diversification almost impossible. As Hattie (2015) argued, the French education system is not exempt from distracting policies that often cost a lot of money and contribute little to student learning. Standardisation of learning, based on a regular revision of the national curriculum, when not correlated with students' real learning situations, is unlikely to improve their cognitive skills. Similarly, the development of early childhood education will have little impact if it does not lead to diversifying student learning (visual, auditory, kinesthetic). Moreover, adding another teacher to support the regular teacher in the classroom will have no added value for learners' performance if the additional teacher is underqualified or poorly qualified. The emphasis on pre-service training rather than in-service training has limited the possibilities for adjusting practices to learning situations and further enhancing students' cognitive or non-cognitive skills. Lastly, the use of digital technologies, a policy strongly emphasised in France, will have little impact if it is not supported by the transformation of teaching practices, initial preparation and a recruitment mechanism.

Overall, despite international evidence, French policymakers have sought to implement solutions that will likely lead to dead ends. As reported across international surveys and national data, French students' scores are not improving, and students' chances of escaping their socioeconomic background remain smaller in France than in many other OECD countries. Despite successive multiple reforms, there is still a long way to go to reduce the inequality gap in the French school system.

REFERENCES

- Chatel, E. (2013), 'L'évaluation de l'éducation dans le contexte de mondialisation: le cas français', *Tréma*, Vol. 40, pp. 28–40.
- Cordingley, P., Bell, M., Evans, D. and Firth, A. (2005), 'The impact of collaborative CPD on classroom teaching and learning – what do teacher impact data tell us about collaborative CPD?', *Research Evidence in Education Library*, Evidence for Policy and Practice Information and Co-ordinating Centre, Social Science Research Unit, Institute of Education, University College London, London.
- DEPP (Directorate of Evaluation, Forecast and Performance) (2016a), 'Note d'information No 37 – Décembre 2016 – Les élèves de 15 ans en France selon PISA 2015 en culture scientifique: des résultats stables, toujours marqués par de fortes inégalités', DEPP, Paris.
- DEPP (2016b), 'Note d'information No 38 – Décembre 2016 – PISA 2015: l'évolution des acquis des élèves de 15 ans en compréhension de l'écrit et en culture mathématique', DEPP, Paris.
- DEPP (2017), 'Note d'information No 17.24 – Décembre 2017 – PIRLS 2016: évaluation internationale des élèves de CM1 en compréhension de l'écrit – Évolution des performances sur quinze ans', DEPP, Paris.
- DEPP (2019a), 'Note d'information No 19.49 – PISA 2018: stabilité des résultats en compréhension de l'écrit', DEPP, Paris.
- DEPP (2019b), 'Note d'information No 19.50 – PISA 2018: culture mathématiques, culture scientifique, et vie de l'élève', DEPP, Paris.
- DEPP (2020), 'Note d'information No 20.46 – Décembre 2020, TIMSS 2019 – Évaluation internationale des élèves de CM1 en mathématiques et en sciences: les résultats de la France toujours en retrait', DEPP, Paris.
- Dobbins, M. and Martens, K. (2012), 'Towards an education approach à la finlandaise? French education policy after PISA', *Journal of Education Policy*, Vol. 27, No 1, pp. 23–43.
- Eurydice (2021), 'Key features of the education system – France' (https://eacea.ec.europa.eu/national-policies/eurydice/content/france_en).
- Gendron B. (2008), 'Emotional capital and emotional work: emotional competencies as teacher professional competencies to teach for learning', in Deitmer, L., Kämäräinen, P., Manning, S. (eds), ECER Vetnet Conference 2008 Proceedings – From Teaching to Learning?, Gothenburg (<https://www.eera-ecer.de/ecer-programmes/conference/1/contribution/535/>).
- Gendron, B. (2017), 'Emotional capital: the set of emotional competencies as professional and vocational skills in emotional works and jobs', *Revista Española de Educación Comparada*, No 29, pp. 44–61.
- Hattie, J. (2012), *Visible Learning for Teachers: Maximizing impact on learning*, Routledge, London.
- Hattie, J. (2015), *What Doesn't Work in Education: The politics of distraction*, Pearson, London.

- Heurdiere, L. and Clément, P. (2016), 'Aux origines du Conseil national des programmes (1983–1990): instruments et acteurs d'une nouvelle gouvernance des contenus d'enseignements', *Revue française de pédagogie*, Vol. 2, pp. 5–24.
- Hood, C. C. and Margetts, H. Z. (2007), *The Tools of Government in the Digital Age*, Palgrave Macmillan, London.
- Jonas, N. (2013), 'Les capacités des adultes à maîtriser des informations écrites ou chiffrées – résultats de l'enquête PIAAC 2012', *Insee Première*, No 1467, National Institute of Statistics and Economic Studies, Paris.
- Lascoumes, P. and Le Galès, P. (2007), 'Introduction: understanding public policy through its instruments – from the nature of instruments to the sociology of public policy instrumentation', *Governance*, Vol. 20, No 1, pp. 1–21.
- Mons, N. (2013), *Rapport pour la préfiguration du conseil national d'évaluation du système scolaire*, Ministry of National Education, Youth and Sports, Paris.
- Mons, N. (2016), *Comparaison des évaluations PISA et TIMSS. –Acquis des élèves – Comprendre les évaluations internationales*, CNESCO, Ministry of Education and Higher Education, Paris.
- Normand, R. (2013), 'The doubts and uncertainties of French educators in the face of travelling policies', in Seddon, T. and Levi, J. S. (eds), *World Yearbook of Education 2013 – Educators, professionalism and politics: Global transitions, national spaces and professional projects*, Routledge, London and New York, pp. 184–200.
- Normand, R. (2016), 'France: between civil service and republican ethics – the statist vision of leadership among French principals', in Årlestig, H., Day, C. and Johansson, O. (eds), *A Decade of Research on School Principals*, Springer, Cham, Switzerland, pp. 357–374.
- Normand, R. (2020a), 'France: the French state and its typical "agencies" in education – policy transfer and ownership in the implementation of reforms', in Årlestig, H. and Johansson, O. (eds), *Educational Authorities and the Schools, Educational Governance Research*, Vol. 13. Springer, Cham, Switzerland, pp. 151–168.
- Normand, R. (2020b), 'The long march towards school improvement in France: paradoxes, tensions and adjustments between bottom-up innovations and top-down policies', in Donovan, J., Trimmer, K. and Flegg, N. (eds), *Curriculum, Schooling and Applied Research: Challenges and tensions for researchers*, Palgrave Studies in Education Research Methods, Palgrave Macmillan, London, pp. 21–40.
- Normand, R. and Derouet, J. L. (2016), 'Top-down accountability and local management: tensions and contradictions experienced by French principals as leaders', in Easley, J. and Tulowitzki, P. (eds), *Educational Accountability: International perspectives on challenges and possibilities for school leadership*, Routledge, London, pp. 155–164.
- OECD (Organisation for Economic Co-operation and Development) (2019a), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris.
- OECD (2019b), *PISA 2018 Results (Volume II): Where all students can succeed*, OECD Publishing, Paris.

Pons, X. (2011), 'What do we really learn from PISA? The sociology of its reception in three European countries (2001–2008)', *European Journal of Education*, Vol. 46, No 4, pp. 540–548.

Pons, X. (2015), *Évaluer l'action éducative: des professionnels en concurrence*, Presses Universitaires de France, Paris.

Schleicher, A. (2018), *PISA 2018: Insights and interpretations*, OECD ([PISA 2018 Insights and Interpretations FINAL PDF.pdf](#))

Van Zanten, A. (2015), *Choisir son École: Stratégies familiales et médiations locales*, Presses universitaires de France, Paris.

CHAPTER 4

Cross-national achievement surveys and educational monitoring in Italy

Orazio Giancola and Luca Salmieri

Abstract

This chapter provides insights into the use of cross-national achievement surveys to monitor educational outcomes and develop and implement education policies in Italy. The full roll-out of the national assessment of students and schools has taken place only over the past 10 years in Italy, and so far weak reciprocities can be reported between data from cross-national surveys and those from the national assessment programme. Furthermore, educational policies and reforms have been informed not by the findings of international assessments, but rather by multiple uncoordinated and sometimes contradictory attempts to mitigate old deficiencies in the education system without using evidence-based research capable of empirically demonstrating correlations between structural problems and defined targets. The chapter begins by introducing the Italian education system, its governance and reforms, as well as organisational shifts in the recent past. It then highlights the main performance patterns of the Italian education system in relation to international achievement surveys. A third section outlines the main features of the monitoring processes applied to the Italian education system, their marked incorporation of the Organisation for Economic Co-operation and Development's Programme for International Student Assessment methods and limited use of policy drivers. Finally, we highlight intended and unintended outcomes that derive from recent reforms.

Introduction

In accordance with the Constitution of the Italian Republic, education is accessible to all and compulsory education is free. The Italian state shares some of its responsibilities with the 20 regions, and education is among them. However, many aspects of education, especially the task of setting general rules, fall under the exclusive legislative power of the state, even though organisational aspects have been decentralised to local level. Together with several European Union Member States such as France, Italy is part of a broader educational model based on state intervention, a rigid and centralised national curriculum, and widespread public schools. This model has paved the way for a late but massive democratisation process in ensuring access to upper secondary education and the relative equality of opportunities and achievements (Barone and Guetto, 2020; Brunello and Checchi, 2005). Educational attainment in Italy is low by international standards, but it has increased significantly among the more recent cohorts. The Programme for International Student Assessment (PISA) coordinated by the Organisation for Economic Co-operation and Development (OECD) currently ranks the overall skills of Italian 15-year-olds 34th in the world in reading, literacy and mathematics, significantly below the OECD average (OECD, 2019).

The Italian system is based on compulsory education provided for at least 10 years, which covers the age group between 6 and 16 years. The completion of compulsory education involves earning a certificate attesting to the completion of 2 years of upper secondary school by the age of 18. After successfully completing primary education (International Standard Classification of Education (ISCED) level 1), all students progress to lower secondary level (ISCED level 2), where they follow the same common core curriculum. The final 2 years of compulsory education correspond to the first 2 years of upper secondary school, amounting to 5 years of compulsory education in total. Therefore, the compulsory education system is characterised by the provision of a common core curriculum that lasts from the first year of primary school to the first examination occurring after 3 years of lower secondary school. Compulsory education also continues after this point, however, when students choose an upper secondary track from among three differentiated offerings – (i) *liceo*, (ii) technical school or (iii) professional school – during which they must complete at least two school years to cover the compulsory period. This structure, in which a compulsory period falls in the middle of upper secondary education, is atypical among European education systems. Almost all Italian students proceed with a further 3 years of upper secondary education to sit the final exam and obtain the state certificate. Once they earn an upper secondary qualification, they are eligible to enrol in tertiary education provided by universities no matter which school track they embarked on. Alternatively, after lower secondary school, students may opt for vocational education and training courses lasting at least 3 years (*istruzione e formazione professionale*). These courses fall under the exclusive legislative purview of individual regions and follow specific curricula related to local labour market needs. Students choosing this course of study are not eligible to enrol in university (for a complete overview of the Italian system see Figure 4.1).

Early childhood education and care (ECEC 0–3) and pre-primary childcare (3–6) comprise a single system called the ‘integrated system’, which is not compulsory. Despite forming part of the same system, ECEC is organised by local authorities (regions and municipalities), whereas the pre-primary school offering falls under the responsibility of the Ministry of Education.

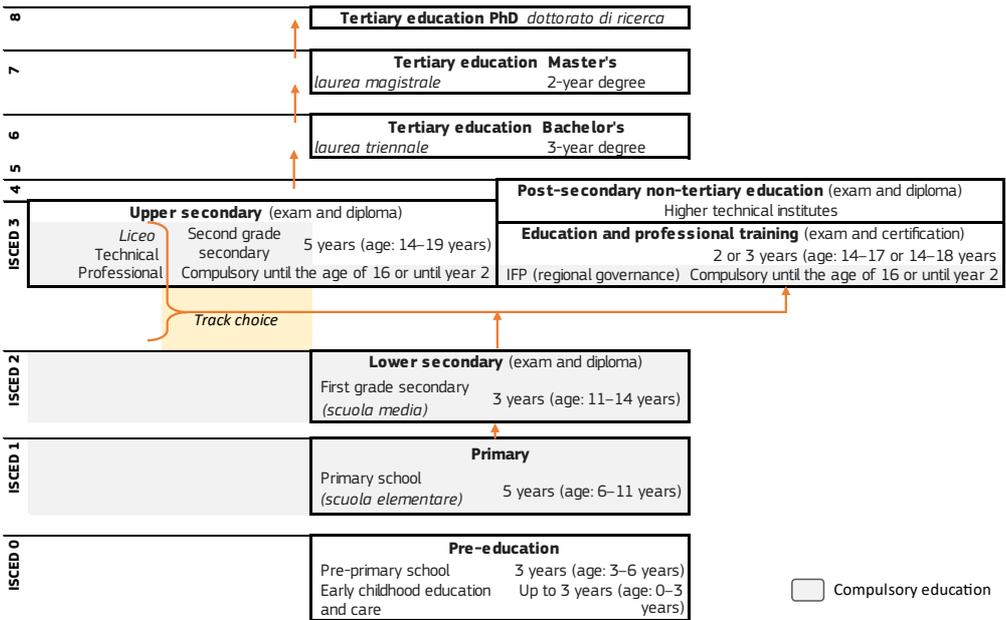
Primary education, which is called elementary school in Italy, begins when pupils have completed pre-primary childcare and are aged 5 or 6. Primary education lasts for 5 years and is free. Students pass from one year to the next and from primary to lower secondary education without exams. The first national examination does not occur until the transition from lower to upper secondary school, when students are aged 13. State public primary schools account for 96.4 % of enrolment, whereas government independent and dependent private primary schools account for 1.7 % and 1.9 % of enrolment, respectively (Eurydice, 2020). The national curriculum on which primary education is based involves a minimum number of weekly teaching hours (from 24 to 40 hours, depending on the regime chosen by parents) dedicated to the Italian language, one foreign language (English, or more rarely French or Spanish), history, geography, maths, sciences, technology and information technology, music, art, physical education, and Catholic religion or alternative activities. Each school is also free to provide some additional activities at its discretion, but these cannot replace any of the core subjects.

Secondary education lasts for 8 years and is divided into two stages: lower secondary school (*scuola secondaria di primo grado*), commonly known as middle school (*scuola media*), which corresponds to ISCED level 2, and upper secondary school (*scuola secondaria di secondo grado*), which corresponds to ISCED level 3. Lower secondary school lasts for 3 years, from 11 to 13, and upper secondary school lasts for 5 years, from 14 to 18.

Since 2009, lower secondary school has followed one of two weekly plans, the first, ‘basic time’, amounting to 30 hours per week, and the other, ‘extended time’, ranging from a minimum of 36 hours to a maximum of 40 hours per week. In both cases, the national curriculum is binding: the subjects taught are Italian, history and geography, maths and sciences, English, a second EU language (usually French or Spanish), technology, art, physical education and sports, music, and Catholic religion or alternative activities, with the addition of computer and digital technologies plus other extracurricular activities in the case of extended time.

By passing the state examination, lower secondary students obtain the national certification required to enrol in upper secondary school. The exam is formulated directly by the Ministry of Education and administered uniformly across Italian schools. Students face three written tests (Italian, maths and a foreign language) and an oral interview on several subjects. To be eligible to sit the exam, students must have previously taken national tests administered by the National Evaluation Institute for the School System (INVALSI) that form part of the country’s schooling assessment tools. Finally, the upper secondary cycle is divided into the first biennium required to complete compulsory education, a second biennium and a final fifth year at the end of which students take the second national exam, which grants them access to university if they pass.

Figure 4.1: The structure of the education system in Italy



Source: Eurydice (2020).

NB: IFP, *istruzione e formazione professionale* (vocational education and training).

As mentioned above, students entering the upper secondary cycle must choose either *licei*, technical schools or professional schools. This tripartite instead of bipartite track system represents Italy's second point of atypicality in relation to the most common forms of bifurcation found among European education systems (Azzolini and Vergolini, 2014). The *liceo* has historically been conceived as fostering general education and providing students with the competencies and knowledge needed for higher-level studies, as well as cultural and methodological skills: students already planning to pursue tertiary education are very likely to opt for a *liceo*, studying one of six curricula (sciences, classical studies, human sciences, arts, languages, or music and dance). Technical schools offer technical and applied education and provide students with a scientific background in the economic and technological professional sectors. Professional schools offer vocational education and provide students with a vocational background in the sectors of services, industry and handicraft. The latter are intended mainly to facilitate access to the labour market.

The general organisation of the education system (national curricula, minimum standards, staffing, evaluation and monitoring, financial resources) falls exclusively to the Italian state (Fornari and Giancola, 2011). The 20 Italian regions hold joint responsibility with the state in some sectors of the education system, such as providing ECEC, and establishing the school calendars for primary and secondary education and the distribution of schools in their territory. Regional authorities also organise facilities maintenance, merging or establishing new schools, and student transport from ECEC to

upper secondary education, at local level. Municipalities (local authorities corresponding to cities and towns) have specific responsibilities in pre-primary, primary and lower secondary education, ensuring the ordinary and extraordinary maintenance of school premises and potentially supplying services such as transport, canteens, and textbook vouchers and financial grants according to the general regulations stipulated jointly by the state and regions.

Italian schools have a certain degree of autonomy, especially in administrative and accountancy domains. They may define some elements of national curricula and widen educational offerings, but they are required to adhere to the national curriculum. Furthermore, they depend on central, regional and provincial procedures for staffing, with school principals hiring new teachers from predetermined pools.

Education at all levels is open to Italian citizens, as well as foreign minors from both EU and non-EU countries. In the latter case, individual schools take specific measures to personalise and adapt the curriculum at local level and, for immigrants who have not yet mastered Italian, provide linguistic support.

Recent non-radical reforms

Over the past 20 years, the Italian education system has undergone a series of non-radical reforms, gradually shifting several facets of governance, autonomy and accountability while leaving the backbone of compulsory education unchanged. An initial step to reform school autonomy was launched in the late 1990s and completed in 2000. Although schools remain part of a national education system, they now enjoy administrative, organisational and, to some extent, educational autonomy (Landri, 2009). Nonetheless, schools are currently required to operate in compliance with the state's general norms on education issued at national level.

In 2003, the Moratti reform brought about a series of changes in the organisation of primary and secondary schools: the teaching of English and the use of computers were introduced and promoted in primary schools, whereas a final exam at the end of 5 years of study was abolished. The portfolio, an individual dossier documenting and certifying the student's curricular and extracurricular activities, was introduced in lower secondary schools. In the same period, a national system of school evaluation and accountability was set up, and INVALSI was established to carry out these standardised assessments. Following a series of pivotal experimental projects, school evaluation and monitoring was fully rolled out in the school year 2005/2006 with the administration of national standardised tests in maths and Italian for primary and secondary school pupils.

In 2010, a second reform reintroduced one-teacher classrooms in the first years of primary school in place of the former dual-teaching method, and class time was reduced from 33 to 30 hours per week in lower secondary schools. The national standardised test on Italian and maths to assess school quality was integrated into the final lower secondary exam. The real innovation of this reform, however, came fully into effect only in the school year 2014/2015, rearranging the curricula for upper secondary schools.

Previously, *licei*, technical schools and professional schools in Italy offered more than 800 experimental courses and many other independent paths and options, varying significantly from school to school. Implementation of the 2010 reform streamlined and drastically cut this multifaceted range of curricular paths, resulting in 20 uniform, binding and unitary curricula.

Finally, the most recent reform, called ‘La Buona Scuola’ or ‘The Good School’ (Law 107 of 201⁽⁸⁾), has increased the autonomy and responsibility of school principals, strengthened and extended evaluation mechanisms for teaching staff, scheduled a mass recruitment of new teachers and introduced mandatory employability enhancement activities, called ‘school–work alternance’, in upper secondary education.

International achievement surveys

The Italian education system has displayed conflicting trends over the past two decades. On the one hand, there has been a marked improvement in students’ access to and their remaining in upper secondary education, with a constant and vigorous increase in enrolment numbers over the years as well as a recent decline in the country’s historically high dropout rate. On the other hand, there has been substantial stability in students’ performance and scores on both national and international standardised tests. Concomitantly, social inequalities in both students’ educational choices and their cognitive and educational achievement continue to remain almost unchanged even after several governmental reforms.

Examining the main changes brings compulsory education under the magnifying glass. Two key policy interventions have targeted this essential area, focusing on the effectiveness and equity of the Italian education system. After school autonomy was augmented in the 1990s, Law 53 was enacted in 2003 to consolidate schools’ capacity to ensure the right/duty to receive education and training for all students for at least 12 years and, in any case, up to the achievement of a 3-year professional qualification by the age of 18.

A second crucial intervention, based on Law 296⁽⁹⁾ passed in 2006, reinforced the principle that compulsory education should last 10 years, from 6 to 16 years. Despite these efforts, data show that many students were still dropping out of school without obtaining any degree higher than a lower secondary school diploma. Italian early school leavers suffer

⁽⁸⁾ Delega al Governo per la definizione delle norme generali sull’istruzione e dei livelli essenziali delle prestazioni in materia di istruzione e formazione professionale (Delegation to the Government for the definition of the general regulations on education and of the essential levels of services in the field of education and vocational training).

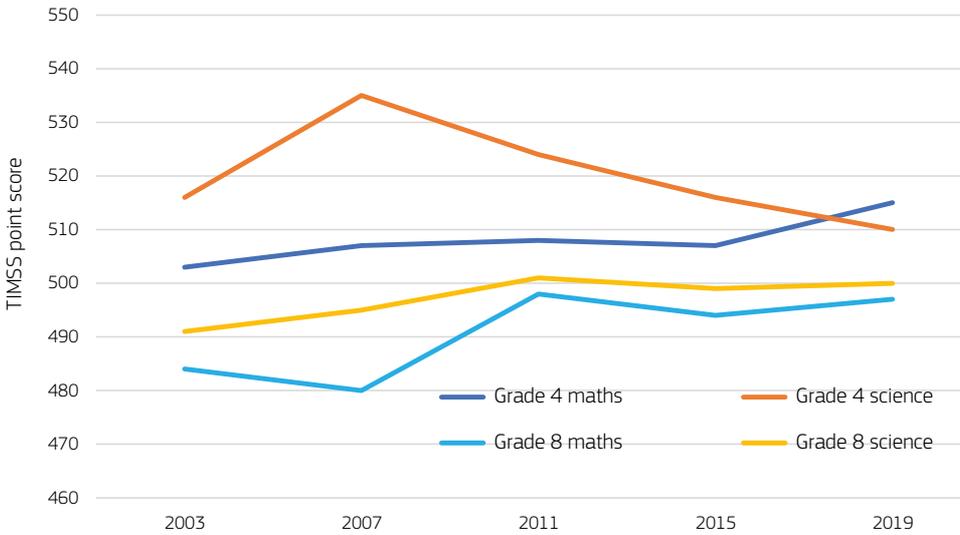
⁽⁹⁾ Legge 27 dicembre 2006, 296, disposizioni per la formazione del bilancio annuale e pluriennale dello Stato, legge finanziaria 2007 (Provisions for the preparation of the annual and multi-year financial statements of the State, 2007 financial law). Article 1, paragraph 622, states that ‘education given for at least ten years is compulsory and is aimed at allowing the achievement of an upper secondary school qualification or a professional qualification lasting at least three years by the age of eighteen’.

a bigger socioeconomic disadvantage in both the short term and the long term than their schooled peers (Borgna and Struffolino, 2017). However, as mentioned above, dropout rates have been decreasing in recent years: the share of individuals aged 18–24 who left secondary education in 2019 was 13.5 %, almost the same percentage reported at the end of the 1990s before a significant surge took place in the new millennium. In addition, although in the past school dropout commonly took place between the second and third year of lower secondary education, it has gradually shifted to upper education. Therefore, although recent data indicate a downward trend in school dropout rates, policies enacted over the past 20 years have primarily mitigated and not eradicated this phenomenon, placing Italy among the EU Member States with the highest incidence of school dropout.

It is also worth noting that students with a migrant background have been entering Italian schools at a faster pace in the past few decades. This growing level of participation from 2000 to the present initially involved primary (10.8 %) and lower secondary schools (9.7 %), but then upper secondary schools were also included (7.3 % of 15-year-old students in 2018) (Giancola and Salmieri, 2018). This shift has significantly affected school organisation, teaching needs (especially in primary and lower secondary schools) and teacher–parent relationships. However, although Italian schools have so far displayed a strong voluntary drive to adopt a multicultural model of education at local level (Santagati, 2021), there has been no radical centralised school reform implementing a broad-based vision for improving the opportunities of foreign-born students. Local adaptations to changing classroom composition that are highly varied and sometimes even contradictory have proved marginal and inadequate to meet the ever-growing array of diverse linguistic, learning and integrated support needs (Azzolini and Barone, 2013).

In terms of learning outcomes, data from national and international assessment surveys show substantial stability over time (or only slight variation). Italian primary school students (fourth grade) perform slightly above the Trends in International Mathematics and Science Study (TIMSS) 2015 average in maths literacy, and this level has remained constant over time. Nevertheless, there are considerable achievement differences between northern and southern regions, reflecting the influence of the country's traditional north–south socioeconomic divide on the educational landscape.

Historically, the Italian education system has produced positive learning outcomes, especially in primary education since, unsurprisingly, this is the education cycle that has undergone fewer radical changes over the years, and a setting in which there is less space for familial strategies of social differentiation and the pursuit of distinction (Grimaldi and Serpieri, 2012). The higher the education cycle, however, the more national and international survey data show a deterioration in the quality, equity and efficacy of education. The performance of Italian students in the third grade of lower secondary education (grade 8) ranks just below the TIMSS average and has tended to remain constant over time after a minor spike in 2011 (see Figure 4.2).

Figure 4.2: TIMSS average scores in maths and science in 2003–2019 in Italy

Source: Created by the authors based on TIMSS 2019, 2015, 2011, 2007 and 2003 databases (<https://timssandpirls.bc.edu/timss-landing.html>).

The influence of family background on student learning outcomes is already notable in primary schools, albeit to a lesser degree. Pupils with direct and indirect access to better material (the availability of books at home) and cognitive resources (parents' educational and cultural awareness) perform better than pupils from disadvantaged backgrounds or lower social classes. These gaps tend to become wider at the next level of education (eighth grade). The correlation between socioeconomic inequalities and disparity in educational performance increases as students transition from lower to upper secondary school, when the period based on the national core curriculum ends (see Table 4.1). At that point, the so-called tracking system begins to exert a significant influence on the stratification of the education system. In the Italian system, tracking consists of the tripartite curricular choice between *licei*, which are chosen extensively by children from higher and middle classes; technical schools, with their greater focus on employability through technical professions, which are more likely to be attended by children from middle and lower classes; and professional schools, which are primarily attended by children with migrant backgrounds, those with poorer performance histories and the underclasses (Benadusi and Giancola, 2014). Students attending technical and professional schools usually do not plan to pursue tertiary education after obtaining the upper secondary certificate (Azzolini and Vergolini, 2014; Contini and Scagni, 2011).

Table 4.1: TIMSS educational outcomes according to cultural resources available to students, 2019 Italy and TIMSS averages

	High level of resources		Medium level of resources		Low level of resources	
	% of students	Average achievement	% of students	Average achievement	% of students	Average achievement
Grade 4 maths – Italy	10 (0.8)	552 (4.4)	83 (0.9)	515 (2.5)	7 (0.8)	491 (5.3)
Grade 4 maths – average	17 (0.1)	562 (0.7)	75 (0.2)	498 (0.5)	8 (0.1)	433 (1.5)
Grade 8 maths – Italy	15 (0.9)	536 (4.4)	72 (1.0)	498 (2.5)	13 (1.0)	452 (5.3)
Grade 8 maths – average	14 (0.1)	546 (1.0)	73 (0.2)	488 (0.5)	13 (0.1)	433 (1.2)
Grade 4 science – Italy	10 (0.8)	551 (4.0)	83 (0.9)	511 (2.9)	7 (0.8)	478 (5.8)
Grade 4 science – average	17 (0.1)	557 (0.8)	75 (0.2)	488 (0.5)	8 (0.1)	414 (1.7)
Grade 8 science – Italy	15 (0.9)	546 (3.4)	72 (1.0)	501 (2.4)	13 (1.0)	449 (5.5)
Grade 8 science – average	14 (0.1)	549 (1.0)	73 (0.2)	489 (0.6)	13 (0.1)	431 (1.2)

NB: Figures in brackets denote the standard error.

Source: Created by the authors based on TIMSS 2019 database (https://timss2019.org/international-database/downloads/T19_G4_SPSS%20Data.zip)

The latest edition of the PISA assessment provides an in-depth intertemporal analysis of student performance in the upper secondary cycle (OECD, 2019). Italian students' scores in reading literacy have been stable since 2000. Average scores in mathematical literacy have increased, with a tendency to flatten out in recent years, whereas scores in science literacy initially improved and then began a downward trend in 2012. Mean performance in Italy after 2012 declined in reading and science, and remained stable (and above the levels observed in 2003 and 2006) in mathematics. Reading performance declined among girls in particular (while remaining stable among boys). Science performance declined most markedly among the highest-achieving students, by a similar amount for both boys and girls.

Overall, Italian students' average PISA scores are consistently below both OECD and EU averages. In this context, there are two patterns to highlight: differences in student performance according to both geographical area and school track – which have been recurring features of the Italian educational landscape. Although 23.3 % of the overall national sample of 15-year-olds fell under the minimum proficiency level in reading (low performers) in the most recent survey – slightly higher figure than the OECD average of 22.6 % – schools in central and northern Italy reported a smaller share of low performers in reading, maths and science. Low performers in southern schools totalled 35.1 % for reading, 37.7 % for mathematics and a distressingly high 40.1 % for science. Several research studies, repeated over time, show that Italy's north-south divide in educational outcomes is not the result of quality in teaching and schools' organisational and didactical proficiency *per se*; rather, it derives from the cumulative convergence of multiple societal factors: disparities in general quality of life, structural inequalities in cultural resources, gaps in the educational levels of the overall population and, last but not least, average public expenditure per student at regional level (Argentin and Triventi, 2015; Ferrer-Esteban, 2011). Individual schools and the school system as a whole seem to be highly sensitive to local socioeconomic variables. Previous national education system reforms failed to consider the social features of schools as institutions that are fully integrated into and relatively dependent on their local ties to the economy, welfare and families.

PISA test scores also demonstrate significant disparities between students in Italy's three upper secondary school tracks (see Table 4.2). What is at work is not only the marked impact of social origin (calculated through the PISA index of economic, social and cultural status (ESCS)) on upper secondary students' scores in reading, mathematics and science, but also the even more intense impact of track type. There is thus a clear chain reaction in which social origins and school track choice concatenate to affect learning outcomes (Giancola and Salmieri, 2020; Hanushek and Wößmann, 2006).

Table 4.2: Average PISA scores and differences among school tracks, 2000–2018 Italy

	2000	2003	2006	2009	2012	2015	2018
Reading							
Italy	487	476	469	486	490	485	476
1. <i>Liceo</i>	544	525	525	541	537	526	521
2. Technical school	478	474	463	476	476	473	458
3. Professional school	429	409	391	417	415	410	395
Gap: general versus non-general (*)	90	84	98	94	92	84	95
Maths							
Italy	—	466	462	483	485	490	487
1. <i>Liceo</i>	—	503	499	520	521	521	522
2. Technical school	—	472	467	488	486	490	482
3. Professional school	—	408	400	423	414	424	405
Gap: general versus non-general (*)	—	63	65	64	70	64	79
Science							
Italy	—	—	475	489	494	481	468
1. <i>Liceo</i>	—	—	518	531	530	513	503
2. Technical school	—	—	475	489	491	480	460
3. Professional school	—	—	414	427	425	411	394
Gap: general versus non-general (*)	—	—	73	74	72	67	75

(*) Denotes the difference between the *liceo* average score and average score for technical and professional schools.

Source: Created by the authors based on PISA 2018, 2015, 2012, 2009, 2006, 2003 and 2000 databases (<https://www.oecd.org/pisa/data/>).

Based on the PISA data set, we can analyse trends over time in students' distribution among the three education tracks. There has been a constant increase in enrolment rates in the general school track (*liceo*) over time, with a consequential decline in enrolment rates in technical and professional secondary schools. *Licei* accounted for 35.6 % of all enrolment in upper secondary schools in 2000, whereas technical schools attracted a higher share (40 %) and professional schools attracted the remaining 24.4 %. Fifteen years later, the distribution was reversed: half the student population (50.5 %) were attending *licei*, whereas enrolment in technical and professional schools had dropped to 30.7 % and 18.8 %, respectively.

The recent steady growth in the share of students attending *licei* plausibly suggests a democratisation of the population choosing this school track, attesting to the power of social influence and the cross-class diffusion of the idea that general upper secondary education offers a better social environment and the first step towards social mobility

(Ballarino et al., 2008). Nevertheless, the increase in *licei* enrolment does not automatically suggest that the entire tracking system is moving towards social heterogeneity. Instead, it may well be that the proportion of students at *licei* rises over time as more and more middle- and upper-class families push their children towards this kind of upper secondary school, increasingly shaping school choice as a process of class connotation. Using the ESCS index to measure the social heterogeneity of the PISA student population (average ESCS values in the three tracks), we find that the 2010 reform standardising the three tracks has had a significant redistributive effect only for *licei*, and not the other two school tracks. The *licei* population has been undergoing social heterogenisation since before the 2010 reform, following a trend that has proceeded independently of educational policies. In contrast, the social heterogeneity of technical and professional school attendees, as measured by the ESCS index, has fluctuated over time (Giancola and Salmieri, 2020).

In summary, the first key point is that social origins exert a statistically significant impact on school track choices. This process generates three outcomes: a high level of social segregation between school tracks and, concomitantly, considerable variability – in terms of social composition – between individual schools. Second, both school track type and socioeconomic composition at school level impact on students' educational outcomes as assessed through standardised tests. Third, a mix of social origin, track type effect and school socioeconomic composition influences educational outcomes and is reflected in students' motivations and expectations in relation to tertiary education. Nevertheless, school track type remains one of the key factors explaining inequalities in student learning outcomes. The transition from lower to upper secondary school and the choice of school track stand out as the core factors driving inequality in the Italian education system, and it is specifically this nexus that is not fully targeted by either educational policies or school counselling practices (Azzolini and Vergolini, 2014; Giancola and Salmieri, 2020).

Several empirical studies have demonstrated the mechanism through which educational inequalities in Italy are socially reproduced, showing that the indirect effect of social origins – school track choice – is stronger than direct effects – parents' educational, cultural and economic achievements affecting children's learning outcomes (Pensiero et al., 2019). Middle- and upper-class families strategically exploit the track system and latent school ranking in terms of social composition in order to reinforce their social status (Giancola and Colarusso, 2020).

Cross-national achievement surveys also suggest that gender gaps and differentiation in students' educational choices and performance are noteworthy in the Italian context. More specifically, Italian girls are more likely (in terms of the odds ratio) to pursue general education at *licei* than at technical or professional schools, whereas boys (in percentage terms) are almost equally distributed among *licei* and technical schools, with a smaller share choosing professional schools (Giancola and Colarusso, 2020). In the latest PISA assessment, Italian boys outperformed girls in mathematics by 16 points, a difference that is greater than the average gender gap in mathematics across OECD countries (5 points). Conversely, Italian girls and boys performed similarly in science, with girls only slightly outperforming boys, by 2 points, in PISA 2018. Lastly, although girls significantly outperformed boys in reading in all the countries and economies that participated in PISA 2018 (by 30 points on average across OECD countries), in Italy the gender gap in reading

(25 points) is lower than the average. This gap has also shrunk over time (compared with 46 points in 2009). Overall, boys' performance has remained stable, whereas girls' performance has declined over the same period. These gender gaps are important because they seem to predict students' choices in tertiary education: among high performers in mathematics or science, approximately one in four boys in Italy expects to work as an engineer or science professional at the age of 30, whereas only one in eight girls has this expectation. Reading and mathematics preferences and abilities developed at an early schooling stage also have a significant impact on the other types of PISA literacy assessment measures: girls' lower financial literacy performance than that of boys, for instance, depends more directly on the mathematics gender gap in Italy than in other European countries (Salmieri and Rinaldi, 2020).

Education policy monitoring

In this section, we describe monitoring processes in Italy and the ways in which cross-national achievement survey results are exploited (or not) and sometimes (mis)used to guide policies, adjust education governance and intervene in the policy agenda. International large-scale assessments, such as PISA and TIMSS as discussed in the previous section, are carried out consistently in Italy. Unsurprisingly, national assessments in Italy have increasingly incorporated cross-national achievement surveys into their body of assessment practices. Nonetheless, the potential of assessment findings as evidence-based tools for interpreting the mechanisms and dynamics of education quality remains largely untapped. A growing body of media coverage and empirical research stemming from international large-scale assessments has become part of public discourse on the quality of the Italian education system. However, the results of these assessments play a much smaller and sometimes unintended role in triggering structural reforms in governance, methods, curricula, teaching organisation and students' learning activities.

Broadly speaking, international large-scale assessments have not influenced the Italian education system in terms of either policy configuration or public debate on the strengths and weaknesses of learning outcomes, curricula efficacy and inequalities among students. This is not, however, because Italian experts and policymakers recognise the inherent limits of using international large-scale assessment results to steer specific educational policies. Although Italy has participated in almost all the main waves of cross-national achievement surveys, there is a lack of familiarity with analysing this evidence to improve student performance through changes in the education system. However, what is still missing is both the *ex ante* and the *ex post* use of evidence-based approaches. Indeed, it is not uncommon for one Italian education reform to be replaced by another wholly contradictory intervention before the first one has been fully implemented, and without any move to assess its measurable effects (Grimaldi and Serpieri, 2012).

Over the past 40 years, policy interventions in the Italian education system have been designed to solve or at least mitigate structural problems and modernisation delays that

policymakers and stakeholders (families, economic organisations, trade unions) do not conceive of as directly responsible for the learning outcomes measured by cross-national surveys. Therefore, the main perceptions are that primary education is efficient, highly accessible to and equal for all, generating positive performance with adequate learning outcomes and based on a form of governance that has survived the major social changes of recent decades. In contrast, deficiencies begin to emerge more and more clearly in lower and upper secondary education: student performance inequalities increase while the average quality of learning outcomes decreases; the share of low achievers spikes and dropout rates rise; gender gaps in learning and subject preferences appear; performance begins to suffer as a result of residential inequalities as well as latent discrimination against students with lower socioeconomic positions and migrant backgrounds; deficiencies pertaining to students with special needs surface; and, finally, inequalities between school tracks and regional educational performance levels can be seen in the fact that schools and students in the north display almost the same achievement levels as top-ranked countries in OECD PISA assessments, whereas schools and students in the south end up at the bottom of the ranking of advanced countries and economies.

Lower and upper secondary schools also suffer from outdated infrastructure and buildings, a lack of teaching and methodological skills for information and communications technology and digital literacy, and difficulties in attracting and hiring teachers in scientific and technical subjects. These secondary education problems converge with the longstanding shortcomings of Italy's teaching staff, one of the oldest, worst-paid and least-trained (and -evaluated) populations in the European landscape, who are also weakened by a polymorphic and contradictory selection and recruitment process (Abbiati et al., 2017). It is also worth noting that Italy primarily hires new teachers based on national competitions to fill vacancies that are traditionally held every 4–5 years – rather than utilising a more flexible system that adapts to emerging needs. Given these structural deficits, even the most comprehensive, integrated and ambitious of education reforms have taken little account of the results from international large-scale assessments.

The tool that instead exerts a more direct influence in steering educational policy, and particularly improvement efforts at school and district levels, is the newly introduced national assessment system (*Sistema Nazionale di Valutazione* (SNV)). Policies intentionally or unintentionally pushing for (i) decentralisation, (ii) school autonomy, (iii) school accountability and (iv) competition among schools in a quasi-market framework have played a key role in transforming the educational system in the light of the practices and outputs of national assessments. On the one hand, general reforms of the entire public sector (rationalisation, accountability and public spending cuts), together with the need to integrate national and regional governance, have been aimed at optimising the organisational structure of education as a sector. On the other hand, internal drivers of change within schools and in teaching and learning methods have set localised targets for student outcomes (Grimaldi et al., 2016).

Two main macro vectors of change have redefined the Italian education system in the past two decades: a very general influence from supranational organisations (EU agencies above all) has connected Italian educational policies to the mainstreaming of test-based assessment, whereas decentralisation and the move from government by politics to governance by technicalities have opened up local schooling spaces to a more

complex decision-making arena. These vectors of change seem to be not only thoroughly entangled but also interdependent. Transnational organisations such as the OECD (Lewis, 2020; Volante and Fazio, 2017) normatively convey a push towards ‘performance comparison’. At the same time, given the greater autonomy of individual schools and districts as a result of the move to decentralise policy production beyond the nation state and government, these local actors seek to reshape local governance models to both face novel competitive dynamics and respond to the new accountability regimes established by state legislation. Moreover, the primacy of school autonomy and the centrality of local governance as expressed in recent reformist rhetoric can be interpreted in two different ways and translated into two different policy approaches. One tends to reinforce the cycle of reforms aimed solely at fostering competition and prioritising quasi-market relations; the other tends to stress the autonomy and performance of individual schools. Considering the extent of these changes, the effects of macro, meso and micro policies converge to generate a form of ‘heterarchy’ (Ball and Junemann, 2012).

There is no doubt that international large-scale assessments based mainly, although not exclusively, on testing have provided a powerful impetus in Italy to move towards comparative and analytical evaluation, with the use of large data sets (Giancola and Viteritti, 2014). PIRLS, TIMSS, PISA, PIAAC, the International Adult Literacy Survey and its successor, the Adult Literacy and Life Skills Survey, have shifted the focus of analysis from the former object – students’ careers and achievements – to the new object of skills, literacies, abilities, and familial, contextual and micro and macro institutional factors. Indeed, the PISA findings have been under the spotlight in media debates – although not on the political agenda, as Italian governance of education rarely relies on them. More often, it is data from national rather than international assessments that are processed to generate parameters for the Italian education system in relation to education goals defined at European or UN level: for instance, the Italian Ministry of Education recently convened a special INVALSI board for the 2030 agenda for sustainable development (Falzetti, 2021). The theoretical, methodological and technical apparatus of international large-scale assessments has been absorbed into the methodologies of the SNV. The resulting isomorphism in administering tests, collecting and mining data, and sorting and presenting results affects national practices both symbolically and practically, especially when data (census results) are incorporated into national exams at the end of an education cycle through INVALSI tests or when data guide schools’ self-evaluation reports and improvement plans.

Analysis of policies, practices and political reforms

Access to education in Italy has undergone a large but extremely recent expansion, which took place later than similar shifts in other countries. This process has left educational poverty largely untouched, especially among adults, since education policies have not been effective in reaching this population despite repeated international large-scale assessments attesting to the need to do so. The other area left untouched by this shift is the pressing need to modernise the country’s education system and fill in its gaps. Instead of addressing these areas, recent government policies have invested

in students' employability (Gremigni, 2019), redefined the managerial roles performed by school principals (Taglietti et al., 2018), and launched a series of inconsistent and disjointed projects to train educators in new teaching and digital tools (Salmieri, 2019).

The most recent reform (The Good School), launched in 2015 by a centre-left government, constitutes a long-term cultural and organisational plan with multiple goals: to empower principals and teachers within a meritocratic framework (Serpieri and Grimaldi, 2015); to eliminate precarious employment conditions among teachers by establishing a new recruitment procedure; to call on schools to develop tighter links with the labour market; and to inject additional resources into the sector of education after an extended period of limited spending (Argentin and Barone, 2016). This reform does not make decisive or significant changes to curricula or teaching schedules, thus limiting its impact on key areas such as combating educational inequalities, ensuring equal opportunities for disadvantaged students, improving learning outcomes and/or further reducing dropout rates.

Before the reform (and still now to a great extent), teacher recruitment was wholly based on formal bureaucratic criteria, and educators' remuneration was differentiated according to standards based on years of service and engagement in additional tasks. With the recent reform, school principals are now authorised to select from among the accredited teachers assigned as candidates to their school district. These contracts last 3 years, and it is up to the principal to decide whether to renew them. Nevertheless, the pools of candidates from which principals can select still depend largely on allocative mechanisms beyond their local control because of the national and regional recruitment procedure. The 2015 reform also introduced the requirement that teachers engage in lifelong learning while giving those with open-ended contracts an annual EUR 500 bonus to be spent on professional education activities.

The Good School reform explicitly aimed to complete the process of expanding local school autonomy that had been initiated years before, in 1997. Although in the past schools were given only limited room to manoeuvre in terms of budget allocations, principals are now tasked with governing their own financial expenditure. They also supervise the triennial plan for educational activities that each school is required to draft. This plan sets educational objectives for the period and establishes the curricular and extracurricular activities needed to achieve those objectives; it is the principal's task to identify the financial and human resources to be used. Three-year plans have also come to serve as instruments for school accountability, since they must be drafted on the basis of schools' self-assessment reports (*Rapporto di Autovalutazione* (RAV)) containing data and figures on each school's performance as measured by the SNV. Indeed, this development represents a topical example of the cross-national achievement survey concept being adopted on a national scale. It is explicitly designed to 'allow for a comparative assessment by the pupils and their families' (Article 2 of Law 107 of 2015), which has definitively paved the way for quasi-market competition among schools. Statistical indicators thus provide a comparable starting point for identifying which educational outcomes the school should target most urgently and what levers may be used to that end. This self-diagnosis, presented in the school's RAV, constitutes the starting point for an improvement plan in which the school specifies how it intends to improve the quality and range of its educational activities.

Despite the previous reforms, the most consequential change in Italy remains the implementation of the SNV. The country's path to introducing nationwide standardised test-based practices for student assessment has been rather tortuous, with moments of inertia, acceleration and stagnation. It is only in recent years that the regulatory framework has been completed, with legislative and operational guidelines established to overcome initial uncertainties and initiate the roll-out of national testing. The process through which assessment made its way into the Italian education system parallels that of school evaluation: both are simultaneously endogenous and exogenous processes, the former stemming from rapid changes in educational policy and the growing need for data-based accountability, and the latter reflecting the pressure exerted by international surveys based on large-scale assessment. However, it is only at analytical level that these two spin-offs are separate; in practice, they are closely if not inextricably intertwined.

As mentioned above, the evaluation of the Italian education system has been conceived as necessary to reconcile school autonomy with accountability, transparency and effectiveness in relation to public users (i.e. families and pupils). Embracing the legacy of the European Education Centre established in the early 1970s, INVALSI represented a landmark of the late 1990s reform period. The policy framework is focused on counterbalancing school autonomy and transitioning towards a more marked structure of governance through a new standardised national evaluation system designed to link school assessment to large-scale international surveys while fostering national test-based surveying. Although INVALSI was established at the same time as school autonomy (Law 258 of 1999), it did not begin operating until 2007–2008 when student performance testing was rolled out nationwide. Since that time, the National Service for the Evaluation of Student Learning has made significant progress thanks to INVALSI participation in international surveys and the continuing improvement of its tools and technicalities.

The turning point in the trajectory of test-based evaluation in Italy came with the first adoption of International Association for the Evaluation of Educational Achievement and OECD survey techniques. The structure of INVALSI tests and analyses has progressively come to resemble that of PISA tests, as both rely on cognitive problems, questionnaires on students' family backgrounds, school contexts, attitudes and motivation towards education and subjects, expectations, etc. The tests administered by INVALSI therefore share the PISA approach even though they differ in certain respects: they reflect national curriculum guidelines, are based on a census and administered annually, involve lower and upper secondary grades, and are used as information sources not only for drafting national policies but also and especially by individual schools and regional aggregations. From 2008 onwards, INVALSI standardised national tests have been regularly administered in the second and fifth grades of primary schools, in the eighth grade of secondary schools (converging with the national exam for the lower secondary certificate) and in the second grade of upper secondary schools.

INVALSI testing is a political issue, however, and has been met with protest, contestation and rejection by a substantial minority of Italian parents, students and teachers (the teachers being mostly union delegates). The main critique levelled against this attempt to measure students' literacies and skills levels is deeply rooted in the idea that learning

is a process of improving individual capabilities and that these capabilities, and thus the learning process, vary from person to person and are shaped by heterogeneous cognitive methods. Critics also argue that it would be fairer and more useful to evaluate individual progress rather than measuring outputs as such – that is, to focus on the process instead of the results. Opponents of standardised national testing call for considering each student in terms of her or his distinct trajectory (which is indeed the main evaluation approach applied in classrooms during the school year) and not in order to generate problem-solving outputs leading to comparison (and therefore competition) between students, teachers and schools.

In addition to opposition from actors on the front lines, there is a large group of Italian pedagogists, intellectuals and researchers who also question standardised national testing on the grounds that excessive attention to standards, data and performance risks marginalising particular types of knowledge that are not performance oriented. A second possible risk critics identify is that the increased focus on school assessment and accountability will translate into hypercompetition between schools, classes, students and families while orienting the teaching ethos towards exasperated individualism and driving students towards insecurity or even a sort of neurotic pursuit of success.

Without delving deeper into this highly complex and competitive zeitgeist, we note that test participation has risen sharply over time, and these thorny controversies have not invalidated rounds of assessment at national level: samples remain representative even in the sites of fiercest resistance. However, this growth in participation has also been accompanied by widespread and deleterious teaching to the test practices on the part of educators; to do so, teachers must choose narrowly from among the otherwise wide range of subjects making up the national curricula, with the aid of booklets providing ready-made solutions (Grimaldi et al., 2016).

The results from standardised national student and school performance tests have gradually been integrated into the complex RAVs that principals and assigned teachers draft annually as part of new accountability-based governance. The regulations surrounding self-evaluation reports are meant to ensure transparency and public oversight. Drafters must specify the goals and targets to be achieved, and this element affords greater school marketing. However, the risk lies in the uncontrolled processing of data that underpin the self-evaluation report. Self-evaluation has thus been gradually leading the education system to a shift towards a governance model based partly on autonomy, partly on external or internal but at any rate hetero-directed processes, and partly on a quasi-market setting. A school that achieves disappointing results and fails to improve in subsequent years will not have its funding cut, nor will its teachers be subject to sanctions or wage reductions. Nevertheless, principals will be rewarded or punished on the basis of their success or failure in achieving the objectives listed in the improvement plans. The dynamics of accountability weigh directly on the school principal in that she or he holds primary responsibility 'for the efficiency and effectiveness of the management of the human, financial, technological, and material resources, as well as for the educational outcomes achieved' (Article 16 of Law 107 of 2015). At the same time, parents and families now have access to comparative information about schools and may thus choose a better-ranked school regardless of where they actually reside.

In the light of these critical points, Italy's current autonomy–evaluation pairing has been interpreted as representing an evaluating state and quasi-market scenario, that is, a variously proportioned mix of these two models. In this paradigm, assessments and evaluations can thus take on such significant weight as to essentially eliminate local autonomy and give rise to a sort of technocratic neo-bureaucratism. It is no coincidence that the Italian assessment system has given rise to a naively hypermanagerial approach to government policies (Grimaldi et al., 2016), creating an initial widespread climate of suspicion. The fact is that test-based assessment of student performances and guided school self-evaluations have become core elements of today's school life, involving principals, teachers, students and families. National tests are now an accepted national rite, producing both national data rarely used for evidence-based policymaking and local data that are widely utilised in self-assessment reports for didactic arrangements, as in the triennial plan of educational activities, and in schools' efforts to advertise their achievement results in order to attract motivated students.

In summary, educational reforms have failed to improve student achievement and performance, as they have not mitigated the systematic bias through which ascriptive factors (social origin, gender, migrant background) continue to determine school choices (Pensiero et al., 2019). This set of variables has a notable effect in predicting students' attitudes towards their educational careers. From the initial introduction of local autonomy to the Good School reform, school policies have not reversed either inequalities or lower levels of performance. We are unable to assess whether this outcome might have been different had the political actors and experts taken more and better account of the implications of cross-national achievement surveys. We can, however, state that educational policies in Italy have had little impact on educational performance in terms of macro effects

REFERENCES

- Abbiati, G., Argentin, G. and Gerosa, T. (2017), 'Different teachers for different students? Evidence on teacher-student matching and its consequences in the Italian case', *Politica Economica*, Vol. 33, No 1, pp. 13–58.
- Argentin, G. and Barone, C. (2016), 'School reform: innovation and the rhetoric of change', *Italian Politics*, Vol. 31, pp. 135–154.
- Argentin, G. and Triventi, M. (2015), 'The north–south divide in school grading standards: new evidence from national assessments of the Italian student population', *Italian Journal of Sociology of Education*, Vol. 7, No 2, pp. 157–185.
- Azzolini, D. and Barone, C. (2013), 'Do they progress or do they lag behind? Educational attainment of immigrants' children in Italy: the role played by generational status, country of origin and social class', *Research in Social Stratification and Mobility*, Vol. 31, pp. 82–96.
- Azzolini, D. and Vergolini, L. (2014), 'Tracking, inequality and education policy – looking for a recipe for the Italian case', *FBK-IRVAPP Working Paper Series*, No 2014-08 (<https://irvapp.fbk.eu/wp-content/uploads/2017/09/FBK-IRVAPP-Working-Paper-No.-2014-08.pdf>).
- Ball, S. and Junemann, C. (2012), *Networks, New Governance and Education*, Policy Press, Bristol.
- Ballarino, G., Bernardi, F., Requena, M. and Schadee, H. (2008), 'Persistent inequalities? Expansion of education and class inequality in Italy and Spain', *European Sociological Review*, Vol. 25, No 1, pp. 123–138.
- Barone, C. and Guetto, R. (2020), 'Education and social fluidity in contemporary Italy: an analysis of cohort trends', in Breen, R. and Müller, W. (eds), *Education and intergenerational social mobility in Europe and the United States*, Stanford University Press, Boston, pp. 196–223.
- Benadusi, L. and Giancola, O. (2014), 'Secondary school systems: “comprehensive” versus “selective” – a comparison in terms of equity', *Scuola Democratica*, Vol. 2, pp. 461–482.
- Borgna, C. and Struffolino, E. (2017), 'Pushed or pulled? Girls and boys facing early school leaving risk in Italy', *Social Science Research*, Vol. 61, pp. 298–313.
- Brunello, G. and Checchi, D. (2005), 'School quality and family background in Italy', *Economics of Education Review*, Vol. 24, No 5, pp. 563–577.
- Contini, D. and Scagni, A. (2011), 'Inequality of opportunity in secondary school enrolment in Italy, Germany and the Netherlands', *Quality & Quantity*, Vol. 45, No 2, pp. 441–464.
- Eurydice (2020), *Compulsory Education in Europe – 2020/21: Eurydice – Facts and figures*, Publications Office of the European Union, Luxembourg.
- Falzetti, P. (2021), *Agenda 2030: Survey on sustainable development goals through INVALSI data*, Franco Angeli, Milan.
- Ferrer-Esteban, G. (2011), 'Beyond the traditional territorial divide in the Italian education system – effects of system management factors on performance in school', *FGA Working Paper Series*, No 43, Giovanni Agnelli Foundation, Turin.

- Fornari, R. and Giancola, O. (2011), 'Policies for decentralization, school autonomy and educational inequalities among the Italian regions', *Italian Journal of Sociology of Education*, Vol. 3, No 2, pp. 114–150.
- Giancola, O. and Colarusso, S. (2020), 'Gender, social origins, and educational choices: how it really works', in Colombo, M. and Salmieri, L. (eds), *The Education of Gender, the Gender of Education – Sociological research in Italy*, Associazione "Per Scuola Democratica", Rome pp. 95–110.
- Giancola, O. and Salmieri, L. (2018), 'Education and the inclusion of immigrants – a cross-national analysis among five European countries', *Scuola Democratica*, Vol. 2, pp. 310–334.
- Giancola, O. and Salmieri, L. (2020), 'Family background, school-track and macro-area: the complex chains of education inequalities in Italy', *DiSSE Working Paper Series*, No 4, Sapienza University of Rome, Rome.
- Giancola, O. and Viteritti, A. (2014), 'Distal and proximal vision: a multi-perspective research in sociology of education', *European Educational Research Journal*, Vol. 13, No 1, pp. 47–57.
- Gremigni, E. (2019), 'Alternating school-work pathways in Italy – some remarks on the competence society', *Partecipazione e Conflitto*, Vol. 12, No 1, pp. 217–238.
- Grimaldi, E. and Serpieri, R. (2012), 'The transformation of the education state in Italy: a critical policy historiography from 1944 to 2011', *Italian Journal of Sociology of Education*, Vol. 4, No 1, pp. 146–180.
- Grimaldi, E., Landri P. and Serpieri, R. (2016), 'NPM and the reculturing of the Italian education system: the making of new fields of visibility', in Gunter, H. M., Grimaldi, E., Hall, D. and Serpieri, R. F. (eds), *New Public Management and the Reform of Education: European lessons for policy and practice*, Routledge, London, pp. 96–110.
- Hanushek, E. A. and Wößmann, L. (2006), 'Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries', *The Economic Journal*, Vol. 116, No 510, C63–C76.
- Landri, P. (2009), 'A temporary eclipse of bureaucracy – the circulation of school autonomy in Italy', *Italian Journal of Sociology of Education*, Vol. 1, No 3, pp. 76–93.
- Lewis, S. (2020), 'New topological spaces and relations of the OECD's global educational governance', in Lewis, S. (ed.), *PISA, Policy and the OECD*, Springer, Singapore, pp. 171–188.
- OECD (Organisation for Economic Co-operation and Development) (2019), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (<https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>).
- Pensiero, N., Giancola, O. and Barone, C. (2019), 'Socioeconomic inequality and student outcomes in Italy', in Volante, L. Schnepf, S. and Klinger, D. (eds), *Socioeconomic Inequality and Student Outcomes – Education policy and social inequality*, Springer, Singapore, pp. 81–94.
- Salmieri, L. (2019), 'The rhetoric of digitalization in Italian educational policies: situating reception among digitally skilled teachers', *Italian Journal of Sociology of Education*, Vol. 11, No 1, pp.162–183.

- Salmieri, L. and Rinaldi, E. E. (2020), 'Gender gaps in financial education – the Italian case', in Colombo, M. and Salmieri, L. (eds), *The Education of Gender, the Gender of Education – Sociological research in Italy*, Associazione Per Scuola Democratica, Rome, pp. 141–168.
- Santagati, M. (2021), 'Interculturalism, education and society: education policies for immigrant students in Italy', *Australian and New Zealand Journal of European Studies*, Vol. 8, No 2, pp. 6–20.
- Serpieri, R. and Grimaldi, E. (2015), 'Leader as policy device: the hybridization of head teachers in Italy', *Journal of Educational, Cultural and Psychological Studies*, Vol. 11, pp. 71–87.
- Taglietti, D., Grimaldi, E. and Serpieri, R. (2018), 'The good school / the bad head teacher: neo-managerialism and the re-making of the head teacher', in Samier, E. and Milley, P. (eds), *International Perspectives on Maladministration in Education – Theories, research, and critiques*, Routledge, London, pp. 151–168.
- Volante, L. and Fazio, X. (2017), 'PISA, policy and global educational governance', in Volante, L. (ed.), *The PISA Effect on Global Educational Governance*, Routledge, London, pp. 3–14.

CHAPTER 5

Cross-national achievement surveys and educational monitoring in the Netherlands

Anneke C. Timmermans, Harm M. Naayer, Martina R. M. Meelissen, Joyce Gubbels, Jaap Scheerens and Greetje van der Werf

Abstract

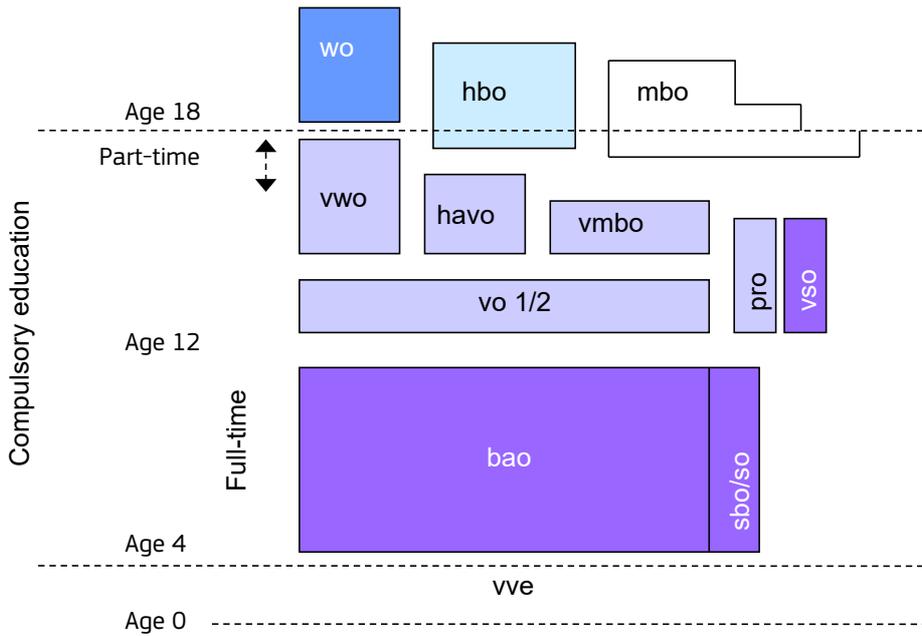
The results of Dutch students who have participated in international large-scale assessment studies have usually been above the international averages but have, nevertheless, pinpointed a number of weaknesses in the Dutch educational system. The Trends in International Mathematics and Science Study, the Progress in International Reading Literacy Study and the Programme for International Student Assessment show patterns of declining or, at best, stable performance levels for the subjects mathematics, science and reading (literacy) over a 20-year period. In contrast, the performance of Dutch students in the International Civic and Citizenship Education Study improved between 2009 and 2016. We show that Dutch policymakers are well aware of results of educational achievement surveys, as these results are presented frequently in yearly evaluations of the quality of the Dutch educational system by the Dutch Inspectorate of Education and are often named in political debates. Moreover, these results are often mentioned in problem statements for educational policy changes. Due to the governance structure, with high levels of autonomy for schools and school boards, educational policies, including those based on large-scale educational achievement results, are almost impossible to evaluate. Nevertheless, the declining trend has sometimes been changed into a situation in which the performance levels have remained stable; however, a systematic increase in performance has not (yet) been observed.

Introduction

In this chapter, we provide an overview of results from international large-scale achievement studies in which the Netherlands has taken part during the past 20 years. Before we give an overview of the highlights of the results and development of Dutch students in the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), the Programme for International Student Assessment (PISA) and the International Civic and Citizenship Education Study (ICCS), we explain the structure and governance of the Dutch educational system. In describing the results of international achievement surveys, general trends in performance levels over time will be discussed, as well as variation in performance and differences between groups. Thereafter, we describe a selection of educational policies implemented during the past two decades that are related to or a response to the results of one or more achievement surveys. The chapter concludes with a reflection on the effects of these educational policies and the difficulties in evaluating educational policies in the Netherlands.

The Dutch educational system

The structure of the educational system in the Netherlands is depicted in Figure 5.1. There are two main features of the Dutch educational system that are particularly worth mentioning (Scheerens et al., 2019). First, there is a strong strand of vocational education, manifested in specialised vocational schooling at both junior and senior secondary levels (pre-vocational secondary education (vmbo) and vocational education (mbo)) with good labour market outcomes (OECD, 2016a). Second, the Dutch system is characterised by strong structural differentiation or tracking at secondary level, combined with relatively early selection for a particular track, at the age of 12 (van de Werfhorst et al., 2015).

Figure 5.1: The Dutch school system

NB: **bao**, mainstream primary education; **havo**, general secondary education; **hbo**, higher professional education; **mbo**, vocational education; **pro**, practical training; **sbo**, special primary education; **so**, special needs education; **vmbo**, pre-vocational secondary education; **vso**, special needs education secondary level; **vve**, early childhood education; **vwo**, pre-university education; **wo**, university education.

Source: Derived from Scheerens et al. (2019).

Full-time education in the Netherlands is compulsory from the age of 5 to the age of 16. Almost all children (98 %) enter primary education voluntarily at the age of 4 (Ministry of Education, Culture and Science, 2016), and many children make use of preschool facilities from the age of 2.5 years. Students who have not acquired a starting qualification for the labour market when they leave full-time education are required to attend part-time education until they have acquired a starting qualification (International Standard Classification of Education (ISCED) level 3) or until the age of 18.

The Dutch education system is divided into three levels: primary education, secondary education (general secondary, and junior and senior vocational education) and tertiary education (higher vocational education and university) (Scheerens et al., 2019). Given the scope of this chapter, we focus on primary and secondary education only.

Primary schools in the Netherlands cater for children aged 4–12 years. Primary school classes are usually based on eight year groups. Grade retention during primary education is possible, and whether a child repeats a grade is decided by the school the child attends (Driessen et al., 2014). Compared with the average retention rate across Organisation for Economic Co-operation and Development (OECD) countries (7.3 %), grade retention occurs rather frequently in the Netherlands (22.4 %) (Ikeda and Garcia, 2014; OECD, 2010, 2011).

During primary education, all schools are required to monitor the development of their students by means of regular testing and administration of monitoring systems. Based on the capabilities of pupils as apparent from test results from monitoring systems in the final years of primary education, teachers give a recommendation on the type of secondary school that would be the most appropriate for the student (Timmermans et al., 2018). This recommendation is decisive in students' transfer to different types of secondary education. In addition, at the end of primary education, all schools must administer a standardised primary school leavers' attainment test to assess the educational potential of their pupils (Scheerens et al., 2019). When the students' test scores indicate that they could go to a higher level of secondary education, primary schools are required to reconsider and potentially adjust the track recommendation to a higher level.

The system of secondary education is strongly stratified, both within and between schools (Scheerens et al., 2019). Dutch secondary education is divided into the following tracks.

- Practical training (pro) is for students aged 12–16 years with special educational needs (ISCED level 2).
- vmbo is for students aged 12–16 years (ISCED level 2). This type of education consists of four different tracks: the basic vocational track, the advanced vocational track, the theoretical track and the mixed track. Depending on the track, students prepare for various levels of mbo.
- Senior general secondary education (havo) is for students aged 12–17 years (grades 1–3, ISCED level 2; grades 4 and 5, ISCED level 3). It prepares students for higher professional education (hbo).
- Pre-university education (vwo), the highest track level, is for students aged 12–18 years (grades 1–3, ISCED level 2; grades 4–6, ISCED level 3). It prepares students for future educational careers at universities.

In secondary schools, students are placed into one or sometimes two adjacent tracks (Korpershoek et al., 2016). For many students, the track in which they were placed in grade 1 of secondary education is strongly decisive in their educational careers in secondary school and beyond (de Boer et al., 2010; Timmermans et al., 2013). Although intermediate upward and downward mobility is possible, its extent is limited, as after 3 years in secondary education 79.3 % of students are still in the track that the teachers recommended (Inspectorate of Education, 2021a).

The main policy for children who are in need of special care is to have them attend regular primary or secondary schools. Under this inclusive education policy, every school board has the responsibility to provide appropriate education for every enrolled pupil, regardless of his or her specific needs for support. By collaborating with other school boards, mostly at regional level, schools are required to arrange educational provisions in such a way that every child can be educated taking into account their special educational needs (van Leeuwen et al., 2009). Depending on the type and severity of the special educational needs, a child can be admitted to special primary education (sbo) schools or the even more specialised special needs education (so) schools in primary and secondary education.

Governance in Dutch education

School autonomy is grounded in the principle of ‘freedom of education’, which has been guaranteed by Article 23 of the Dutch Constitution since 1917 (Ministry of Education, Culture and Science, 2016). Freedom of education gives the right to any natural or legal person to set up a school, to organise teaching, and to determine the educational, religious, or ideological principles on which teaching is based. This constitutional arrangement puts public and private schools on an equal footing, with all schools receiving public funding provided that they meet the requirements for schools in the primary or secondary education sector. Parents have free school choice, and funding ‘follows the student’; both of these lay the foundation for potentially strong competition among schools (Nusche et al., 2014, pp. 20–21).

The Netherlands has one of the OECD’s most devolved education systems, with schools enjoying a high degree of autonomy (Kelly et al., 2020; OECD, 2016a). Public and private schools (or school boards) may autonomously decide how and, to a large extent, when to teach the core objectives of the Dutch curriculum based on their religious, philosophical or pedagogical views and principles. Responsibility for education is shared almost entirely between school boards and the central government. Councils representing school boards at various levels form an intermediate level of educational administration and have considerable influence on collective labour agreements and educational policies. With regard to ensuring teaching quality, the Dutch government describes the distribution of responsibilities for educational reform as follows: the government will establish the objectives of the policy measures, whereas the field itself will decide how best to pursue those objectives.

Quality assurance at school and system levels

The Ministry of Education, Culture and Science has responsibility for the overall quality of the education system, but it does not interfere with the organisation of individual schools. The ministry’s responsibilities relate mainly to setting legislation and determining the structure and funding mechanisms of the education system. It can also control the system by setting quantitative or qualitative standards, attainment targets and examinations. The Minister for Education, Culture and Science is also responsible for the evaluation of the quality of education, which is carried out by the Dutch Inspectorate of Education (Inspectorate of Education, 2010; 2020a).

Schools and school boards are responsible for ensuring quality at school level, and the Inspectorate of Education checks that they do so effectively by monitoring student performance on a yearly basis, and less frequently by conducting school visits (Inspectorate of Education, 2017a,b). The Dutch Inspectorate of Education visits each school at least once every 4 years to ascertain whether schools, both public and private, provide the expected quality of education. The Dutch inspection framework is based on a risk-oriented approach, in which high-risk schools are inspected more frequently, whereas low-risk schools are inspected less frequently. The results of the school visits are communicated to the school boards and published online for horizontal accountability purposes (Ehren et al., 2005).

Quality assurance at system level is also one of the main tasks of the Inspectorate of Education. There is a well-developed system, coordinated by the inspectorate, of conducting national assessments to measure the performance levels, attitudes and curricula in the final year of primary education in various subjects (some of which are not covered by international achievement surveys, such as the arts, English and physical education). Two national assessments are conducted every school year, each covering a specific subject, with the aim of assessing the current performance level of the school leavers in regular primary schools and comparing performance with previous assessments on the same subject. Both national assessments include a sample of regular primary schools, but one also includes a sample of special primary schools. National assessments are not yet available in secondary education, although calls for national assessments in secondary education are becoming stronger. Together with results of school visits and thematic studies conducted by the Inspectorate of Education, national assessments and international comparative studies form the basis for monitoring the quality of the Dutch educational system (Inspectorate of Education, 2021a).

International achievement surveys

The Netherlands participates frequently in various international achievement surveys. Therefore, this chapter describes only a modest selection of the highlights from over 20 years of participation. We will first describe the results for students in primary education (TIMSS and PIRLS) and thereafter the results for students in secondary education (PISA and ICCS).

Trends in International Mathematics and Science Study

The main results (national means, standard errors and international benchmarks) of Dutch grade 4 children in the TIMSS data collections between 1995 and 2019 in the subjects mathematics and science are provided in Table 5.1. Dutch students consistently score, for mathematics and science, above the centre point of the TIMSS scale (500), which is based on the international average achievement of pupils in a specific group of countries set in 1995 (Mullis et al., 2020). Compared with several other countries frequently participating in TIMSS, the scores of Dutch students seem relatively stable over time. Nonetheless, compared with 2011, the scores in 2015 were significantly lower for mathematics and science (Mullis et al., 2016), although the differences were relatively small. Although the mathematics performance of Dutch students in 2019 was significantly higher than in 2015 and almost back to the level of 2011, the performance levels of 1995 were never reached again. Moreover, the level of science performance in 2019 remained relatively low (Meelissen et al., 2020a).

In addition, compared with many other participating countries, the variance in Dutch students' achievement is relatively small. This is particularly clear from the results regarding the international benchmarks set in TIMSS (see Table 5.1); a common pattern is that the percentages of Dutch students who at least reached the lowest and intermediate mathematics benchmark levels are higher than the international average, whereas the percentages of Dutch students who reached the advanced and high mathematics

benchmark levels are relatively low (Meelissen and Punter, 2016). In 2019, however, the percentage of Dutch students who reached the advanced mathematics benchmark level was similar to the international median.

With regard to subgroups of students, boys consistently outperform girls in mathematics. This performance difference has been stable since the Netherlands first participated in TIMSS. In contrast, boys and girls have performed at the same level in science since 2015. Before 2015, boys outperformed girls (Mullis et al., 2020).

Table 5.1: Maths and science achievement results of Dutch fourth-grade students, 1995–2019 TIMSS

Mathematics	1995	2003	2007	2011	2015	2019
Mean	577 ⁽¹⁾	540 ⁽²⁾	535 ⁽²⁾	540 ⁽²⁾	530 ⁽²⁾	538 ⁽²⁾
Standard error mean	3.4	2.1	2.1	1.7	1.7	2.2
Percentage above benchmark: advanced ⁽³⁾	—	6	7	5	4	7
Percentage above benchmark: high ⁽³⁾	—	29	42	44	37	44
Percentage above benchmark: intermediate ⁽³⁾	—	65	84	88	83	84
Percentage above benchmark: low ⁽³⁾	—	82	98	99	99	98
Science	1995	2003	2007	2011	2015	2019
Mean	557 ⁽¹⁾	525	523 ⁽²⁾	531 ⁽²⁾	517 ⁽²⁾	518 ⁽²⁾
Standard error mean	3.1	2.0	2.6	2.2	2.7	2.9
Percentage above benchmark: advanced ⁽⁴⁾	—	3	4	3	3	4
Percentage above benchmark: high ⁽⁴⁾	—	32	34	37	30	33
Percentage above benchmark: intermediate ⁽⁴⁾	—	83	79	86	76	76
Percentage above benchmark: low ⁽⁴⁾	—	99	97	99	97	96

(1) Significantly above the international average: mathematics mean 529 with standard error 0.7 (Mullis et al., 1997); science mean 524 with standard error 0.7 (Martin et al., 1997).

(2) Significantly higher than the centre point of the TIMSS scale (500 points).

(3) Achievement in mathematics is described in terms of four points along the scale that act as international benchmarks: advanced (applying and justifying their mathematical understanding, score of 625), high (score of 550), intermediate (score of 475) and low (basic mathematics knowledge, score of 400) (Mullis et al., 2020).

(4) Achievement in science is described in terms of four points along the scale that act as international benchmarks: advanced (understanding of a variety of topics in life science, physical science and Earth science, score of 625), high (score of 550), intermediate (score of 475) and low (limited knowledge of science facts, score of 400) (Mullis et al., 2020).

Progress in International Reading Literacy Study

The main results (national means, standard errors and international benchmarks) of the Dutch grade 4 children in the PIRLS data collections between 2001 and 2016 for reading are provided in Table 5.2. Besides the national means, we also provide the separate scores for two reading purposes (reading for literary experience and reading to acquire information) and two reading processes (retrieving and inferencing, and interpreting, integrating and evaluating). Dutch students consistently score above the centre point of the PIRLS scale, which is based on the international average achievement of students in a group of countries set in 2001 (Mullis et al., 2017). After a significant decline in performance between the measurements in 2001 and 2006 (Mullis et al., 2007), Dutch students show a steady pattern of performance levels, with only very minor variations over subsequent measurements.

With regard to the two reading purposes, Dutch students' scores are at similar levels in the domains 'reading for literary experience' and 'reading to acquire information'. With regard to the reading processes, in 2006 and 2011, Dutch students scored significantly higher on retrieving information and inferencing than on the more complex interpreting, integrating and evaluating (Mullis et al., 2017).

Also in relation to reading, and compared with many other participating countries, Dutch students' achievement variation is relatively small. The same pattern as that observed for mathematics and science in TIMSS is also apparent for reading in PIRLS; the percentages of students who at least reached the lowest and intermediate benchmark levels are higher than the international average, whereas the percentages of Dutch students who reached the advanced and high reading benchmark levels are relatively low. Since 2006, the percentages of Dutch students reaching the international benchmarks have been relatively stable, with only minor changes (Mullis et al., 2017). Moreover, Dutch students' motivation to read is lower than that of students in (most) other participating countries.

With regard to student subgroups, since 2001 girls have scored higher in reading during every data collection than boys. It seems that the gap in performance of approximately 10 points is relatively stable. More influential factors are home resources (a combined index based on the number of (children's) books available in the house and parental education and occupation) and whether the language of the test is spoken at home. In 2016, a 36-point difference was observed between the 33 % of students with a high level of resources and the 67 % of students with an average level of resources, which is a substantially smaller difference than the participating countries' average difference of 63 points observed in PIRLS between these groups. Moreover, in PIRLS, the mean score of the group of students who sometimes speak the test language at home is similar to that of the students who (almost) always speak the test language, with only those who never speak the test language at home having a lower score. In the Netherlands, a 22-point difference was observed between the 82 % of students who always or almost always spoke Dutch at home and the 18 % of students who only sometimes or never spoke Dutch at home (Mullis et al., 2017).

Table 5.2: Reading results of Dutch fourth-grade students, 2001–2016 PIRLS

Subject			2001	2006	2011	2016
Reading (comprising all reading dimensions)	Mean		554 ⁽¹⁾	547 ⁽¹⁾	546 ⁽¹⁾	545 ⁽¹⁾
	Standard error mean		2.5	1.5	1.9	1.7
	Percentage above benchmark: advanced ⁽²⁾		10	6	7	8
	Percentage above benchmark: high ⁽²⁾		54	49	48	48
	Percentage above benchmark: intermediate ⁽²⁾		92	91	90	88
	Percentage above benchmark: low ⁽²⁾		9	99	100	99
Reading purposes dimension	Literary experience	Mean	555	546	545	546
		Standard error mean	2.6	1.8	2.4	1.7
	Acquire and use information	Mean	554	549	547	545
		Standard error mean	2.6	1.6	1.9	1.9
Reading processes dimension	Retrieving and inferencing	Mean	559	554	549	546
		Standard error mean	2.6	1.7	2.1	2.0
	Interpreting, integrating and evaluating	Mean	552	542	543	544
		Standard error mean	2.5	1.7	1.9	1.7

⁽¹⁾ Significantly higher than the centre point of the PIRLS scale (500).

⁽²⁾ PIRLS describes achievement in terms of four points along the scale that act as international benchmarks: advanced (score of 625), high (score of 550), intermediate (score of 475) and low (score of 400).

Programme for International Student Assessment

The Netherlands has participated in all PISA cycles since 2003. The main results (national means) of the Dutch grade 9 students in the PISA data collections between 2003 and 2018 for the subjects mathematics, science and reading literacy are provided in comparison with the means of the OECD and the first 15 Member States of the European Union (EU-15) in Table 5.3. Dutch students showed a significant decline in performance in all three subjects; however, the decline was not equally strong in all three subjects and did not occur at the same time. Therefore, we will briefly describe the trend in each of the three subjects measured in PISA.

Dutch 15-year-olds scored, on average, 519 points in mathematics in the latest PISA 2018 (Gubbels et al., 2019; OECD, 2019a). This score is significantly higher than the OECD average (mean: 492 points) and the average of the EU-15 (mean: 496 points). Similar to the TIMSS results, the level of maths achievement of Dutch 15-year-olds gradually declined between 2003 and 2012; the decline in performance between

2012 and 2015 was significant and slightly stronger. Since 2003, the mathematics performance level has decreased by one quarter of a standard deviation.

Dutch 15-year-olds scored 503 points on average in science (Gubbels et al., 2019; OECD, 2019a). This score is significantly higher than the OECD average (mean: 491 points) and the average of the EU-15 (mean: 492 points). The performance levels of Dutch students in science were rather stable between 2003 and 2012 but showed a significant decline of 13 points in 2015. Because the decline in science performance in the Netherlands since 2013 has been stronger than the general decline in OECD countries (van der Molen et al., 2019), Dutch students are no longer in the lead.

In PISA, reading literacy is defined as students' capacity to understand, use, evaluate, reflect on and engage with texts in order to achieve their goals, develop their knowledge and potential, and participate in society (OECD, 2019a); this is different from the curriculum-based interpretation of reading in PIRLS (Mullis et al., 2017). In reading literacy, Dutch 15-year-olds scored 485 points on average (Gubbels et al., 2019; OECD, 2019a), which is comparable to the OECD average of 489 points but significantly lower than the average score of 492 points of the EU-15. In addition, between 2015 and 2018 there was a significant decline of 18 points in the reading literacy performance of Dutch students. In line with the PIRLS results, reading motivation tends to be lower in the Netherlands than in other countries.

The decrease in performance of Dutch students in all three subjects seems to be moderated by the level of parental education (Aalders et al., 2020). A stronger decline in performance between 2003 and 2018 was found for students whose parents have a low level of education, which implies that the differences in performance between students whose parents have a low level of education and those whose parents have a high level of education increased over time. Despite growing differences, the strength of the association between performance and socioeconomic status is not significantly different from the OECD average (OECD, 2019b). Moreover, the number of resilient students – those who are in the bottom quarter of the PISA index of economic, social and cultural status in their own country but who score in the top quarter for reading in that country – is above the OECD average (OECD, 2019b). Furthermore, the decline in performance was stronger for boys with an immigrant background than for boys without an immigrant background (Aalders et al., 2020). This applied to all three subjects and indicates that, for boys, the influence of an immigrant background increased over time. For girls, the pattern was found only for reading literacy.

Table 5.3: Maths, science and reading results of Dutch 15-year-olds and the OECD and EU-15 means, PISA 2003–2018

Subject		2003	2006	2009	2012	2015	2018
Mathematics	OECD	499	494	495	494	490	492
	EU-15 ⁽¹⁾	501	498	495	496	496	496
	The Netherlands	538	531	526	523	512	519
Science	OECD	499	498	501	501	493	491
	EU-15	500	503	502	504	497	492
	The Netherlands	524	525	522	522	509	503
Reading literacy	OECD	494	489	493	496	493	489
	EU-15	497	492	494	499	498	492
	The Netherlands	513	507	508	511	503	485

(¹) EU-15 refers to the 15 countries that belonged to the European Union in 1995 or before and that have participated in all PISA data collections since 2006. They are the following countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.

International Civic and Citizenship Education Study

The Netherlands participated in the first cycle of ICCS in 2009 and in its follow up in 2016. Table 5.4 contains the main results (national means, standard errors and proficiency levels) of Dutch grade 8 students, aged approximately 14 years, in both ICCS data collections. The International Association for the Evaluation of Educational Achievement studies of civic and citizenship education have recognised that, in order to participate effectively as citizens, young people need to possess a knowledge base and the capacity to reason about the institutions, events, actions and processes that exist in their civil and civic communities, as well as to develop and justify views and attitudes towards those things (Schulz et al., 2016). Measurements in ICCS include four content domains: civic society and systems, civic principles, civic participation, and civic identities. In addition to the mean score measuring civic and citizenship knowledge and understanding, ICCS also measures a wide range of topics concerning attitudes, skills and (future) intentions related to adolescent civic and citizenship development. As an illustration, we have included the scores for three attitudinal scales and two scales concerning ‘civic engagement’ in Table 5.4 (Daas and Mennes, 2018; Isac et al., 2012; Schulz et al., 2017).

In the measurement cycle of 2009, Dutch students scored significantly lower than the average (514) of all participating European countries (Kerr et al., 2010; Maslowski et al., 2012; Schulz et al., 2010). The results from ICCS 2016 show that the average score of Dutch students increased substantially (from 494 to 523; Munniksma et al., 2017, Schulz et al., 2017). It seems that the large increment in performance can mostly be attributed to students in the senior general and pre-university tracks in secondary

education. Students in the vocational tracks still perform very weakly compared with the international average (Munniksmas et al., 2017).

Gender differences in the citizenship knowledge of Dutch students are in line with those in all other countries. Girls outperform the average score of boys in both the 2009 and the 2016 measurements. Compared with other participating countries, the difference in performance between girls and boys is ranked in the top five of the smallest gender gaps (in absolute value) (Schulz et al., 2010, 2017).

The 2009 and 2016 measurements among Dutch students show that the scores of most of the attitudinal components (scales) can also be considered to be subpar compared with the ICCS average. In this brief overview, we highlight the scores considering the scales 'Equal rights for all ethnic and racial groups' (within Dutch society) and 'Importance of conventional citizenship'. For both scales, Dutch students show a more negative attitude than students in other (European) countries. This also applies to the civic engagement of Dutch students: they seem to have less interest in political and social issues than their peers in other (European) countries. Moreover, they report lower scores on the citizenship self-efficacy scale (Maslowski et al., 2012; Munniksmas et al., 2017; Schulz et al., 2010, 2017).

All surveys taken together indicate that the results of Dutch students are usually above the international averages but, nevertheless, pinpoint a number of weaknesses in the Dutch educational system. A pattern of declining or, at best, stable performance levels for the subjects mathematics, science and reading (literacy) over a 20-year period was observed. ICCS shows contrasting findings, with scores that are improving over time but still remaining below the international average.

Table 5.4: Achievement results of Dutch eight graders, ICCS 2009 and 2016

Subject		2009	2016	
General knowledge and understanding	Mean	494	523	
		Standard error mean	7.6	4.5
	Percentage above proficiency level	A ⁽¹⁾	24	36
		B ⁽¹⁾	47	68
		C ⁽¹⁾ (basic)	85	91
		D ⁽¹⁾	100	99
		Below D ⁽¹⁾	n.a.	100
Civic attitudes	Equal rights for all ethnic and racial groups	Mean	47 ⁽²⁾	49 ⁽²⁾
		Standard error mean	0.3	0.3
	Equal gender rights	Mean	51 ⁽³⁾	52 ⁽³⁾
		Standard error mean	0.5	0.3
	Importance of conventional citizenship	Mean	47 ⁽²⁾	48 ⁽²⁾
		Standard error mean	0.3	0.3
Civic engagement	Interest in political and social issues	Mean	46 ⁽²⁾	n.a.
		Standard error mean	0.3	n.a.
	Citizenship self-efficacy	Mean	48 ⁽²⁾	48 ⁽²⁾
		Standard error mean	0.6	0.2

⁽¹⁾ General knowledge was described in terms of four levels: D with scores between 311 and 394 points, C with scores between 395 and 478 points, B with scores between 479 and 562 points, and A with scores of 563 points or higher. Students with scores below 311 points were categorised as 'below level D' (Munniksma et al., 2017).

⁽²⁾ Significantly below ICCS average (the centre point for all scales in 2009 was set at $M = 50, SD = 10$).

⁽³⁾ Significantly above ICCS average (the centre point for all scales in 2009 was set at $M = 50, SD = 10$).

Education policy monitoring

The Dutch government aims to be among the top five knowledge economies in the world (Dekker, 2013; Plasterk and van der Hoeven, 2009). High-quality education and well-educated students, especially in mathematics and science, are necessary to achieve this goal. This objective is one of the main reasons why the Netherlands participates in international achievement studies (Meelissen et al., 2020b). Policymakers in education in the Netherlands are well aware of the results of Dutch students in international comparative studies, as they are frequently referred to in yearly reports about the state of education in the Netherlands (see, for example, Inspectorate of Education, 2020b, 2021a), in letters presenting summaries of results from the Minister for Education, Culture and Science to the House of Representatives (see, for example, Slob, 2019) and in debates about educational policy in the House of Representatives.

Given that multiple sources are available on the performance of students in the Netherlands (e.g. international comparative studies and national assessments), it is very difficult to pinpoint the exact changes in Dutch educational policies that are explicitly related to the results from one or more international comparative studies in which the Netherlands participated. Nevertheless, we have tried to provide an overview of responses and policy changes in which the results of educational achievement surveys were explicitly mentioned during the policy design process or in the problem description leading to policy change. In this section, we discuss three strands of educational policies that have been explicitly associated with international survey results.

Improving performance in mathematics and reading literacy, and interest in reading

A relatively small decrease in reading and mathematics achievement in PIRLS and TIMSS in 1995–2007 (Meelissen et al., 2020b), in combination with declining performance levels in PISA in the Netherlands, drove policymakers to pay more attention to the basic subjects of reading and numeracy in primary and lower secondary education. This approach led to the introduction of reference levels for mathematics and the Dutch language in 2010 by law (SLO, 2009). Reference levels, including minimum and ambitious levels, were applied to primary education, secondary education and senior secondary vocational education, and describe what knowledge and skills students should have acquired with regard to the Dutch language and mathematics at several points in their academic careers (Inspectorate of Education, 2012). The reference levels were expected to promote more efficient and effective programmes in schools, to create better connections between subsequent levels of education, to prevent unnecessary repetition and to detect gaps in the curriculum (Brandt et al., 2010; SLO, 2009). In particular, setting ambitious reference levels was expected to challenge talented students.

The results of PIRLS in 2016 and PISA in 2018 triggered initiatives to improve Dutch students' reading literacy. The results of these surveys were explicitly described in the problem statements of these initiatives (Gubbels et al., 2017, 2019). The steady but worryingly poor performance of Dutch students in PIRLS 2016, together with a

decrease in the performance of Belgian students (in Flanders), led to the establishment of a Flemish–Dutch language council in 2018. With regard to the Dutch students, the low percentage of students reaching advanced benchmark levels (8 %) and low levels of interest in reading compared with most other countries in PIRLS 2016 were the main triggers to conclude that the quality of Dutch reading comprehension education needed to be improved. The task of this language council was to assess bottlenecks in reading curricula and to formulate concrete recommendations to strengthen education in reading comprehension based on relevant and recent scientific evidence. The council implemented an action plan to improve reading comprehension and interest in reading (Pereira and Nicolaas, 2019). In short, schools should choose an approach to teaching reading comprehension that is guided by a sense of urgency; reading comprehension is a crucial skill for school success and participation in society. The approach of the schools is described in a school policy document focusing on the subjects language and reading, with concrete and shared plans to stimulate the development of the reading comprehension of all children. In the classes, effective reading instruction should be used, and teachers should be aware that stimulating an interest in reading is an essential component of effective reading comprehension instruction (Pereira and Nicolaas, 2019).

In addition, the results of PIRLS 2016 and PISA 2018 initiated the launch of the Leesoffensief (Reading Offensive) initiative in 2020 (Raad voor Cultuur and Onderwijsraad, 2019; van Engelshoven and Slob, 2020). This initiative is aimed explicitly at improving students' interest in reading and only indirectly at improving reading comprehension. Moreover, this initiative is specifically targeted at students lagging behind in reading comprehension in primary education and students in the vocational tracks in secondary education, and is focused on key points in the educational careers of students at which an interest in reading is at risk. The Leesoffensief aims to secure a systematic place for reading in a new national curriculum (see 'Rethinking the Dutch curriculum and educational system reforms' below), with a particular focus on 'interpreting, integrating and evaluating' (see PIRLS results), assessing whether the reference levels with respect to reading are at a sufficiently ambitious level and supporting programmes designed to encourage an interest in reading.

Promoting excellence among primary and secondary school students

Despite various programmes to stimulate excellence among students in primary and secondary education having been implemented since 2008 (Dijkma, 2008), this has not yet resulted in growing rates of students reaching the higher proficiency levels in comparative international achievement surveys. For example, PIRLS 2011 and TIMSS 2011 showed relatively very low percentages of students in the Netherlands reaching high and advanced benchmarks compared with other high-achieving countries. These results boosted the discussion on supporting talented and high-achieving students and led to the introduction of several projects focused on talent development in education (see, for example, Doolaard and Harms, 2014; Meelissen et al., 2020b). For example, the Secretary of State made explicit reference to the TIMSS 2011 and PIRLS 2011 results when announcing plans to stimulate excellence in primary and secondary education (Dekker, 2013). The strong focus on low- and average-performing students at both levels of education may have meant that the most talented students did not get the opportunity to excel, since too many students were not sufficiently challenged at school.

Excellence is stimulated through a variety of financial and communicative instruments, for example by offering grants to collaborating schools to implement (or develop) programmes identifying talented students early, promoting national projects including conferences, stimulating collaboration between schools, universities and companies, recognising performance, and ensuring the professional development of teachers and school leaders (Dekker, 2013; Dijkma, 2008). The most recent call for proposals was in 2019. These financial and communicative instruments are expected to stimulate and offer support to primary and secondary schools to increase the attention given to talented students and to realise more evidence-based programmes for talented students.

Rethinking the Dutch curricula and educational system reforms

The whole curriculum (core objectives) for primary education and the curriculum for the first 2 years of secondary education are currently undergoing significant revision and are expected to be updated (e.g. by including digital literacy) in the coming years. The current curricula are considered to be no longer up to date, and several councils have requested an overall revision of the Dutch curricula (SLO, 2014; WRR, 2013). However, international achievement survey results are not the direct reason for this revision. Suggestions for the review were collected from the public by a large group of teachers and school principals, with support from education specialists and curriculum developers. Currently, these suggestions are being reviewed and have not yet been formally implemented.

In this larger debate about the Dutch curricula, the somewhat disappointing results of ICCS 2009 and to a lesser extent those of ICCS 2016 (see, for example, Maslowski et al., 2012, Munniksma et al., 2017) provided input into a broader political and societal discussion about the state of Dutch citizenship education and the goals that should be pursued. Although Dutch schools have been required by law to actively promote students' citizenship and social inclusion since 2006, a shared conclusion from both cycles of ICCS is that a large proportion of schools have no clear vision about how to translate civic and citizenship topics into curricula. Some schools offer separate subjects (i.e. the subject 'study of society' or 'citizenship studies') in secondary education, but most schools address citizenship through a cross-curricular programme in subjects such as geography, history and economics, and/or through extracurricular activities and events (Dijkstra et al., 2018; Maslowski and van der Werf, 2013). The Dutch government therefore issued a new 'Citizenship law', which was formally implemented in late 2021. The general assignment for primary and secondary schools has been extended and now includes, for instance, 'the promotion of knowledge about and respect for: diversity in religion, ethnic background, political orientation, handicaps, and sexual orientation'⁽¹⁰⁾. Moreover, the proposed law includes addenda that should support and facilitate the subject within schools and offer teachers resources and further opportunities for professionalisation.

Only very recently, the Education Council (2021) published advice on a more rigorous restructuring of the educational system. Concerns regarding equity of education are the primary basis for this plea. The analyses of Aalders et al. (2020) based on the

⁽¹⁰⁾ Law on Primary Education, 2021 (<https://wetten.overheid.nl/BWBR0003420/2021-10-01>).

Dutch PISA results, which show increasing gaps in performance between Dutch students whose parents have a low level of education and Dutch students whose parents have a high level of education, form a central part of the problem description. When put in the international context, despite the highly tracked character of Dutch secondary education and perhaps due to strong central examination, the level of inequity in the performance of Dutch students is still below the international mean (OECD, 2016b; Scheerens et al., 2019; van de Werfhorst et al., 2015). The council recommends a transition from primary to secondary education without selection. Students with various performance levels should stay together for the first 3 years of secondary education. This does not imply that all students should receive education at the same level. Teachers are expected to differentiate between students and personalise their instruction. Selection will take place only after these 3 years, based on standardised testing and a broad spectrum of information. Based on these data, school teams may formulate a recommendation for (potentially fewer) tracks in the upper levels of secondary education.

Analysis of policies, practices and political reforms

When it comes to an evaluation of the effectiveness of the educational policies in the Netherlands, there is a striking consensus among all evaluators and reviewers on these policies (Scheerens et al., 2019). They all agree on a lack of coherence, clear frameworks and a basis in empirical evidence, and limited evaluability of the way schools use extra funding and work towards the rather general policy objectives (see, for example, Driessen and Mulder, 1999; Mulder and Meijnen, 2013). The implicit message is that school autonomy has gone too far in the Netherlands and is preventing policies that are effective and efficient (Scheerens et al., 2019). For example, Mulder and Meijnen (2013) are very explicit in their recommendations on having clearer policy targets from the central government, more explicitly planned programmes, stricter accountability requirements and better conditions for programme evaluation. The inefficiency in educational policy is part of a larger syndrome in Dutch education, in which innovation and reform are framed as ‘bottom up’, leading to many fragmented local initiatives in which the wheel is reinvented over and over again (Scheerens, 2013). Despite the recommendations in practically all evaluation studies, the counsel to make better use of evidence-based comprehensive school reform programmes has never been followed up in a consistent way (Scheerens et al., 2019). The Dutch government sticks to the idea that the ‘what’ of educational policies is determined by the government, whereas the ‘how’ is to be determined by schools and school boards.

These evaluation problems are also explicitly noted in relation to several of the educational policies that were described in the previous section (see, for example, Dijkstra et al., 2018; Doolaard and Harms, 2014). For example, the finding that a large proportion of schools has no clear vision about how to translate civic and citizenship topics into curricula is explicitly linked to the high degree of autonomy of schools in implementing citizenship education (Dijkstra et al., 2018). Moreover, Doolaard and Harms (2014) indicated that the financial and communicative instruments used for educational policies to improve the quality of education for talented students maybe logical given the

Dutch governance structure but that it will inevitably lead to variation between schools in the focus on education for talented students and unequal benefits from funding opportunities. The same 'bottom-up' approach is also clearly visible in the initiatives to improve reading comprehension and interest in reading. Despite recommendations on several key themes, schools are encouraged to develop an approach that fits with their own vision on education and their specific student population (Pereira and Nicolaas, 2019). We will briefly illustrate how such policies lead to variation between schools and limit the potential effectiveness of educational policies for promoting excellence and improving equity.

Regarding policies promoting excellence among students, in the most recent large-scale achievement surveys the percentage of students with an excellent performance level remains relatively low compared with other high-achieving countries, notwithstanding years of policies that provided additional funding to schools and school boards. Schools in Dutch secondary education often spend the additional financial resources for stimulating excellence on a variety of interventions, including opportunities for students to join national competitions and attend masterclasses at universities and ICT facilities, projects to broaden the curriculum, and additional support with respect to socioemotional development and learning how to learn (Boers-Müller et al., 2013). Only a few of these interventions are explicitly related to the subjects reading literacy, mathematics and science, which may explain why the percentages of students reaching advanced benchmark levels in achievement surveys have not increased. Moreover, the relatively low performance expectations of schools and parents (Mullis et al., 2020), and a focus on the easiest tasks (van Zanten, 2020) or the minimum reference level (Inspectorate of Education, 2021b) may explain why talented students still do not get adequate attention and opportunities to learn.

When considering educational equity, the differences in performance between students from families with a high level of education and students from families with a low level of education are still growing, despite several decades of educational policy to reduce educational inequity. This trend, observed when comparing performance levels in subsequent data collections of PISA (Aalders et al., 2020) and register data on education (Inspectorate of Education, 2018), is worrisome and very much a topic of societal and political debate. To date, policies to increase equity in education consist of financial incentives for schools that serve more challenging student populations (IBO, 2017). A relatively recent survey of interventions implemented in primary schools (Cebeon, 2016) has shown that schools tend to spend the additional financial resources on a wide variety of interventions, including hiring teaching assistants, reducing class size, summer schools, preschool programmes and pre-teaching. More troublesome are the findings that the most popular interventions implemented by primary schools are costly and based on very limited empirical evidence on effectiveness in general or for the specific target groups.

Only very few of the policy initiatives in education in the Netherlands have been piloted (reference levels; Oomens et al., 2012) before their actual implementation or have been formally and systematically evaluated. All the initiatives presented in the previous section of this chapter (i) are multicomponent policies, (ii) provide substantial

autonomy for schools in choosing a specific form of implementation, (iii) are aimed at various (sub)populations of students and (iv) were implemented in quick succession or only very recently. Moreover, these policies were implemented in an educational landscape characterised by scaling up (Blank and van Heezik, 2015). Schools, especially in secondary education, increased in size and school boards joined together to create larger and professionally supported organisations, which potentially can improve the productivity of education (Dijkgraaf and van der Geest, 2008). In addition, the field of education has become increasingly complex, with the councils suggesting various policy changes and commercial organisations collaborating with schools more often. All in all, this implies that we cannot make strong inferences of the effectiveness of the various components of the educational policies linked to international achievement surveys, for particular subgroups or for specific subjects.

To conclude, the results of Dutch students who have participated in international large-scale achievement surveys have usually been above the international averages but have, nevertheless, pinpointed a number of weaknesses in the Dutch educational system. A pattern of declining or, at best, stable performance levels for the subjects mathematics, science and reading (literacy) over a 20-year period has been observed. Nonetheless, the declining trend has sometimes been changed into a situation in which the performance levels have remained stable; however, a systematic increase in performance has not (yet) been observed.

REFERENCES

- Aalders, P., van Langen, A. M. L., Smits, K., van den Tillaart, D. and Wolbers, M. H. J. (2020), *PISA-2018 De verdieping: Kansenongelijkheid in het voortgezet onderwijs*, KBA Nijmegen, Nijmegen.
- Blank, J. L. T. and van Heezik, A. A. S. (2015), *Productiviteit van overheidsbeleid Deel I Het Nederlandse onderwijs, 1980–2012*, Centrum voor Innovaties en Publieke Sector Efficiëntie Studies, The Hague / Delft.
- Boers-Müller, A.-M., Janssen, Y. and Steenbergen-Punterman, N. (2013), *Focus op talentontwikkeling en excelleren in het vo Een inventarisatie naar behoeften, aanbod en begeleidingsmogelijkheden voor begaafde en excellerende leerlingen binnen het voortgezet onderwijs*, Netherlands Institute for Curriculum Development, Enschede.
- Brandt, R., van Deelen, L., Franssen, A., Gerrits, P., Groenewegen, P., van Hardeveld, J. et al. (2010), *De referentieniveaus taal en rekenen in de praktijk Het vormgeven van doorlopende leerlijnen: waarom, wat en hoe?*, CPS Onderwijsontwikkeling en advies, Amersfoort.
- Cebeon (2016), *Besteding middelen terugdringen onderwijsachterstanden in het primair en voortgezet onderwijs*, Cebeon, Amsterdam.
- Daas, R. J. M. and Mennes, H. I. (2018), *Inhoudelijke vergelijking domeinbeschrijving burgerschap in het basisonderwijs en ICCS*, University of Amsterdam, Amsterdam.
- de Boer, H., Bosker, R. J. and van der Werf, M. P. C. (2010), 'Sustainability of teacher expectation bias effects on long-term student performance', *Journal of Educational Psychology*, Vol. 102, No 1, pp. 168–179, doi:10.1037/a0017289.
- Dekker, S. (2013), *Kamerbrief over toptalent in het funderend onderwijs*, Ministry of Education, Culture and Science, The Hague.
- Dijkgraaf, E. and van der Geest, S. A. (2008), *Schaalgrootte en de kwaliteit van het voortgezet onderwijs*, School of Economics, Erasmus University of Rotterdam, Rotterdam.
- Dijkma, S. A. M. (2008), *Kamerbrief staatssecretaris over de concrete uitwerking van de maatregelen om excellentie in het basisonderwijs te stimuleren*, Ministry of Education, Culture and Science, The Hague.
- Dijkstra, A. B., ten Dam, G. T. M. and Waslander, S. (2018), 'Sturing van burgerschapsvorming door de overheid? Tussen staatspedagogiek en persoonlijkheidspedagogiek', *Pedagogische Studiën*, Vol. 95, No 5/6, pp. 315–328.
- Doolaard, S. and Harms, T. (2014), *Omgaan met excellente leerlingen in de dagelijkse onderwijspraktijk*, GION onderwijs/onderzoek, University of Groningen, Groningen.
- Driessen, G. and Mulder, L. (1999), 'The enhancement of educational opportunities of disadvantaged children', in Bosker, R. J., Creemers, B. P. M. and Stringfield, S. (eds), *Enhancing Educational Excellence, Equity and Efficiency*, Kluwer, Dordrecht, Boston and London.
- Driessen, G., Mulder, L., Leest, B. and Verrijt, T. (2014), 'Zittenblijven in Nederland: een probleem?', *Tijdschrift voor Orthopedagogiek*, Vol. 53, pp. 297–311.

- Education Council (2021), *Later selecteren, beter differentiëren*, Education Council, The Hague.
- Ehren, M. C. M., Leeuw, F. L. and Scheerens, J. (2005), 'On the impact of the Dutch Educational Supervision Act: analyzing assumptions concerning the inspection of primary education', *American Journal of Evaluation*, Vol. 26, No 1, pp. 60–76, doi:10.1177/1098214004273182.
- Gubbels, J., Netten, A. and Verhoeven, L. (2017), *Vijftien jaar leesprestaties in Nederland: PIRLS-2016*, Expertisecentrum Nederlands, Nijmegen.
- Gubbels, J., van Langen, A. M. L., Maassen, N. A. M. and Meelissen, M. R. M. (2019), *Resultaten PISA-2018 in vogelvlucht*, University of Twente, Enschede, doi:10.3990/1.9789036549226.
- IBO (Interdepartementaal Beleidsonderzoek) (2017), *Onderwijsachterstandenbeleid, een duwtje in de rug? Interdepartementaal beleidsonderzoek naar het onderwijsachterstandenbeleid*, Ministry of Finance, The Hague.
- Ikeda, M. and García, E. (2014), 'Grade repetition: a comparative study of academic and non-academic consequences', *OECD Journal: Economic Studies*, Vol. 2013/1, doi:10.1787/eco_studies-2013-5k3w65mx3hnx.
- Inspectorate of Education (2010), *Risk-based Inspection as of 2009*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2012), *Achterstandbestrijding en referentieniveaus voor taal en rekenen in het vo*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2017a), *Inspection Framework Primary Education*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2017b), *Inspection Framework Secondary Education*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2018), *De staat van het onderwijs 2018*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2020a), *Onderzoekskader 2017 voor het toezicht op de voorschoolse educatie en het primair onderwijs – Geldig per 1 augustus 2020*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2020b), *De staat van het onderwijs 2020*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2021a), *De staat van het onderwijs 2021*, Inspectorate of Education, Utrecht.
- Inspectorate of Education (2021b), *Peil.Rekenen-Wiskunde – Einde (speciaal) basisonderwijs 2018–2019*, Inspectorate of Education, Utrecht.
- Isac, M. M., Maslowski, R. and van der Werf, G. (2012), 'Native students' attitudes towards equal rights for immigrants – a study in 18 European countries', *Journal of Social Science Education*, Vol. 11, No 1, pp. 7–22.

- Kelly, D. L., Centurino, V., Martin, M. O. and Mullis, I. V. S. (eds) (2020), *TIMSS 2019 Encyclopedia: Education policy and curriculum in mathematics and science*, TIMSS & PIRLS International Study Center, Chestnut Hill, MA (<https://timssandpirls.bc.edu/timss2019/encyclopedia/>).
- Kerr, D., Sturman, L., Schulz, W. and Bethan, B. (2010), *ICCS 2009 European Report – Civic knowledge, attitudes and engagement among lower secondary school students in twenty-four European countries*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Korpershoek, H., Naaijer, H. M. and Bosker, R. J. (2016), *De inrichting van de onderbouw: Onderzoek naar de motieven en de beweegredenen van vo-scholen naar soort brugklas*, University of Groningen, GION onderzoek/onderwijs, Groningen.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A. and Kelly, D. L. (1997), *Science Achievement in the Primary School Years: IES's Third International Mathematics and Science Study*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Maslowski, R. and van der Werf, G. (2013), 'The Netherlands', in Ainley, J., Schulz, W. and Friedman, T. (eds), *ICCS 2009 Encyclopedia: Approaches to civic and citizenship education around the world*, International Association for the Evaluation of Educational Achievement, Amsterdam, pp. 293–303.
- Maslowski, R., van der Werf, M. P. C., Oonk, G. H., Naayer, H. M. and Isac, M. M. (2012), *Burgerschapscompetenties van leerlingen in de onderbouw van het voortgezet onderwijs: Eindrapport van de International Civic and Citizenship Education Study (ICCS) in Nederland*, GION onderzoek/onderwijs, University of Groningen, Groningen.
- Meelissen, M. R. M. and Punter, R. A. (2016), *Twintig jaar TIMSS. Ontwikkelingen in leerlingprestaties in de exacte vakken in het basisonderwijs 1995–2015*, University of Twente, Enschede.
- Meelissen, M., Hamhuis, E. and Weijn, L. (2020a), *Leerlingprestaties in de exacte vakken in groep 6 van het basisonderwijs: Resultaten TIMSS-2019*, University of Twente, Enschede, doi:10.3990/1.9789036551090.
- Meelissen, M., Hamhuis, E. and Weijn, L. (2020b), 'The Netherlands', in Kelly, D. L., Centurino, V. A. S., Martin, M. O. and Mullis, I. V. S. (eds), *TIMSS 2019 Encyclopedia: Education policy and curriculum in mathematics and science*, TIMSS & PIRLS International Study Center, Chestnut Hill, MA (<https://timssandpirls.bc.edu/timss2019/encyclopedia/>).
- 'Ministry of Education, Culture and Science (2016), Key Figures – Education, Ministry of Education, Culture and Science, The Hague
- Mulder, L. and Meijnen, W. (2013), *Onderwijsachterstanden in de BOPO-periode, 2009–2012 – Een review (Inequality in Primary Education in the Period 2009–2011 – A review study)*, ITS, Nijmegen.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. and Smith, T. A. (1997), *Mathematics Achievement in the Primary School Years: IES's Third International Mathematics and Science Study*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M. and Foy, P. (2007), *PIRLS 2006 International Report*, International Association for the Evaluation of Educational Achievement, Amsterdam.

- Mullis, I. V. S., Martin, M. O., Foy, P. and Hooper, M. (2016), *TIMSS 2015 International Results in Mathematics*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Mullis, I. V. S., Martin, M. O., Foy, P. and Hooper, M. (2017), *PIRLS 2016 International Results in Reading*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L. and Fishbein, B. (2020), *TIMSS 2019 International Results in Mathematics and Science*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Munniksma, A., Dijkstra, A. B., van der Veen, I., Ledoux, G., van der Werfhorst, H. and ten Dam, G. (2017), *Burgerschap in het voortgezet onderwijs. Nederland in vergelijkend perspectief*, University of Amsterdam, Amsterdam.
- Nusche, D., Braun, H., Halasz, G. and Santiago, P. (2014), *OECD Reviews of Evaluation and Assessment in Education: Netherlands 2014*, OECD Publishing, Paris.
- OECD (Organisation for Economic Co-operation and Development) (2010), *Results: What makes a school successful? Resources, policies and practices*, OECD Publishing, Paris.
- OECD (2011), 'When students repeat grades or are transferred out of school: What does it mean for education systems?', *Pisa in Focus*, No 6, OECD, Paris.
- OECD (2016a), *Netherlands 2016: Foundations for the future, Reviews of National Policies for Education*, OECD Publishing, Paris, doi:10.1787/9789264257658-en.
- OECD (2016b), 'Selecting and grouping students', in *PISA 2015 Results (Volume II): Policies and practices for successful schools*, OECD Publishing, Paris, pp. 155–181.
- OECD (2019a), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (<https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>).
- OECD (2019b), *PISA 2018 Results (Volume II): Where all students can succeed*, OECD Publishing, Paris (<https://www.oecd.org/publications/pisa-2018-results-volume-ii-b5fd1b8f-en.htm>).
- Oomens, M., Kieft, M., van Haalen, J., Emmelot, Y. and Ledoux, G. (2012), *Evaluatie pilot project referentieniveaus*, Oberon, Utrecht.
- Pereira, C. and Nicolaas, M. (2019), *Effectief onderwijs in begrijpend lezen Acties voor beter leesbegrip en meer leesmotivatie*, Taalunie, The Hague.
- Plasterk, R. H. A. and van der Hoeven, M. J. A. (2009), *Nota 'De kenniseconomie in zicht'*, Ministry of Education, Culture and Science, The Hague.
- Raad voor Cultuur and Onderwijsraad (2019), *Lees! Een oproep tot een leesoffensief*, Raad voor Cultuur and Onderwijsraad, The Hague.
- Scheerens, J. (2013), *OECD review on evaluation and assessment frameworks for improving school outcomes – Update to the country background report for the Netherlands*, Ministry of Education, Culture and Science, The Hague (<https://www.oecd.org/education/school/Netherlands%20CBR%20Update.pdf>).

- Scheerens, J., Timmermans, A. C. and van der Werf, G. (2019), 'Socioeconomic inequality and student outcomes in the Netherlands', in Volante, L., Schnepf, S. V., Jerrim, J. and Klinger, D. A. (eds), *Socioeconomic Inequality and Student Outcomes – Cross-national trends, policies, and practices, Education Policy & Social Inequality*, Vol. 4, Springer, Singapore, pp. 111–132.
- Schulz, W., Ainley, J., Fraillon, J., Kerr, D. and Losito, B. (2010), *ICCS 2009 International Report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Schulz, W., Ainley, A., Fraillon, J., Losito, B. and Agrusti, G. (2016), *IEA International Civic and Citizenship Education Study 2016 assessment framework*, International Association for the Evaluation of Educational Achievement, Amsterdam.
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G. and Friedman, T. (2017), *Becoming Citizens in a Changing World – IEA International Civic and Citizenship Education Study 2016 international report*, Springer International Publishing, Cham Switzerland.
- SLO (Netherlands Institute for Curriculum Development) (2009), *Referentiekader taal en rekenen*, SLO, Enschede.
- SLO (2014), *Leerplan voor het funderend onderwijs voor de toekomst*, SLO, Enschede.
- Slob, A. (2019), *Kamerbrief over rapportage resultaten PISA-2018 in vogelvlucht*, Ministry of Education, Culture and Science, The Hague.
- Timmermans, A. C., Kuyper, H. and van der Werf, G. (2013), *Schooladviezen en onderwijsloopbanen: Voorkomen, risicofactoren en gevolgen van onder- en overadvisering*, GION onderzoek/onderwijs, University of Groningen, Groningen.
- Timmermans, A. C., de Boer, H., Amsing, H. T. A. and van der Werf, M. P. C. (2018), 'Track recommendation bias: gender, migration background and SES bias over a 20-year period in the Dutch context', *British Educational Research Journal*, Vol. 44, No 5, pp. 847–874.
- van der Molen, P., Schouwstra, S., Feskens, R. and van Onna, M. (2019), *Vaardigheidsontwikkelingen volgens PISA en examens*, Stichting Cito, Arnhem.
- van de Werfhorst, H., Elffers, L. and Karsten, S. (2015), *Onderwijsstelsels vergeleken: leren, werken en burgerschap*, Didactief Onderzoek, Utrecht.
- van Engelshoven, I. K. and Slob, A. (2020), *Meerjarenplan Alfabetisering*, Ministry of Education, Culture and Science, The Hague.
- van Leeuwen, B., Thijs, A. and Zandbergen, M. (2009), *Inclusive Education in the Netherlands*, National Institute for Curriculum Development, Enschede (<https://www.slo.nl/publish/pages/3072/inclusive-education-in-the-netherlands.pdf>).
- van Zanten, M. A. (2020), 'Opportunities to learn offered by primary school mathematics textbooks in the Netherlands', dissertation, Freudenthal Institute, Faculty of Science, Utrecht University, Utrecht, doi:10.33540/81.
- WRR (Scientific Council for Government Policy) (2013), *Naar een lerende economie. Investeren in het verdienvermogen van Nederland*, Amsterdam University Press, Amsterdam.

CHAPTER 6

Cross-national achievement surveys and educational monitoring in Sweden

Linda Rönnerberg and Christina Wikström

Abstract

Since international student achievement surveys were introduced in the 1960s, Sweden has participated in a large number of them. These surveys initially revealed unexpectedly low achievement levels in Sweden relative to other participating countries; however, there were some more positive findings and interpretations in the 1970s and 1980s. The results from the 1990s onwards have been more discouraging for Swedish educational policy and practice. In recent years, some improvements have been seen, although there is some controversy surrounding these interpretations. In Sweden, international achievement surveys play an important role in framing political discussions about educational outcomes, and they are used for education policy monitoring. Nevertheless, it is hard to establish causal links between particular reforms and such uses of international achievement data. Moreover, the Swedish case illustrates how selective national uptake of these international measures is embedded in and contingent on contextual and historical national circumstances and reform trajectories.

Introduction

Sweden has traditionally been labelled a social democratic welfare state (Esping-Andersen, 1990), characterised, like its Nordic neighbours, by generous and universal welfare in combination with an emphasis on social justice and empowerment. Education is tax funded, from compulsory education to university, and educational institutions are not allowed to charge fees. Sweden has a long history of public education aimed at both preparing students for future employment and facilitating participation in a democratic society (Antikainen, 2006), and the importance of lifelong learning has been strongly emphasised (Isaksson, 2012). Although the structure and administration of the school system has changed in many ways over time, some fundamental elements have remained. To provide equal opportunities, equivalence and ‘a school for all’ have often been highlighted as important education policy aims since the 1950s (Møller and Rönnerberg, 2021). The Education Act and national curricula are key pieces of legislation that govern education from preschool to upper secondary level. There is strong centralisation in terms of governance, regulations and monitoring, with central guidelines and standards, but the education system is also highly decentralised, with extensive local responsibility and authority, allowing the principal organisers – both municipalities and private so-called independent free school operators – freedom that is quite extensive. Schools and teachers enjoy extensive autonomy when it comes to, for instance, approaches to teaching and assessment, and interpretation of criteria.

Structure of the Swedish school system

In Sweden, the Ministry of Education and Research is responsible for the implementation of the Education Act and policy decisions. The 290 municipalities have overall responsibility for organising and delivering education, whereas the Swedish National Agency for Education (NAE) and the Schools Inspectorate (SI) are key government agencies responsible for developing, supporting and monitoring Swedish education. Since 2018, there have been 10 years of compulsory education, starting with a preschool class at age 6, followed by nine additional years of compulsory elementary education, divided into primary and secondary levels. Up to this point, education is fairly standardised, with little streaming or tracking. Three years of upper secondary education then follow (for students aged 16–19 years). Students choose between academic and vocational tracks and can attend different programmes within these tracks. Overall, 82 % of Swedish 25- to 34-year-olds have been educated to at least upper secondary level, and this share is close to the Organisation for Economic Co-operation and Development (OECD) average (OECD, 2017).

Swedish schools, both public and privately operated, are fully tax funded. They are not permitted to charge parents fees, and a form of voucher system is applied, whereby each student brings a sum of money to the school of their choosing. This goes back to the 1990s, when a range of marketisation and privatisation reforms were implemented, including school choice and privately owned so-called free schools. The overall rationale for these reforms was that, by ‘breaking the state monopoly’ in education (Lundahl, 2002), independent schools and school choice would promote pedagogic alternatives, innovation and quality, and competition would reward successful schools (whereas the less successful

schools would be weeded out, although this ambition was more implicit). Today, there is intense national debate in which both national and international policy actors have argued that these reforms have led to inequality, the reinforcement of social gaps and a deterioration in the quality of education (see, for example, Andersson et al., 2019; OECD, 2015; Swedish government, 2017). Other central policy issues concern how increased competition between schools (both public and private) has affected the validity of national instruments, such as national tests and school grades, and the extent to which they can be used to monitor quality (Gustafsson et al., 2014), as well as how international student achievement surveys have been positioned in such policy debates. In this chapter, such aspects of Swedish education policy development and debate will be further described and discussed. We begin by providing some additional information on policies relating to student achievements in the Swedish education setting by looking a bit more closely at assessment and grading, before moving on to Sweden's participation in international student achievement surveys, and how these are connected to education monitoring and policy.

Student assessment: grading and national testing

The current grading system is criterion referenced, or 'goal referenced', as it is commonly described in Sweden. It was changed from a norm-referenced system ('grading on a curve') that was used in the 1960s as a way to achieve comparable grading that would make selection decisions fairer in the transition to further education. The norm-referenced system received criticism, especially because of its limitations in evaluating performance levels and monitoring performances over time (Wikström, 2006). A criterion-referenced system was introduced in 1994, with a letter-grading system that currently ranges from F to A, with E being the pass grade.

Although the grading system and the use of national tests to assess student performance may seem similar to those in many other countries, there are some fundamental differences (Lundahl et al., 2015), including in a Nordic context (Tveit, 2018).

First, with a focus on grading, Swedish students are graded relatively late in their school career, namely in grade 6 (when students are approximately 12 years old). Over the years, policies on grading have been objects of both political discussion and reform. A policy decision was made in the early 1970s allowing schools to choose whether to grade their students as early as their first year at school or whether to wait until they were about 15 years old. About half the student population was graded, and the rest was not. Research on how this turned out showed that grading did not make much difference for the majority of students in terms of scholastic achievement. For very low- and high-achieving students, however, it seemed to make a difference. Low-ability students who were graded performed worse later on than non-graded students, when assessed in terms of grades and graduation. In addition, there was a gender difference, with girls who were graded in primary elementary school performing better later on than both boys and girls who were not graded (Klapp et al., 2016). In the 1980s, it was decided that students should not be graded until year 8 (14 years old), and since 2012/2013 students have been graded from year 6 onwards. It has recently been decided that schools can voluntarily decide to grade their students 2 years earlier, at the end of year 4 (10 years old). This has been a controversial decision, and so far very few schools have actually taken up this option.

Views on assessment also differ, which is reflected in a reluctance to summatively assess and grade students, particularly young students. There is a history of a general mistrust of grading among educationalists in Sweden, interestingly in parallel with a long tradition of psychometric research and practice (Tveit, 2018). Teachers receive no or very little training in assessment and grading during their teacher training, but they are then expected to assess and grade their own students with a lot of authority. The grades are to be based on classroom assessment, but in some subjects there are so-called national tests for grade calibration. At local level, this often works well, but it is also well known that grading practices vary between teachers and schools, as teachers have different understandings of the grading criteria and vary in their assessment expertise. This problem has been enhanced by teachers being pressured to grade leniently as a consequence of school competition, and grade inflation has been a problem ever since the criterion-referenced system was introduced (Vlachos, 2019; Wikström, 2005). This obviously also makes school grades problematic for making comparisons (Gustafsson et al., 2014), raising questions about fairness in school rankings and in the competition for attractive study positions.

Second, national testing has been discussed and implemented in different ways in the Swedish school system over time. At present, these tests, known as national tests, are administered by the NAE, which commissions content experts in higher education institutions to develop them. These tests were first implemented as low-stakes tests, but this has gradually changed (Lind Pantzare, 2018; Tveit, 2018). Previously, the national tests and their reporting were optional, implemented as tests to provide formative information and only loosely aligned with summative assessment and grading. The tests are now clearly summative and it is compulsory to use and report them. National testing has intensified in recent years, and mandatory testing of all students is now implemented in years 3, 6 and 9 in compulsory education, with an emphasis on Swedish, mathematics and English. There are also national tests in upper secondary education, in both basic and more advanced courses, as well as in core subjects in elementary school. Even if the stakes have risen, the tests cannot be compared with high-stakes external examination internationally. The Swedish national tests are intended to be important for guiding – but not determining – teachers' practices. There is no external monitoring or scoring. The teachers score the tests themselves, and there is some room for subjective interpretations, although the scoring guidelines have become more precise over time.

As instruments for educational evaluation and monitoring, the Swedish national tests have several limitations, although evaluation and monitoring are expected outputs from a policy perspective (Tveit, 2018). First of all, the tests are not strictly comparable between test administrations, which makes them difficult to use for evaluation on a larger scale, between schools or over time. Since they are administered by the schools and scored by the teachers, the pass scores are set in advance, which makes it hard to adjust for variations in difficulty between test administrations (Lind Pantzare, 2018). This means that there are both reliability and validity issues that are difficult to control for. Inter-rater reliability can be problematic, and teachers do not always interpret rubrics in a similar way, leading to more systematic differences. The possibility of using external markers has been discussed for some time, but no practical solution has been presented. Currently, there is political pressure to find alternatives, but there is still no

formal legislation pertaining to this. It has also been suggested that the national tests could be redesigned to provide a valuable complement to the international surveys, if the development and scoring of the tests were adjusted to work for such purposes (Gustafsson et al., 2014). Such changes are also still to be implemented, and this means, as this chapter will show, that the international achievement surveys have become quite central to longitudinal educational performance monitoring in the Swedish education policy setting, along with all the strengths and limitations attached to such tests and how the surveys are designed.

International achievement surveys

Sweden has, to date, been a frequent participant in various international achievement surveys; it has participated in International Association for the Evaluation of Educational Achievement (IEA) studies, predominantly the Trends in International Mathematics and Science Study (TIMSS) and previous versions of this survey, but also the Progress in International Reading Literacy Study (PIRLS), as well as OECD studies such as the International Adult Literacy Survey and the Programme for International Student Assessment (PISA). During the first 40 years of participation, up to 2004, Sweden was involved in more than 40 international comparative studies (Skolverket, 2004), and now, some 20 years later, many more can be added to this list (Gustafsson and Blömeke, 2018).

The 1960s and onwards

Sweden participated in the first IEA study – the First International Mathematics Study (FIMS) – in 1964. This was due, to a great extent, to the involvement of Swedish scholars in the planning and implementation of the first survey, led by Professor Torsten Husén, the first secretary-general of the IEA (Husén, 1967). FIMS was administered at a time when there was general concern about knowledge levels in the country and the ability to compete internationally. Research was focused on how to take advantage of the hidden ‘ability reserve’, that is, students with high cognitive ability, in the expansion of the educational system (Husén and Härnquist, 2000). The outcome of FIMS was, however, a disappointment, since Swedish students ranked low in comparison with the other 11 countries, especially the younger students, who were at the bottom. These results led to an intense debate about the quality of Swedish education, in which the centralised and unified school system that had been implemented in full during the 1960s was criticised for holding talented students back (Skolverket, 2004). A few years later, in 1970, Sweden participated in the First International Science Study. This time, the results were more positive, especially when it came to reading comprehension among the younger students. They did, however, reveal gender differences, with male students outperforming female students, especially in science. These findings caused some debate but did not lead to any major changes in education or policy (Skolverket, 2004). It should be noted that, from the late 1960s onwards, support for using educational measurement as a tool in the educational system declined and was replaced by a mistrust of the measurement of students and schools. Evaluation was, however, more popular, and extensive evaluation programmes focusing on education practice were undertaken. The focus of concern at

this time was social aspects of school policy rather than performance evaluation. The reason for this was, to a large extent, due to the change in the landscape in relation to Swedish schools, as a result of urbanisation, immigration and a general restructuring of society. Many schools were large and heterogeneous, which brought a number of challenges that had not been experienced before. The quality of education in the comprehensive school system was extensively discussed but approached differently in the educational and political debates (Isaksson, 2012).

Views on the relevance of educational measurement and evaluation gradually changed, as did society in general. In 1976, Sweden elected a liberal government after 40 years of social democratic rule, albeit for only 6 years. Nevertheless, the political shifts were more frequent from then on, which made more long-term work on educational reforms increasingly complicated, with reforms often built on compromise and sometimes based on more than one political ideology. In 1980, a new national curriculum was introduced. Among other things, the option to grade the younger pupils was revoked, and all pupils were graded from the end of year 8. This coincided with Sweden's participation in the Second International Mathematics Study. Yet again, the results were disappointing, with poor performances in arithmetic, algebra and geometry. The outcome was referred to as the 'mathematics crisis' and led to calls for a number of improvements, focusing on teachers, including a more selective admissions policy for teacher education, changes to teacher training and professional development for practising teachers (Ministry of Education, 1986, p. 5). According to Emanuelsson (2001), ambitions were high, but only a few of these requests or recommendations were actually implemented. In subsequent years, the debate about poor performance and gender gaps calmed down somewhat, especially after Sweden had participated in the Second International Science Study in 1984, this time with improved results. In particular, the poor- to medium-performing students did better, and the gap between male pupils and female pupils had narrowed. Overall, compared with the first surveys, a small but consistent increase in the level of achievement could be observed during this decade (Gustafsson and Blömeke, 2018). The results were interpreted as supporting a school system with little differentiation, which had been the aim of the reforms implemented some decades earlier and was now fully in place (Skolverket, 2004).

The 1990s and early 2000s

The beginning of the 1990s was also quite uncontroversial. Sweden participated in the first PISA study, administered in 2000, and, apart from some disappointing performances in mathematics, the results did not cause any immediate alarm. In fact, in PIRLS, in which Sweden also participated, the 10- to 11-year-olds performed really well, being ranked top in some areas. One concern, however, was an additional group difference, this time the gap between students with a Swedish background and students with an immigrant background; this was an inequality issue that the NAE promised to focus on in the future.

In spite of the positive changes in results, this was a time of political and ideological change, with neoliberal trends in society. This, in combination with changes in political governance, led to profound changes for Swedish schools during this decade. A number of reforms were implemented in the early 1990s, and several of these reflected a belief

in the effectiveness of decentralisation and marketisation in society. As mentioned earlier, the responsibility for schools was moved from the state to the municipalities. Furthermore, tax-funded independent free schools and a voucher system were established to encourage school choice and school competition. A new so-called goal-oriented, that is, criterion-referenced or standards-based, curriculum was introduced, along with the new criterion-referenced grading scale described earlier. A standards-based system was expected to ensure that all students would meet knowledge requirements and that the schools could be evaluated based on performance – namely the degree to which the schools succeeded in delivering equal and high-quality education to their students.

In the midst of this period of reforms, Sweden participated in TIMSS 1995, and the results were better than before. This was interpreted as a positive effect of the improvements that had been made after the previous mathematics crisis (Emanuelsson, 2001). From then on, Swedish participation in international studies increased, due to a great extent to PISA. The PISA 2000 results were, however, somewhat disappointing, especially in mathematics. The government gathered a team of experts to determine the reason for this and to make recommendations on what to change and focus on. The experts identified several problems and presented a number of recommendations, including on how content and aims could be improved, on how clearer criteria for teaching and assessment for all levels could be formulated and on relevant professional development for teachers (Swedish government, 2004).

In the same year in which the expert report was published, Sweden participated in PISA and TIMSS, and also in a national evaluation of Swedish schools, the NU-03. The PISA results had not changed much, with fairly good results in reading but poorer results in problem-solving and science. In TIMSS, however, Sweden ended up scoring below the OECD average, with the largest decline among all participating countries since the first TIMSS (Skolverket, 2004). The PISA 2006 results did not really add to these alarming findings from TIMSS and were in line with the OECD average. However, there had been some changes in the PISA survey that made comparisons with previous results difficult (Skolverket, 2007). Finland stood out with an excellent performance, which caused discussion since it is a neighbouring country with many similarities to Sweden. It was argued that this was to do with Finland having a more equal and traditional school system, since Finland had been reluctant to carry out reforms such as those implemented in Sweden and had maintained a traditional school system in which the status of teachers was still high (Gustafsson and Blömeke, 2018).

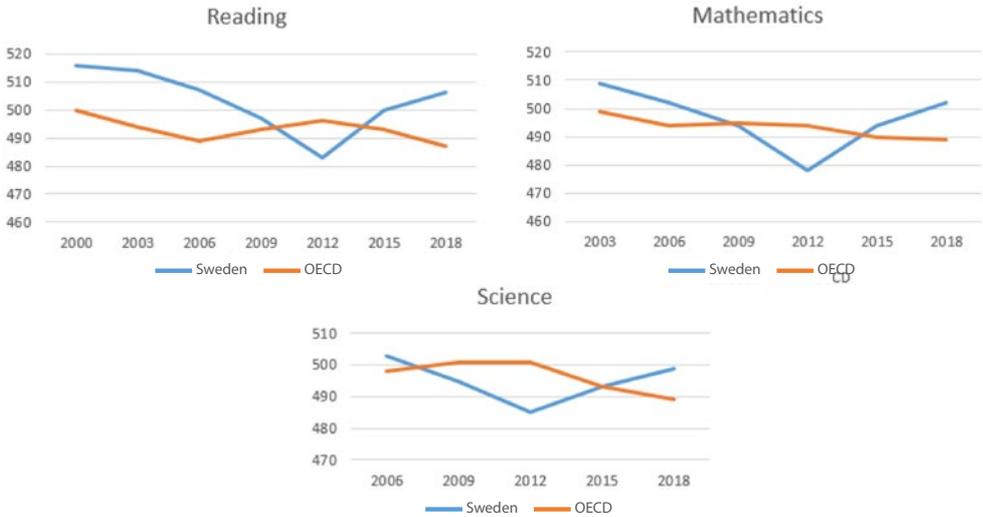
The 2010s and the ‘PISA shock’

In 2011, a new Education Act was introduced in Sweden, including stricter regulation, especially for independent schools, and a higher degree of school monitoring. Other important changes were revised national curricula for elementary and upper secondary level education, and a grading scale with more grade levels. The TIMSS 2011 and PIRLS 2011 surveys confirmed that improvements were called for, as they indicated a continuing decline in results. Still, it was the PISA 2012 release in December 2013 that led to the real so-called PISA shock. The results from this study showed an unprecedented decline in knowledge and skills, with scores in all subject areas falling below the OECD

average. They also revealed increasing gaps between schools and groups of students; in particular, the low achievers' results had declined the most (Davidsson et al., 2013). The results were considered a national failure and led to a lot of discussion and political debate.

Following the shock of the steep Swedish decline in PISA 2012, the Minister for Education and Research commissioned the OECD to investigate how Sweden could improve. The OECD report (2015) largely confirmed the gravity of the problems with Swedish education from the perspective of both quality and equality. The OECD stressed the importance of improving the quality of the teaching and school leadership profession, giving more support to disadvantaged students and implementing a national school improvement strategy (OECD, 2015). A commission was appointed and tasked with delivering a set of concrete proposals outlining how to improve student and school performances, enhance the quality of education and promote equal education for all, together with an action plan that was summarised in a comprehensive commission report (Swedish government, 2017). The results from the following PISA surveys in 2015 and 2018 were, overall, perceived as a relief to the government and to the country, since they showed improved results and smaller gender gaps, even though the problems of inequality and socioeconomic gaps remained. The performance increases were not great, but they were claimed to be statistically significant and in line with the 2006 results, that is, around the OECD average (Skolverket, 2019). The improved results were interpreted as signs of the beginning of a positive trend and an improvement in school quality, but, as we will see in our final section, PISA 2018 exclusions were also debated and questioned.

Figure 6.1 depicts Sweden's PISA performance in relation to the OECD average and serves to illustrate the varying results over time, with the 'shock' in 2012 particularly striking; the abovementioned improved performance in the two most recent PISA surveys is also notable.

Figure 6.1: Sweden's PISA performance and OECD averages

Sources: OECD (2021a,b,c).

Education policy monitoring

Overall, there has been a general expansion of different activities relating to policy monitoring and assessment in education in recent decades. This development is related to the overall education policy direction of the past three decades described above, in which far-reaching decentralisation and marketisation have been accompanied by intensified policy measures in relation to evaluation and accountability. As a result, national education evaluation and monitoring has expanded in both scope and intensity. Policies have been implemented covering, for instance, intensified national school inspection, more and earlier national testing of pupils in different school subjects, earlier grading and formalised assessment of student performance from the start of compulsory education (Jarl and Rönnberg, 2019).

Actors and activities

Initially, it should be noted that, as in several other countries, there is no formally designed framework for education policy monitoring in Sweden (Benedal, 2019; Nusche et al., 2011; Swedish government, 2014). Even so, there are a number of activities undertaken by actors at different levels that encompass diverse efforts and measures relating to system monitoring, evaluation and accountability. Ultimately, the Ministry of Education and Research is responsible for monitoring the education system, but this work is largely carried out by different national agencies, in particular the NAE.

Table 6.1 shows some of the main evaluation and monitoring activities and actors.

Table 6.1: Activities and actors in Swedish education policy monitoring processes

National level	Principal organiser (municipality or independent organiser)	School level
National school inspections, all schools and all principal organisers, results made public (SI)	Annual municipal financial monitoring of the local political council's work as well as monitoring of the work in schools (municipal-level accountants etc.)	Evaluation and assessment of students, including knowledge assessments, grading (years 6–9), developing tests, individual study plans for each student who is too young for grading (teachers)
Annual collection of statistical key indicators on, for instance, students, costs, personnel and student achievement (NAE)	Continuous and systematic evaluation, assessment and follow-up of different aspects of education (municipal central officers in education, independent organiser central management staff)	Evaluation, assessment and follow-up of schoolwork and its organisation, for instance in subjects, teaching or development projects (head teachers and teachers)
National evaluations (by the NAE, and other agencies, and also by public inquiry commissions, university researchers, etc.)	Systematically administering and documenting local quality improvement work (every principal organiser, public as well as private)	Systematically administering and documenting quality work and development (every school and head teacher, as well as teachers and students)
National and mandatory standardised testing, in a selection of subjects in years 3, 6 and 9, and further optional tests (NAE)		
Participation in cross-national student achievement surveys (NAE, university researchers)		

Source: Adapted from Jarl and Rönnerberg (2019).

Table 6.1 makes it clear that there are a number of central activities and processes at national level, but these are often linked to or originate from the level of the principal organiser and/or the school, thus highlighting the important roles of teachers and head teachers. Furthermore, this also reflects the general high level of decentralisation, local autonomy and discretion of municipalities, school principal organisers and school professionals. Taken together, the activities at different levels feed into and are aggregated for system-level policy monitoring purposes. The main part of such activities takes place at national level; below, we will focus on the first column of Table 6.1.

National education policy monitoring features activities carried out by the SI. Inspections have a long history in Sweden, dating back to the early years of compulsory education in the 1860s. After having been abolished for almost a decade after the extensive decentralisation reforms in the 1990s, school inspections were reintroduced in 2003. Today, regular school inspections are undertaken in each school and for each principal organiser – private and public – and the results are published in a publicly available inspection report (Rönnerberg, 2014). Fines can be issued if shortcomings are detected. The SI, in addition, also conducts surveys of parents, teachers and students that form part of

the data used when undertaking the regular inspections, and it presents annual reports to the government on aggregated results and findings from its inspection activities. The SI also issues permits for independent schools and investigates complaints from parents and students against individual schools (SI, 2021).

The NAE is a key actor in monitoring education policy. As is the case with other national agencies, it receives annual guidelines from the government in the form of instructions and appropriation directions, but, at the same time, the agency exercises its own authority within these frameworks. The NAE is responsible for a number of important activities, such as (i) the collection of annual education statistics, (ii) evaluations, (iii) national testing and (iv) administering Sweden's participation in international achievement surveys. The annual education statistics include different measures of students, school staff, education costs and student results – primarily in the form of grades. These indicators are central to any national follow-up and general monitoring. National evaluations, such as the NU-03 mentioned previously, have, however, gradually diminished in importance in recent decades. A commission report (Swedish government, 2014) concluded that both national testing and results from cross-national achievement surveys have become more important data sources for policy monitoring purposes, disregarding the fact that the national tests have limitations for such purposes and that the international surveys have their limits too; for instance, they are not connected to the national curricula and syllabi.

With regard to national testing, different types of national tests have a long history in Sweden and have served different education policy purposes over time. The political motives for intensifying national testing in recent years have included a desire to ensure (more) equal grading practices and early identification of pupils in need of special support (Lind Pantzare, 2018; Skedsmo, et al., 2021). As already mentioned, the assessments of the tests have been widely debated, and the SI has been given the responsibility by the government of rescoring samples of teachers' grading on national tests. These reassessments have repeatedly found considerable differences, with (class) teachers generally being more generous in their assessments than the SI (see Novak and Carlbaum, 2017).

Last but not least, Sweden's participation in cross-national achievement surveys is also central to supporting national education policy monitoring. A key national document is the Swedish annual budget bill, and there are some central statistical indicators that are used to monitor education specifically. These include aggregated national student performance data, for example in the form of grades, but also international data in the form of, for instance, PISA results (on performance and segregation); other international achievement surveys (for instance TIMSS) are also regularly reported in the budget bills. Their status as key indicators in these documents adds particular importance to such measures, and here international achievement surveys are indeed key (National Audit Office, 2021). The government commission report from 2014 explained the vital role of this cross-national student data in evaluating school reforms generally: 'today only the international student achievement surveys have the design and quality to monitor student performance in Swedish schools over time' (Swedish government, 2014, p. 18). In addition, Sweden regularly reports on the strategic framework for European

cooperation in education and training indicators in national policy documents and is a part of the European semester, in which, among other things, educational outcomes are carefully monitored, for instance through international student performance data.

We will return to and expand on these issues below, but next we would like to highlight some issues and debates relating to education policy monitoring in Sweden, as this will provide relevant context in which the uses and misuses of international student achievement data can be understood.

Issues and debates in education policy monitoring

Education system monitoring and, indeed, improving education system performance have increasingly been in the public and political spotlight, and have received considerable attention in recent years. Several of the recent and ongoing Swedish reforms included in the Eurydice database relate to issues of monitoring and educational performance and results in one way or another. These include measures to protect students' safety and health and ensure that they are able to find peace to study; measures to increase equality; curriculum and syllabus revisions; and additional measures to improve student performance and to increase the proportion of students eligible for upper secondary education (Eurydice, 2021). Issues of system and student performance, and how performance is to be assessed, by whom and in what ways are thus undoubtedly issues on the education policy agenda.

In the past decade, the OECD has repeatedly criticised Sweden on issues relating to student performance data and education policy monitoring (Nusche et al., 2011; OECD, 2015, 2017; see also Swedish government, 2014). As the Swedish education system overall relies quite heavily on teacher-based assessments, in combination with a system of intense school competition to attract students, issues of reliability of student achievement data have been raised in different ways. In brief, a central line of reasoning is that grades are not sufficiently robust to aid in national policy monitoring, and national tests, although standardised, still rely on local marking and cover only some selected subjects. As a result, the overall reliability of national student results data and their use in policy monitoring have been and continue to be questioned by the OECD. In the wake of Sweden's poor PISA performance in 2013, the OECD's commissioned report with policy advice for the Swedish government (OECD, 2015) emphasised, among other things, the need to reform accountability arrangements and to address the issue of 'unreliable national student achievement data' (OECD, 2015, p. 143).

These issues have indeed also been subjects of debate in the Swedish national setting, and repeated uncovering of grades higher than the national test results supports claims of grade inflation (Vlachos, 2019; Wikström, 2005). Although this has been a source of both scholarly and political concern, no significant policy measures have been implemented to resolve the issue. The proposals in the Swedish government's official report (2014) referred back to the advice given by the OECD in its report from 2011 (Nusche et al., 2011), and, even though the government's report aimed to set up a national system for education policy monitoring, we must note that this has not yet been implemented either.

To summarise, this section has illustrated that there are several activities and actors involved in education policy monitoring, but there is no systematic and coherent education policy monitoring system. National student assessments and data, in the form of grades and standardised testing, have been both reformed and debated with regard to their reliability, and in this particular context international comparative student assessments become potentially important for monitoring system performance, in particular from a longitudinal perspective. Nevertheless, and as we observe below, this does not mean that cross-national achievement surveys are actually used in policy development and policy evaluation.

Analysis of policies, practices and political reforms

The political and public reactions to surveys such as PISA have varied and also been divided to some extent. Claims have been made that the international surveys have become political tools of legitimation, used to cherry-pick results that support ideological beliefs and policy decisions. Thus, the relationship between international student achievement survey results and political reform is neither straightforward nor easy to disentangle. On the one hand, it is a challenging task to establish any relationships and therefore very hard to identify causal links, and thus to identify reforms that were primarily and/or originally driven or discouraged by results and monitoring of international performance data. On the other hand, it is also clear that international student survey results have indeed become increasingly embedded in the general political and public conversation about education and reform. In this way, these data are central sources for both legitimation and argumentation in policy debates. It is also important to reiterate the centrality of these instruments in monitoring education achievements over time. The international education achievement surveys do not employ a longitudinal design and hence survey different students in every new round, which limits the value of these cross-sectional surveys for measuring change over time. Still, this is one of the main functions of these surveys and also a common (mis) interpretation of these tests in Sweden.

Below, we discuss how the reaction to international performance surveys is influenced by diverse political and contextual conditions. In addition, we provide some illustrations of how the uptake or function of performance data may vary due to these different circumstances. Let us first conclude that there is no doubt that international student achievement data are an integrated part of contemporary Swedish education policy deliberations. The results from these surveys are reported in the annual budget bills and in other central official policy and agency documents. As an illustration of their increasing centrality, the parliamentary document search engine yielded almost 600 hits for documents mentioning PISA from 2015 to 2020, followed by 125 hits for TIMSS and 48 hits for PIRLS (National Audit Office, 2021). This also signals the weight assigned to PISA compared with other international student assessments. Contrasting the past 6 years with the first 6 years of PISA (2000–2006), the corresponding hits in the public policy documents database were 159 for PISA, 39 for TIMSS and 16 for

PIRLS (Riksdagen, 2021). International student achievement surveys thus appear to be receiving increasing attention in policy documents and consequently the policy process. But in what ways are they receiving increasing attention and what are the consequences of this?

Selective legitimization of political reforms post 2000

Initially, the first PISA data releases were not followed by extensive media and political attention, but things changed gradually from the mid 2000s onwards. One change was the new government in office in 2006, which had an extensive education reform agenda; another was the start of the decline in PISA performance (Ringarp and Waldow, 2016). From this point, references to PISA and ‘reference societies’ based on PISA performance started to gain political ground and importance (see Breakspear, 2012). When analysing Sweden’s relationship with the OECD, Grek asserts that ‘this declining performance, given the history of Sweden as a model European education system throughout the twentieth century, in addition to the success of close neighbours (Finland, for example) became instrumental in allowing the OECD to enter the country in order to socialise and “educate” all relevant actors about the critical need for change’ (Grek, 2020, p. 188).

Indeed, there are numerous reports on how the ‘international’ situation in a wider sense but also results and measurements from international student surveys are legitimising reform in grading (Lundahl et al., 2015), ‘standards-based’ curricular revisions (Sundberg and Wahlström, 2016) and other initiatives to, for instance, promote mathematics (by the National Association for Local Authorities and Regions; see Arnesson, 2016). The results from international surveys are used as political legitimization for reform and targeted initiatives in different ways. There is selective use of political referencing of international student surveys in the context of education reform (Ringarp and Waldow, 2016).

Thus, the ‘international student performance argument’ has been decisive in some policy reforms but not others. There are also instances in which the Swedish uptake of the results from international student surveys has not been visible in the form of actual political reform decisions to revise policy. In fact, many of the proposals from the initiatives raised by the OECD (2015) and national commissions post PISA 2012 (Swedish government, 2016, 2017) still have not been adopted politically or implemented. One such issue concerns school choice and the segregation effects in its wake. In its report, the OECD (2015, p. 101) urged Sweden to ‘Revise school choice arrangements to ensure quality with equity’, arguing that Sweden could ‘benefit from managing school choice to prevent segregation and increased inequities. Providing full parental school choice can result in segregating students by ability and/or socioeconomic background and generate greater inequities.’ The issue was then further addressed in the national school commission’s report, and yet another commission was tasked with drafting a reform proposal (Swedish government, 2020). However, all attempts to reform school choice (and/or the liberal conditions for the private free schools) have been met with great resistance from business actors active as school owners and proponents of marketisation, and this has meant that establishing political majority coalitions to implement any restrictions has proved very difficult (Rönnerberg et al., 2019). This also goes for the contested issue of limiting profit-

making in publicly funded but privately delivered education, a policy design feature unique to Sweden. As with profit-making, school choice and admission regulations are still pending, and, at the moment, there is no political majority in Parliament allowing the possibility of revising school choice policy.

The above examples illustrate the diverse and nationally rooted nature of the uptake (or not) of international performance data in different national contexts, and there is no straightforward causal relationship between international survey data, education policy monitoring and national education reform. Nevertheless, these indicators, benchmarks and follow-ups are brought up and used as arguments for change in different ways. When national political agendas join up with international measures, strong political and data alliances are formed; this means that the political institutional set-ups and rules of the game are important to consider when analysing possible policy impacts in different national arenas. Sweden's multiparty electoral system based on proportional representation results, not infrequently, in minority coalition governments; this is critical when formulating policy and getting legislation through Parliament, and has an impact on both the form and content of education policy processes (Baggesen Klitgaard, 2007; Jarl and Rönnerberg, 2019).

The parliamentary arena: who mentions the Programme for International Student Assessment and when?

Analysing references to PISA in parliamentary minutes from 2000 to 2017, Lundahl and Serder (2020) reported that they appeared mostly when arguing in favour of a particular change or reform, and more seldom as an argument against reform proposals; they also noted that PISA was used in arguments by members of Parliament from all political parties. Lundahl and Serder coded more than 350 mentions of PISA in connection with more than 30 proposals or reforms, and the most common reference to PISA occurred in relation to promoting equivalence in Swedish education. Equity and equivalence have a particular political meaning and history in Swedish education policy (Rönnerberg, 2012) and have traditionally been important for all political parties. Since PISA reports on segregation and associated issues, it is not surprising that references to it are made, given the general centrality of the issue in the Swedish education policy context.

In addition to equivalence arguments, PISA was also frequently mentioned with reference to promoting teachers' professional status and development, and it should be noted that most of these references were made after 2012. Other common mentions of PISA in the minutes include (i) calls to invest more in education, (ii) support for early grading, (iii) references to improving order and discipline in classrooms, (iv) mention of students in need of support and (v) support for promoting digital skills. There are a few party political differences: PISA and immigration are mainly referenced by the Swedish Democrats, and order and discipline issues and early grading are primarily referenced by parties of the centre right (Lundahl and Serder, 2020). Overall, the most common political use of references to PISA in these parliamentary debates was to promote reform by referring to declining student achievement. These references are situated and embedded in the political game, and the opposing political blocs (left wing and right wing) use PISA as a reference to discredit the opposition.

The study by Lundahl and Serder (2020) highlights how references to PISA in the national parliamentary arena are linked to both historically rooted and party politically salient issues, and, in that way, serve to illustrate empirically the contextual character of uptake and use of national achievement surveys in national policy formation. In addition, these results also seem to support the overall conclusion of Santos and Centeno (2021), with external references to PISA in the Swedish national policy debate overall seemingly used ‘at moments of political turbulence with the goal of convincing others of the benefits (or uselessness) of a policy idea or proposal’ (Santos and Centeno, 2021, p. 13).

The debate on exclusion – is Programme for International Student Assessment 2018 to be trusted?

Lundahl and Serder (2020) reported that it was less common for references to concern whether and how PISA was actually a valid measurement of the state of Swedish education. Hence, overall, it is assumed that PISA is ‘correct’. However, Lundahl and Serder’s data ended in 2017, and since PISA 2018 there has been intensified debate in Sweden about the rationale and premise for the increase in the number of students excluded from taking the PISA test. Given that issues of validity and the actual implementation of these surveys seldom emerge in political PISA debates, we now devote some attention to this particular issue. We will show the high political stakes attached to PISA for contemporary political legitimacy and reputation, and also how the questioning of compliance with OECD regulations created a space for party political arguments and the need for ‘impartial’ monitoring.

The implementation of PISA 2018 was subject to criticism when it came to the sampling of participating students, and the NAE was questioned for excluding too many students with an immigrant background. The NAE and the Ministry of Education and Research explained that this was an unusual situation, since Sweden had received a very large number of immigrants in 2015, who, according to OECD regulations, can be excluded if they have recently immigrated. The OECD approved the high Swedish exclusion rate (11 % compared with the OECD-stated maximum of 5 %). Nevertheless, one national newspaper looked into the high exclusion rate, for instance by interviewing head teachers in schools who had been selected for PISA 2018, and obtained testimonies describing excessive levels and exclusion on false grounds. As a response, the NAE commissioned the OECD to undertake a second review of the previously approved exclusion rate and its potential consequences for the interpretation of the PISA data. This move received additional criticism, and a central argument, made by the political opposition and the Parliamentary Standing Committee on Education, among others, was that both the NAE and the OECD had interests at stake in this second review, and thus it was inappropriate for them to be responsible for a second scrutiny of the exclusion procedure and its potential effects.

As an outcome of this criticism, the National Audit Office was appointed to review both the second OECD review process and its outcome. This report was made public in 2021 and concluded that incorrect decisions had been made in this respect – the majority of the students who immigrated to Sweden during the European migration crisis in 2015 had been in Sweden too long to be excluded (National Audit Office, 2021). The NAE,

the Minister for Education and Research and the OECD were all criticised for not having dealt with these exclusion processes or the second review appropriately. Representatives of the parliamentary opposition were quick to demand that the minister and the NAE director-general resign, and the implementation of PISA 2018 was thus turned into a tool for discrediting political parties. This situation is still to be resolved, but at the time of writing the Swedish Minister for Education and Research had been summoned to parliament to explain the actions and inactions of the government and the NAE.

Summary

This chapter has described Swedish education policy development and monitoring and how these activities intersect with international student performance data in different ways. In brief, there is increasing emphasis and centrality ascribed to international student survey results in order to (selectively) legitimise and argue for education reform in the Swedish education policy context. There is a nationally distinct reaction to international performance surveys, and to get a full picture selective national uptake needs to be situated in countries' respective political-institutional and historical contexts. In particular, the national reception is founded on and filtered through previous policies, reform trajectories and assessment histories. Highlighting such contextual circumstances shines an additional and useful analytical light on the varying and selective political functions of international student performance data.

As a final note, Swedish education policy has maintained the drive for equivalence and social justice and combined it with more market-oriented elements, and the inclusive and comprehensive 'a school for all' idea is still a viable and widely endorsed policy construct. Although there are clear indications that students' socioeconomic backgrounds matter more for student performance today than previously – a deeply concerning issue – there is still an explicit political will to provide good education to all. Despite the fact that there are indeed many issues that are in urgent need of reform in future education policy, it is important to recall that Sweden, overall, still has a comparatively equitable school system. It remains to be seen if and how international student survey results will make their mark, or, alternatively, will be made invisible or ignored, in future education policy formulation and legitimation.

REFERENCES

- Andersson, E. K., Hennerdal, P. and Malmberg, B. (2019), 'The re-emergence of educational inequality during a period of reforms: a study of Swedish school leavers 1991–2012', *Environment and Planning B: Urban Analytics and City Science*, Vol. 8, No 4, pp. 685–705.
- Antikainen, A. (2006), 'In search of the Nordic model in education', *Scandinavian Journal of Educational Research*, Vol. 50, No 3, pp. 229–243.
- Arnesson, D. (2016), 'PISA i skolan: hur lärare, rektorer och skolchefer förhåller sig till internationella kunskapsmätningar', Department of Applied Educational Science, Umeå University, Umeå, Sweden.
- Baggesen Klitgaard, M. (2007), 'Why are they doing it? Social democracy and market-oriented welfare state reforms', *West European Politics*, Vol. 30, No 1, pp. 172–194.
- Benerdal, M. (2019), 'Nationell utvärderingspolicy: utformning och förändring på grundskoleområdet 1988–2014', dissertation, Department of Political Science, Umeå University, Umeå, Sweden.
- 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*, No 71, OECD Publishing, Paris.
- Davidsson, E., Karlsson, K-G. and Oskarsson, M. (2013), 'Trender och likvärdighet: Svenska elevers resultat på PISA naturvetenskap i en internationell jämförelse', *Utbildning & Demokrati: Tidskrift för didaktik och utbildningspolitik*, Vol. 22, No 3, pp. 37–52.
- Emanuelsson, G. (2001), *Svårt att lära–lätt att undervisa? Om kompetensutvecklingsinsatser för lärare i matematik 1965–2000*, National Centre for Mathematics Education, Gothenburg (http://ncm.gu.se/media/ncm/kup/Svart_att_lara.pdf).
- Esping-Andersen, G. (1990), *The Three Worlds of Welfare Capitalism*, Princeton University Press, Princeton.
- Eurydice (2021), 'National reforms in school education – Sweden' (https://eacea.ec.europa.eu/national-policies/eurydice/content/national-reforms-school-education-71_en).
- Grek, S. (2020), 'Facing “a tipping point” ? The role of the OECD as a boundary organisation in governing education in Sweden', *Education Inquiry*, Vol. 11, No 3, pp. 175–195.
- Gustafsson, J-E. and Blömeke, S. (2018), 'Development of school achievement in the Nordic countries during half a century', *Scandinavian Journal of Educational Research*, Vol. 62, No 3, pp. 386–406.
- Gustafsson, J.-E., Cliffordson, C. and Erickson, G. (2014), *Likvärdig kunskapsbedömning i och av den svenska skolan – problem och möjligheter.*, SNS Förlag, Stockholm.
- Husén, T. (1967), *International Study of Achievement in Mathematics*, John Wiley, New York.
- Husén, T. and Härnqvist, K. (2000), *Begåvningsreserven: en återblick på ett halvsekels forskning och debatt*, Föreningen för svensk undervisningshistoria, UppsalaSweden.

- Isaksson, C. (ed.) (2012), *Grundskolan 50 år: Från folkskola till folkets skola*, Ekerlids Förlag, Stockholm.
- Jarl, M. and Rönnerberg, L. (2019), *Skolpolitik: Från riksdagshus till klassrum*, Liber, Stockholm.
- Klapp, A., Cliffordson, C. and Gustafsson, J.-E. (2016), 'The effect of being graded on later achievement: evidence from 13-year olds in Swedish compulsory school', *Educational Psychology*, Vol. 36, No 10, pp. 1771–1789.
- Lind Pantzare, A. (2018), 'Dimensions of validity: studies of the Swedish national tests in mathematics', dissertation, Department of Applied Educational Science, Umeå University, Umeå, Sweden.
- Lundahl, L. (2002), 'Sweden: decentralization, deregulation, quasi-markets – and then what?', *Journal of Education Policy*, Vol. 17, No 6, pp. 687–697.
- Lundahl, C. and Serder, M. (2020), 'Is PISA more important to school reforms than educational research? The selective use of authoritative references in media and in parliamentary debates', *Nordic Journal of Studies in Educational Policy*, Vol. 6, No 3, pp. 193–206.
- Lundahl, C., Hultén, M., Klapp, A. and Mickwitz, L. (2015), *Betygens geografi – Forskning om betyg och summativa bedömningar i Sverige och internationellt*, Vetenskapsrådet, Stockholm.
- Ministry of Education (1986), *Matematik i skolan – Översyn av undervisningen i matematik inom skolväsendet (DsU 1986:5)*, Liber, Stockholm.
- Møller, J. and Rönnerberg, L. (2021), 'Critical perspectives in and approaches to educational leadership in Norway and Sweden', in Courtney, S., Gunter, H., Niesche, R. and Trujillo, T. (eds), *Understanding Educational Leadership: Critical perspectives and approaches*, Bloomsbury, London.
- National Audit Office (2021), *Pisa-undersökningen 2018 – arbetet med att säkerställa ett tillförlitligt elevdeltagande*, Riksdagen, Stockholm.
- Novak, J. and Carlbaum, S. (2017), 'Juridification of examination systems: extending state level authority over teacher assessments through regrading of national tests', *Journal of Education Policy*, Vol. 32, No 5, pp. 673–693.
- Nusche, D., Halász, G., Looney, J., Santiago, P. and Shewbridge, C. (2011), *OECD Reviews of Evaluation and Assessment in Education: Sweden 2011*, OECD Publishing, Paris.
- OECD (Organisation for Economic Co-operation and Development) (2015), *Improving Schools in Sweden*, OECD Publishing, Paris.
- OECD (2017), *Education Policy Outlook: Sweden*, OECD, Paris (<https://www.oecd.org/education/Education-Policy-Outlook-Country-Profile-Sweden.pdf>).
- OECD (2021a), 'Reading performance (PISA) (indicator)', OECD, Paris, doi:10.1787/79913c69-en.
- OECD (2021b), 'Mathematics performance (PISA) (indicator)', OECD, Paris, doi:10.1787/04711c74-en.
- OECD (2021c), 'Science performance (PISA) (indicator)', OECD, Paris, doi:10.1787/91952204-en.
- Riksdagen (2021), 'Dokument & lagar' (<https://www.riksdagen.se/sv/dokument-lagar/>).

- Ringarp, J. and Waldow, F. (2016), 'From "silent borrowing" to the international argument – legitimating Swedish educational policy from 1945 to the present day', *Nordic Journal of Studies in Educational Policy*, No 1, pp. 1–8.
- Rönnerberg, L. (2012), 'Reinstating national school inspections in Sweden: the return of the state'. *Nordic Studies in Education*, Vol. 32, No 2, pp. 69–83.
- Rönnerberg, L. (2014), 'School inspection in Sweden: historical evolution, resurrection and gradual change', *Revue française de pédagogie*, No 186, pp. 35–45.
- Rönnerberg, L., Lindgren, J. and Lundahl, L. (2019), 'Education governance in times of marketization: the quiet Swedish revolution', in Langer, R. and Brüsemeister, T. (eds), *Handbuch Educational Governance Theorien*, Springer, Dordrecht, the Netherlands.
- Santos, I. and Centeno, V. G. (2021), 'Inspirations from abroad: the impact of PISA on countries' choice of reference societies in education', *Compare: A Journal of Comparative and International Education*, doi:10.1080/03057925.2021.1906206.
- Skedsmo, G., Rönnerberg, L. and Ydesen, C. (2021), 'National testing and accountability in the Scandinavian welfare states: education policy translations in Norway, Denmark and Sweden', in Grek, S., Maroy, C. and Verger, A. (eds), *World Yearbook of Education 2021: Accountability and datafication in the governance of education*, Routledge, London.
- SI (Schools Inspectorate) (2021), 'About the Swedish Schools Inspectorate' (<https://www.skolinspektionen.se/other-languages/english-engelska/>).
- Skolverket (2004), *Internationella studier under 40 år–svenska resultat och erfarenheter*, Skolverket, Stockholm.
- Skolverket (2007), *PISA 2006 – 15-åringars förmåga att förstå, tolka och reflektera – naturvetenskap, matematik och läsförståelse*, Skolverket, Stockholm.
- Skolverket (2019), *PISA 2018 – 15-åringars kunskaper i läsförståelse, matematik och naturvetenskap*, Skolverket, Stockholm.
- Sundberg, D. and Wahlström, N. (2016), 'Den svenska läroplansutvecklingen: Begrepp och tendenser', in Elmgren, M., Folke-Fichtelius, M., Hallsén, S., Román, H. and Wermke, W. (eds), *Att ta utbildningens komplexitet på allvar – En vänskrift till Eva Forsberg*, Uppsala University, Uppsala, Sweden.
- Swedish government (2004), *Att lyfta matematiken–intresse, lärande, kompetens*, SOU 2004:97, Fritzes, Stockholm.
- Swedish government (2014), *Utvärdera för utveckling – om utvärdering av skolpolitiska reformer*, SOU 2014:12, Fritzes, Stockholm.
- Swedish government (2016), *Samling för skolan – nationella målsättningar och utvecklingsområden för kunskap och likvärdighet*, SOU 2016:38, Wolters Kluwers, Stockholm.
- Swedish government (2017), *Samling för skolan – Nationell strategi för kunskap och likvärdighet*, SOU 2017:35, Wolters Kluwers, Stockholm.
- Swedish government (2020), *En mer likvärdig skola – minskad skolsegregation och förbättrad resurstilldelning*, SOU 2020:28, Wolters Kluwers, Stockholm.

- Tveit, S. (2018), 'Ambitious and ambiguous: shifting purposes of national testing in the legitimation of assessment policies in Norway and Sweden (2000–2017)', *Assessment in Education: Principles, Policy & Practice*, Vol. 25, No 3, pp. 327–350, doi:10.1080/0969594X.2017.1421522.
- Vlachos, J. (2019), 'Trust-based evaluation in a market-oriented school system', in Dahlstedt, M. and Fejes, A. (eds), *Neoliberalism and Market Forces in Education: Lessons from Sweden*, Routledge, London.
- Wikström, C. (2005), 'Grade stability in a criterion-referenced grading system: the Swedish example', *Assessment in Education: Principles, Policy & Practice*, Vol. 12, No 2, pp. 125–144.
- Wikström, C. (2006), 'Education and assessment in Sweden', *Assessment in Education: Principles, Policy & Practice*, Vol. 13, No 1, pp. 113–128.

CHAPTER 7

Cross-national achievement surveys and educational monitoring in Finland

Heidi Katarina Harju-Luukkainen, Sari Sulkunen and Minna Maunula

Abstract

This chapter provides information on results of international achievement surveys and their use in monitoring educational outcomes in Finland. The educational monitoring system in Finland differs from that of many other countries. So far, little attention has been paid to its lack of standardised measurements and official control. Furthermore, Finnish education policies are strongly driven not by the aim to reach high rankings in international assessments but rather by the emphasis on quality and equity of education. This chapter describes the Finnish education system, its core curriculum for basic education and teacher training. We also review international achievement survey results and the education policy monitoring of the country in general. At the end of the chapter, we take a closer look at some monitoring results that have had an impact on political reforms and policy changes.

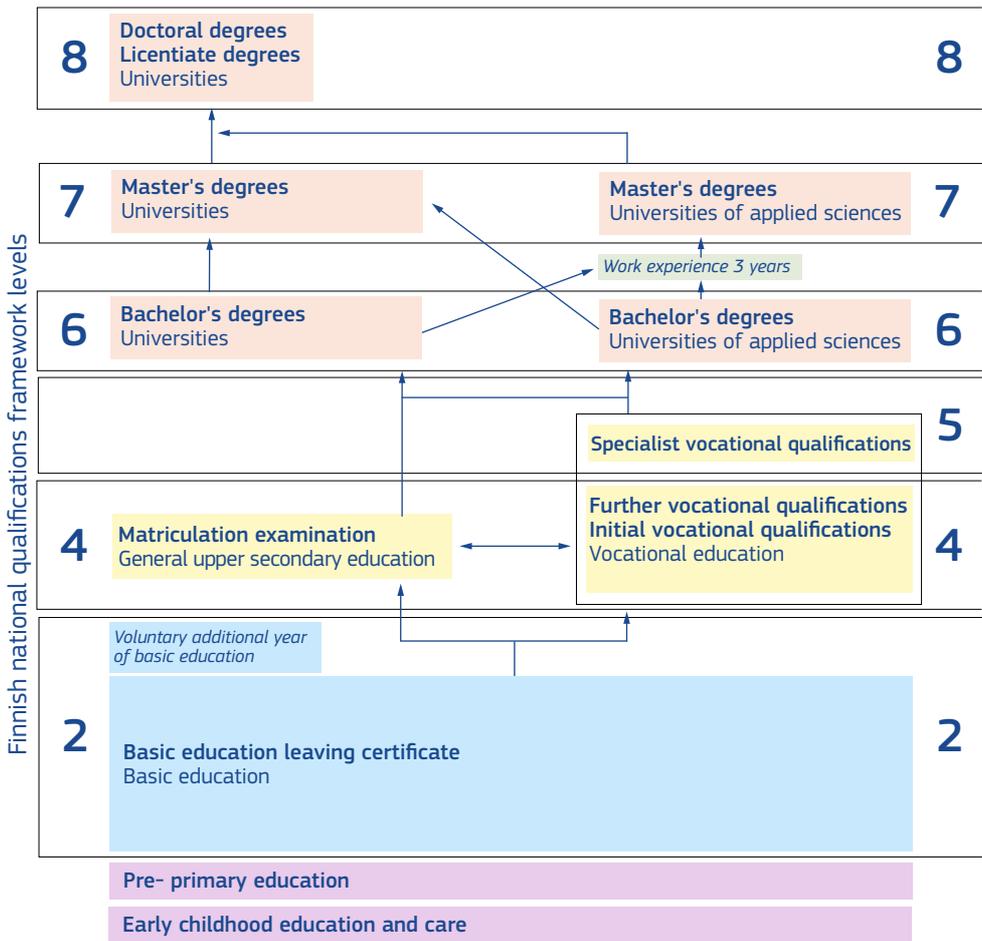
Introduction to the Finnish education system

Along with other Nordic countries (Denmark, Iceland, Norway, Sweden), Finland follows a 'Nordic model', which is also called a 'Nordic dialogue' (Antikainen, 2006; Garvis and Eriksen Ødegaard, 2017; Telhaug et al., 2006). The Nordic countries continually score highly on quality of life measures, and many Nordic countries achieve above-average educational outcomes (Garvis et al., 2019, p. 1). The model aims to provide equal opportunities regardless of students' socioeconomic status or residential area. Although the Nordic education systems are often considered to be alike, there are significant differences between them (Garvis et al., 2019). The previous chapter discussed the Swedish context; this chapter focuses on the Finnish education system.

The Finnish system is regarded as a cornerstone of the welfare of society, and it aims to offer Finnish citizens equal opportunities in education. However, this goal has not been fully met, as is evident in the international assessment results reviewed below (see also Saarinen et al., 2021). The Finnish education system consists of early childhood education and care, and pre-primary, basic, general upper secondary, vocational, higher and adult education. Compulsory education consists of 1 year of preschool, 9 years of basic education and, since August 2021, also some years of upper secondary education. Children enter preprimary education at the age of 6 and start their basic education the year they turn 7 years old. After basic education, students continue in either general or vocational upper secondary school. Students' compulsory education ends when they reach the age of 18 or when they complete their upper secondary qualifications (Ministry of Education and Culture, 2020). The general structure of the current compulsory education system in Finland was founded in the educational reform at the beginning of the 1970s, and since then its core has been the comprehensive 9-year basic education. However, after that reform, some adjustments were made. For instance, the ability grouping that was widely used during the 1970s was officially abolished in the mid 1980s. Figure 7.1 presents the structure of the Finnish education system in more detail. The arrows show the different educational paths the students can take. The educational paths' main feature is that students can move freely from one level to the next.

In contrast to other European countries, in Finland not only is compulsory education free of charge for students, but the costs of textbooks and other materials, a daily meal, in some cases travel expenses, school healthcare and other welfare services are also covered. In most cases, children attend their local school for compulsory education. However, some schools are specialised in specific subject areas, such as music, languages and sports, and these schools may request that the children sit an entrance examination. The emphasis on children attending the local public school was further strengthened in legislation in 2011 (Thuneberg et al., 2013).

Figure 7.1: The structure of the education system in Finland



Source: Finnish National Agency for Education (2021).

Core curriculum for basic education

The first national core curriculum for basic education was published in 1970 in Finland, and it was a document strongly emphasising the centralised character of the system. After multiple reforms, a direction for decentralisation and teacher autonomy was set (Vitikka et al., 2012), building the foundations of the Finnish curriculum, in which teachers and local authorities are given a large degree of autonomy. The curriculum for basic education is renewed approximately every 10 years. In addition, smaller refinements are made when needed. Every time a curriculum is renewed, the relevant stakeholders, including principals, teachers, education experts, organisations, researchers and policymakers, take part in

the drafting process. In the final phases, a larger audience is also consulted so that, for instance, parents can comment on the draft. Curricular reforms aim to reflect on changes and challenges in the globalised world in order to proactively prepare children for their future (Finnish National Agency for Education, 2016). The latest core curriculum for basic education was implemented for grades 1–6 in 2016 and for grades 7–9 in 2017–2019.

The nationwide core curriculum is considered a starting point for equality in and the quality of the entire education system. The national core curriculum provides a common direction and basis for renewing education and instruction in schools. The fundamental value of the core curriculum in basic education is that every pupil is 'unique and has the right to high-quality education. Pupils are heard, valued and encouraged. They feel that their learning and well-being matter. Pupils are guided towards a sustainable way of life and understanding the importance of sustainable development' (Finnish National Agency for Education, 2016, p. 2). It also has an important role in steering the Finnish education system, and it is a framework for local curricula. The local (as well as national) curriculum serves as a concrete guiding tool for teachers in their work, and it also highlights school-specific needs and opportunities. The curriculum sets out pedagogical content and goals, and the teacher can apply pedagogical freedom in its implementation.

The latest core curriculum (Finnish National Agency for Education, 2016) stipulates content for each school subject. The central aim of the new curriculum is to promote instruction with an integrated approach. The intention is to facilitate students' understanding of the relationships and interdependencies between different types of content. Therefore, the new core curriculum describes seven transversal competence areas (Finnish National Agency for Education, 2016, p. 2). These are:

1. thinking and learning to learn;
2. cultural competence, interaction and expression;
3. taking care of oneself and managing daily life;
4. multiliteracy;
5. ICT competence;
6. working life competence and entrepreneurship;
7. participation, involvement and building a sustainable future.

Furthermore, multidisciplinary learning modules support the dialogue between different subjects, and schools have to organise one module at least once every year.

Teacher training guaranteeing high-quality education

One important aspect of the Finnish education system for guaranteeing a good-quality and equal education for all is the high-quality university-level teacher education. A high level of education for teachers is seen as important, because Finnish teachers work fairly autonomously without pedagogical supervision.

Teacher training in Finland has a long history dating back to the country's independence in 1917. In the 20th century, two important steering documents were developed that helped shape the Finnish education system. First, in 1921 the Act on Compulsory

Education (*Laki oppivelvollisuudesta*)⁽¹¹⁾ was implemented, followed by the development of teacher training (Kuikka, 1988). This legislation required all municipalities throughout the country to provide education to all children aged 7–12 years. Consequently, however, these new requirements for compulsory education and teacher training led to a substantial teacher shortage (Määttä et al., 2013). Because of this shortage, teacher training colleges were created throughout Finland. After several decades, in 1971 the responsibility for teacher education was transferred to universities. The purpose of this shift was to both reassure the education research community and implement high-quality teacher training at master’s level (Harju-Luukkainen et al., 2019).

The main goal of Finnish teacher training is to develop enquiry-oriented teachers (Jyrhämä and Maaranen, 2012) who can adapt to their future local settings. The idea is that teachers will be able to combine theoretical knowledge and practice in the classroom. The teacher education programmes conduct research-based teaching that focuses on building student-focused pedagogical skills and decision-making, especially regarding how to justify one’s own pedagogical decisions (see Kansanen, 2006). In contrast to many other European countries, in Finland the teaching profession is highly valued and competitive, meaning that only the best teachers can enter the profession. Teachers in Finland complete a master’s degree from a research university. For classroom teachers (for grades 1–6), the major subject is education. Subject teachers (grades 7–12) have a subject major and on top of that receive teacher training, called pedagogical studies.

International achievement surveys in Finland

The Finnish education system has repeatedly received international attention due to its students’ relatively high performance level in international educational assessment studies, such as the Programme for International Student Assessment (PISA) organised by the Organisation for Economic Co-operation and Development (OECD), and the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) organised by the International Association for the Evaluation of Educational Achievement (IEA) (see, for example, Mullis et al., 2017; OECD, 2016). In Finland, the average results are high despite relatively moderate average economic investments in education (OECD, 2018). Finnish schools also seem to be relatively equitable in terms of the performance differences between socioeconomic student groups (Willms, 2010). In addition, variation in performance between schools is among the smallest in the world (OECD, 2016).

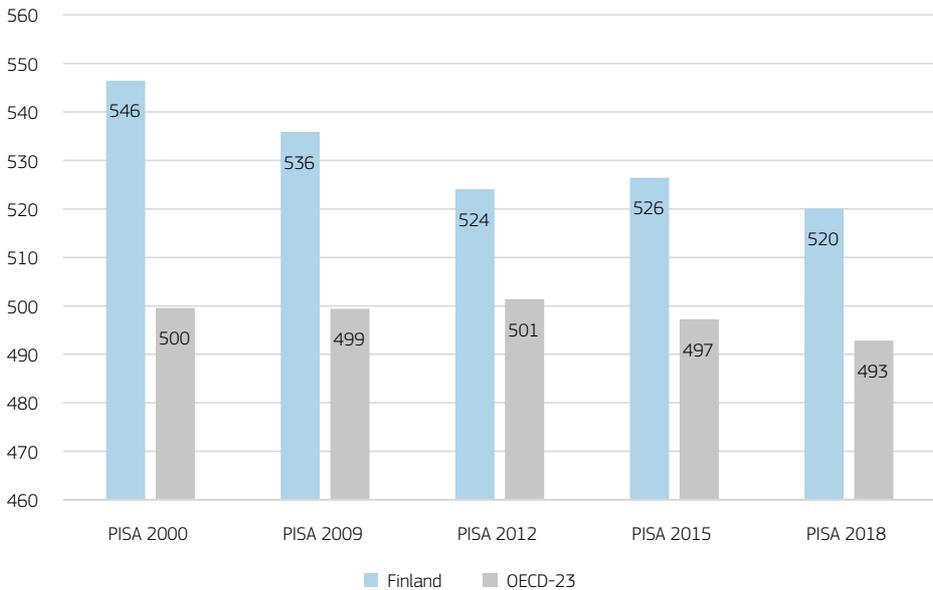
Finland has participated in the PISA assessments since the first cycle in 2000. In addition, Finland took part in the fourth-grade students’ PIRLS reading assessment in 2011 and 2016, and TIMSS in 1999 (seventh graders), 2011 (fourth, seventh and eighth graders), 2015 (fourth graders) and 2019 (fourth and eighth graders). Thus, of all these assessments, PISA provides the most regular data, on 15-year-old students for 20 years, allowing achievement trends to be observed. As PISA started with reading literacy as the

⁽¹¹⁾ <https://www.eduskunta.fi/pdf/saadokset/101-1921.pdf>

main assessment area in 2000 and offers extensive data on main assessment areas at 9-year intervals, the reading achievement of Finnish 15-year-olds can be observed in three cycles, in 2000, 2009 and 2018. Next, we will review the Finnish achievement patterns and trends in PISA, with the main focus on reading. We will then briefly review the trends in the IEA studies PIRLS and TIMSS.

In PISA 2000, the Finnish students outperformed their peers in all other participating countries with a reading mean score of 546, which was well above the OECD mean (500). In 2009, Finnish students' average performance in reading decreased by 10 points, and in 2012 it decreased by a further 12 points (Väljärvi and Sulkunen, 2016). Since then the negative trend has stagnated, and in 2018 the average reading score of Finnish students was 520. Figure 7.2 illustrates the trends in average reading performance in the most recent cycles of PISA (2009–2018) and shows the 2000 mean scores from the first PISA main assessment of reading. Finnish students' average reading performance has been clearly above the OECD average in each cycle. Despite the declining trend, Finland still ranks among the top OECD countries in PISA reading (OECD, 2019a, p. 57).

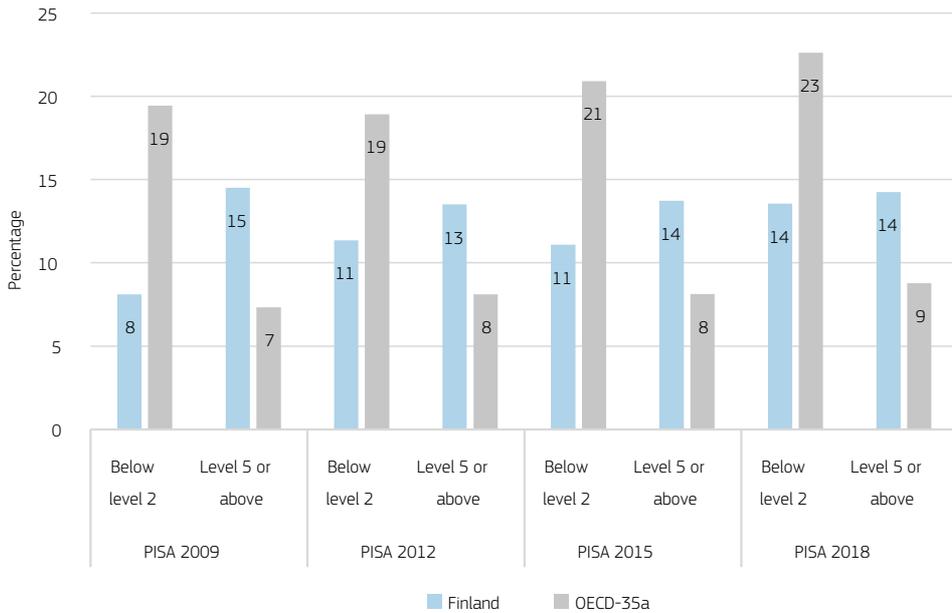
Figure 7.2: Trends in PISA reading scores in Finland and the OECD average (23 countries)



Source: OECD, PISA 2018 database, Table I.B1.10.

In addition to the decline in students' average performance, there has been a steady increase in the proportion of low performers, and in 2018 nearly 14 % of Finnish 15-year-olds were below proficiency level 2 in reading. The increase in the proportion of low performers started in PISA 2009 and has developed since, as illustrated in Figure 7.3. However, the proportion of top performers has remained stable (approximately 14 %). In addition, in the OECD countries on average the trend has been similar, and Finland still has a below-average share of low performers and an above-average share of high performers.

Figure 7.3: Shares of low and high performers in PISA reading in 2009–2018 in Finland and the OECD average (35 countries)



Source: OECD, PISA 2018 database, Table I.B1.7.

Finnish students' performance in mathematics and science in PISA has shown similar trends to those in reading literacy (Väljärvi and Sulkunen, 2016). The average performance in both started to show a significant decline in 2009, which continued in 2012, with the strongest decline in mathematics. Since 2012, the average performance in both mathematics and science has declined further (Leino et al., 2019). Moreover, in both assessment areas the share of low performers has increased and the share of high performers has decreased. In mathematics, Finnish students currently perform above the OECD average, but in science Finland is still among the top OECD countries.

As mentioned above, Finland strives for equal educational opportunities for all. In the early cycles of PISA, the Finnish education system was found to have relatively equitable learning outcomes compared with other OECD countries. Since 2009, however, PISA results have shown slightly less equitable performance. In reading literacy, the standard deviation stayed below the OECD average until 2009 but has since been at the OECD average level (Leino et al., 2019). In addition, in many PISA cycles Finnish students' economic, social and cultural background had a below-average relationship with performance in all areas assessed (OECD, 2010, p. 27; OECD, 2013, p. 37). However, between 2009 and 2018, the performance gap between advantaged students and disadvantaged students increased, for example in reading from 61 points to 79 points. Although the socioeconomic gap in reading in Finland is still below the OECD average (89 points in 2018), it has clearly widened. Furthermore, supporting students with immigrant backgrounds is a challenge for the Finnish educational system, as the gap between these students' reading performance and non-immigrant students' reading performance is among the largest in PISA, even

after accounting for socioeconomic status (OECD, 2019b, p. 185). Further equity concerns arise given the Finnish gender gap in reading, which has been among the widest in PISA since 2000. In 2018, girls outperformed boys by 52 points (OECD, 2019b, p. 143).

Turning to the PIRLS assessment, Finnish fourth graders' performance can currently be observed from 2011 and 2016 results. In both cycles, Finland has been among the top countries, and the average reading performance has been stable (Leino et al., 2017; Mullis et al., 2017). In addition, results related to equity have remained the same since 2011. This holds for the standard deviations illustrating the gap between low- and high-performing students and the gender difference. Furthermore, in the fourth grade, Finnish girls clearly outperform boys as the gender difference (22) is slightly above the international average and one of the largest among European countries. Moreover, the share of students achieving the advanced international benchmark has stayed the same (18 %; the international average is 10 %) since 2011. However, the share of students who did not achieve even the lowest benchmark doubled to 2 % from 2011 to 2016. Still, this was only half of the share of these students on average in PIRLS.

In TIMSS, the most recent Finnish participants are fourth graders in the 2011, 2015 and 2019 rounds and eighth graders in the 2011 and 2019 rounds. In general, in mathematics Finnish students in both age groups have shown above-average performance in each cycle, and in science their performance has consistently ranked among the top countries (Vettenranta et al., 2020a,b). Nevertheless, consistently with PISA there has been a decline since 2011 in both mathematics and science, although for eighth-grade mathematics the decline was not statistically significant. For eighth-grade science, between 2011 and 2019, there was a 9-point decline (from 552 points to 543 points). For fourth graders, there was a 10-point decline in average performance in mathematics from 2011 to 2015 (with mean scores being 545 and 535, respectively) and a 16-point decline in science (with mean scores being 570 and 554, respectively). In both areas, there were no statistically significant changes in fourth-graders' performance from 2015 to 2019 (in 2019, the mathematics mean score was 532 and the science mean score was 555). However, between 2015 and 2019, the share of fourth-grade students below the low international benchmark in both mathematics and science increased by 2 percentage points to 5 % and 3 %, respectively. The share of students reaching the advanced benchmark also increased (by 3 percentage points in mathematics and 2 percentage points in science), resulting in 11 % high achievers in mathematics and 15 % in science in 2019. For eighth graders, there was a 3 percentage point increase in students below the low benchmark in mathematics and a notable 5 percentage point increase in the same figure for science (7 % in maths and 6 % in science in 2019). In science, there was also a 3 percentage point increase in the share of students achieving the advanced benchmark, which resulted in 16 % of Finnish students achieving this benchmark in 2019. All in all, the shares of students achieving TIMSS benchmarks are at a high level internationally.

The results of the assessments, particularly those of early cycles of PISA, showed that the Finnish education system, which aims for educational equity, ranked relatively well in international comparisons, as high-performing students did not do any worse than their peers in other countries, whereas the weakest students outperformed their international counterparts (Hautamäki et al., 2009). Thus, the results may have contributed to basic education legislation changes that have strengthened the main principles of the Nordic

educational ideas even further (Thuneberg et al., 2013). The first PISA results were used as evidence of why the education system did not need extensive reforms. Accordingly, the declining trend that has been observed in both international and national assessments since 2006 (Hautamäki et al., 2013; Leino et al., 2019; Vetterranta et al., 2016) has been taken seriously, and programmes have been launched to turn the trend around. These programmes have included thematic assessments that delve deeper into the details of the national features of the education system, national development programmes on literacy, and additional funding for municipalities and schools to improve their practices. This will be discussed further in the next section.

Education policy monitoring in Finland

In this section, we provide a description of the Finnish monitoring processes at different levels and how achievement results are used to inform and further develop the education system. The Finnish education policy monitoring system differs from that of many other countries. According to Vainikainen and Harju-Luukkainen (2020) very little attention has been paid to the Finnish assessment system and especially to its lack of standardised measurements and official control. However, these factors contribute significantly to the overall functioning of the education system. In practice, the assessment of the education system happens at three different levels. School-based assessments are the most common form of assessment in Finland, and they are conducted by the teachers and schools. These assessments happen across the country at classroom level, every day. The national curriculum for basic education states the following (Finnish National Agency for Education, 2016, p. 46): ‘According to the Basic Education Act, the task of student assessment is to guide and encourage pupils to study and to develop the student’s prerequisites for self-assessment. The student’s learning, work and behaviour must be assessed in a variety of ways. These tasks are the starting point for developing a culture of assessment in primary education. The focus is on assessment that promotes learning.’ This means that teachers assess students summatively and formatively throughout the school year and adjust their teaching methods according to individual students’ needs. Much of the assessment concerns the interaction between teachers and students. In addition, Finnish teachers make sure that students receive feedback that guides and encourages learning from the beginning, as well as information about their progress and skills. Working with students’ families is part of a good evaluation culture. The goals of schoolwork and school assessment practices are discussed with parents and guardians. In grades 1–7, assessment translates into verbal or numerical assessment or a combination thereof according to the decision of the education provider. Numerical assessment is used in grades 8 and 9. The numerical grade is an average, a summative assessment of the student’s level of competencies in relation to the objectives in each subject. In Finland, the national core curriculum for basic education determines the learning objectives for each school subject. Grading guidelines are also given. Objective evaluation at this point is of great importance; the grades obtained in different subjects at the end of compulsory education will largely determine the next steps in the student’s educational pathway.

Furthermore, the Finnish Education Evaluation Centre (FINEEC) has responsibility for conducting national assessments. Among other things, FINEEC is responsible for evaluating

learning outcomes with respect to the distribution of lesson hours and the national core curriculum targets stipulated in the Basic Education Act (628/1998) ⁽¹²⁾. The assessment of learning outcomes is based on sampling the student cohort. Typical sample sizes comprise 5–10 % of the age group, which means that each assessment involves about 4 000–6 000 students (Jakku-Sihvonen, 2013, p. 24). The assessed schools represent around 15 % of all the schools that provide basic education in Finland (Harju-Luukkainen et al., 2016; Ouakrim-Soivio, 2013, p. 21).

According to Harju-Luukkainen et al. (2016), the assessment of learning outcomes can be viewed from many perspectives, and it has different purposes for different target groups. National assessments provide valuable information for the highest educational authorities. In Finland, basic education is expected to secure equal educational opportunities for all students. Therefore, the equity of learning outcomes is studied from several perspectives, such as those of students' gender, region, type of municipality, socioeconomic background and language spoken at school. In principle, reaching the objectives for equal learning opportunities as defined in the national core curriculum should lead to educational equity, so that there would be no statistically significant differences between the learning outcomes of boys and girls, for example, or between different regions in Finland.

From the school's perspective, the national assessments of learning outcomes provide benchmarks for schools to evaluate their own success in reaching their objectives of teaching and learning in different subjects. Schools selected for an assessment receive feedback in the form of reference data on the results and learning-related perceptions of their own students. Because there are no national examinations at the end of basic education, many schools welcome this opportunity to compare their own results and grading practices with the national benchmarks and use the assessment as a tool to develop their instruction in different subjects (Harju-Luukkainen et al., 2016; Ouakrim-Soivio and Kuusela, 2012, p. 13).

Besides the two national monitoring systems, international achievement surveys are conducted regularly in Finland. Nevertheless, their role in the education system's development is still rather small compared with that of national assessments. Reports and research connected to educational achievement surveys are published, and results are discussed publicly. However, these results are not used as actively in the development of the education system as those of national assessments. This issue and the role of different policies and practices in assessment will be discussed in the next section.

Analysis of policies, practices and political reforms

In this section, we will take a closer look at monitoring results that have had an impact on political reforms and/or policy changes, keeping in mind the framework of the entire system. Factors such as a lack of standardised measurements and control (Vainikainen et al., 2017) contribute significantly to the overall functioning of the education system. Furthermore, it is notable that a national central organ that oversees both national

⁽¹²⁾ <https://www.finlex.fi/en/laki/kaannokset/1998/en19980628.pdf>

and international assessment policies, practices and assessment results does not exist. According to Vainikainen and Harju-Luukkainen (2020), this has led in some cases to a situation in which participation in some important international assessments has been irregular and there have been problems with the capability of national sample-based assessments to produce enough comparable data suitable for monitoring trends. Therefore, there are many challenges as well as benefits in the Finnish monitoring system; however, we will focus only on a few of them. According to the Act on Compulsory Education (paragraph 1) ⁽¹³⁾, the aim of compulsory education in Finland is to ‘secure the basic skills and education necessary for everyone in life and society, and to promote equal opportunities to develop oneself in accordance with one’s abilities and needs. In addition, compulsory education aims to raise the level of education and skills, reduce learning gaps and increase educational equality, equity and the well-being of children and young people.’ These premises guide Finnish compulsory education but also contribute to its monitoring policies and practices.

Both national and international assessments inform educational policy in Finland, but their roles vary. The national assessments are curriculum based (see, for example, Kauppinen and Marjanen, 2020), and thus they are relevant, for example, for curriculum development. The international assessments have a different focus. PISA aims to assess competencies that adolescents will need in their future, regardless of the subject area in which these competencies were acquired (OECD, 2019a). As PIRLS and TIMSS are curriculum-based assessments but involve numerous participating countries with varying curricula, the assessments unavoidably cover only the common parts of the curricula. Thus, international assessments can only to a limited degree steer curriculum development, which is an important and in Finland the most essential part of education policy. As mentioned, Finnish educational policies are strongly driven not by the aim of achieving high rankings in international assessments but by the quality and equity of education. Nevertheless, international assessments still have some impact on educational policy, as discussed below.

In the domain of literacy, for example, the excellent results of the PISA 2000 assessment surprised many stakeholders, and the national programme aiming to develop literacy education in Finland (Ministry of Education and Culture, 2000) lost its momentum to some degree. This aptly demonstrates how the top results in PISA were followed by a period during which only little effort was made to improve literacy education. However, this changed after the declining results in PISA 2009. Since then, literacy has become a key topic of discussion, and there have been numerous policy measures to improve the quality of literacy education. One of the key elements in developing high-quality literacy education is the national curriculum for basic education, as it has an impact on all children and adolescents. During curriculum development, concerns arising from declining PISA results contributed, in part, to introducing multiliteracy as a transversal competence area in the national curriculum. In practice, this meant that multiliteracy – the competence to interpret, produce and evaluate various kinds of texts in multiple contexts (Finnish National Agency for Education, 2016, pp. 21–22) – became a core area

⁽¹³⁾ <https://www.finlex.fi/fi/laki/ajantasa/2020/20201214?search%20%5Btype%20%5D=pika&search%20%5Bpika%20%5D=laki%20%20oppivelvollisuudesta>

to be taught in all school subjects. Teaching discipline-specific language and literacy both intensifies literacy education and supports content learning at the same time (Fang and Coatoam, 2013). However, multiliteracy as a transversal competence also shows how curriculum development has not been limited to reading literacy as defined in PISA (OECD, 2019a) but has been influenced by the topical research literature (Cope and Kalantzis, 2009) as well as the needs of society and working life in a globalised world.

In addition to curriculum development, literacy has also been the focus in many other policy initiatives during recent years. Some of the initiatives are strategic and provide guidelines for nationwide literacy stakeholders, including but not limited to those in school education. These initiatives include, for example, the Literacy Forum – a group of literacy researchers, practitioners and other literacy stakeholders who provided guidelines for supporting the reading engagement and literacy competence of Finnish children and adolescents in 2018 (Literacy Movement, 2018; Ministry of Education and Culture, 2017). The work of the Literacy Forum is currently being updated, as the Ministry of Education and Culture has established a task force to prepare the national literacy strategy, which was to be finalised in autumn 2021 (Finnish National Agency for Education, 2021). Based on the guidelines of the Literacy Forum, the National Agency for Education established an ongoing governmental programme – Literacy Movement – to coordinate and support the numerous literacy activities by various stakeholders in Finland.

Other initiatives are shorter development programmes, such as the Literacy Movement at School (2019). These programmes provide schools and teachers with resources and support for developing their literacy activities and innovative practices. The downside of these initiatives is that they benefit primarily those schools that apply and are accepted for the programmes, and they suffer from a limited funding period. However, the practices developed may be disseminated to other schools and thus may benefit a broader target group in the future.

Concern about declining results has turned into vigorous development of policy measures and practices, which can be considered a positive consequence of international assessments. Their role, however, should not be overestimated. Although the results of these assessments, particularly PISA, are often referred to in arguments for new initiatives, they are not the sole contributing factor but part of a larger landscape of educational policy development, and they do not have a straightforward relationship with policy initiatives (Seppänen et al., 2019). In fact, many researchers in Finland (see, for example, Kivinen and Hedman, 2017) caution that PISA does not really tell us anything about the success of educational policy in a country, as it gathers very few data on education as such. Therefore, in Finland, the development of the education system is not limited to comparing competence levels across countries using instruments such as PISA but tied up with the development of the overall quality of the system, including in relation to international large-scale assessments.

Moreover, there are also some negative and unintended consequences of the discussion related to international assessment results. In reading literacy, for example, the large gender difference reported in PISA reading has dominated public discussions in Finland

since PISA 2000. In public discussions, boys are considered low performers and girls are considered high achievers based on the two gender groups' average results (Sulkunen and Kauppinen, 2018). These frequently repeated statements are powerful and even harmful, as they strengthen gender stereotypes and obscure the socially defined nature of gender norms (White, 2007). In education, there is a risk that low-performing girls and high-performing boys will go unnoticed and will be left without support because of the gender stereotypes. Moreover, such portrayals of achievement survey results obscure other social factors related to equity and the real risk groups among both boys and girls (Brozo, 2019). It is also noteworthy that the discussion on boys' literacy has not proved to be fruitful in solving the issue, as the gender gap still persists (OECD, 2019b, p. 147). Furthermore, there has also been criticism of international assessments, particularly those adopting the OECD's discourse on literacy as a prerequisite for an individual's economic success and participation in society – as critics consider it too narrow and one-sided (see, for example, Mäkinen, 2021) – and equating PISA results with the quality of education (Kivinen and Hedman, 2017). This criticism is presented mainly by researchers. However, there is also awareness among policymakers and decision-makers that the international large-scale assessments do not provide enough data to solve educational issues. For example, the large gender gap for Finnish adolescents has been verified in multiple assessments with limited information on the causes. Thus, with a mandate from the Ministry of Education and Culture, FINEEC further studied the factors causing the gender gap to identify norms, structures and practices related to gender equality (Saarinen et al., 2021).

In summary, in the Finnish context, the lack of standardised examinations gives more freedom to schools and teachers to implement the curriculum in a way that supports individual students' educational outcomes in accordance with the aim of the education system (see, for example, Volante et al., 2020). However, the declining trend in international assessments and the increase in regional differences (Harju-Luukkainen et al., 2020; Vettenranta et al., 2016) is challenging the Finnish monitoring system. The Finnish school system is still maintaining a good level of quality nationwide, and it aims to provide an equal starting point for all pupils. However, as illustrated in this chapter, this aim has not been fully met, particularly when we look at equity of learning outcomes (and not only, for example, learning opportunities). Indeed, recent educational achievement results have shown that the education system needs to be developed further to keep its high ranking in international comparisons, if this is what is highlighted, and to support its unique qualities that have been found to be effective. In the future, the Finnish education system could benefit from a broader national discussion and a clearer national framework for monitoring practices on students' educational outcomes and equality.

REFERENCES

- Antikainen, A. (2006), 'In search of the Nordic model in education', *Scandinavian Journal of Educational Research*, Vol. 50, No 3, pp. 229–243.
- Brozo, W. (2019), *Engaging teen and preteen boys in active literacy – Evidence and practice*, Cambridge University Press, Cambridge.
- Cope, B. and Kalantzis, M. (2009), "Multiliteracies": new literacies, new learning, pedagogies', *International Journal*, Vol. 4, No 3, pp. 164–195.
- Fang, Z. and Coatoam, S. (2013), 'Disciplinary literacy: what you want to know about it?', *Journal of Adolescent and Adult Literacy*, Vol. 56, No 8, pp. 627–632.
- Finnish National Agency for Education (2016), *Perusopetuksen opetussuunnitelman perusteet 2014 (Curriculum for Basic Education 2014)* (https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf).
- Finnish National Agency for Education (2021), 'Kansallinen lukutaitostrategia vahvistamaan lukutaitotyötä' ('National literacy strategy to strengthen literacy work') (<https://www.oph.fi/fi/uutiset/2021/kansallinen-lukutaitostrategia-vahvistamaan-lukutaitotyota>).
- Garvis, S. and Eriksen Ødegaard, E. (2017), *Nordic Dialogues on Children and Families – Evolving families*, Routledge, Abingdon-on-Thames, UK.
- Garvis, S., Harju-Luukkainen, H., Sheridan, S. and Williams, P. (2019), *Nordic Families, Children and Early Childhood Education*, Palgrave Macmillan, London.
- Harju-Luukkainen, H., Vettenranta, J., Oukrim-Soivio, N. and Bernelius, V. (2016), 'Differences between PISA reading literacy scores and grading for mother tongue and literature at school: a geostatistical analysis of the Finnish PISA 2009 data', *Education Inquiry*, Vol. 7, No 4, pp. 463–479.
- Harju-Luukkainen, H., Wang, J. and La Torre, D. (2019), 'Using content analysis to compare a U.S. urban teacher residency to a Finnish teacher education program', *Urban Review*, Vol. 51, pp. 247–269.
- Harju-Luukkainen, H., Vettenranta, J., Wang, J. and Garvis, S. (2020), 'Family related variables effect on later educational outcome: a geospatial analysis on TIMSS 2015 Finland', *Large-Scale Assessments in Education*, Vol. 8, No 3.
- Hautamäki, J., Hautamäki, A. and Kupiainen, S. (2009), 'Educational equity account in Nordic countries', in Matti, T. (ed.), *Northern Lights on PISA 2006: Differences and similarities in the Nordic countries*, TemaNord 2009:547, Nordic Council of Ministers, Copenhagen, pp. 157–167.
- Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M-P. and Hotulainen, R. (2013), *Oppimaan oppiminen peruskoulutun päättövaiheessa. Tilanne vuonna 2012 ja muutos vuodesta 2001*, Tutkimuksia 347, University of Helsinki, Helsinki.
- Jaku-Sihvonen, R. (2013), 'Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista' ('On assessment systems for learning outcomes and developmental challenges of these systems'), in Räisänen, A. (ed.), *Oppimisen arvioinnin kontekstit ja käytännöt (Contexts and Practices for the Assessment of Learning)*, Reports and studies 2013:3, Finnish National Agency for Education, Helsinki, pp. 13–36.

- Jyrhämä, R. and Maaranen, K. (2012), 'Research orientation in a teacher's work', in Niemi, H., Toom, A. and Kallioniemi, A. (eds), *Miracle of Education: The principles and practices of teaching and learning in Finnish schools*, Sense Publishers, Rotterdam, pp. 97–114.
- Kansanen, P. (2006), 'Constructing a research-based program in teacher education', in Oser, F. K., Achtenhagen, F. and Renold, U. (eds), *Competence Oriented Teacher Training – Old research demands and new pathways*, Sense Publishers, Rotterdam and Taipei, pp. 11–22.
- Kauppinen, M. and Marjanen, J. (2020), *Millaista on yhdeksäsluokkalaisten kielellinen osaaminen? Suomen kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2019 (Achievement in mother tongue and literature at the end of basic education)*, No 13:2020, Finnish Education Evaluation Centre, Helsinki.
- Kivinen, O. and Hedman, J. (2017), 'Moniselitteiset PISA-tulokset ja niiden ongelmalliset koulutuspoliittiset tulkinnat' ('Ambiguous PISA results and their problematic interpretations in education policy'), *Politiikka*, Vol. 59, No 4, pp. 250–263.
- Kuikka, M. T. (1988), 'Opiskelijoiden rekrytointi maamme kansakouluseminaareihin 1920- ja 1930-luvuilla' ('The recruitment of the students for our country's primary schoolteacher colleges in the 1920s and 1930s'), in Jouko, K. (ed.), *Tutkimuspohjaista koulutusta kohti – Professori Veikko Heinosen juhla kirja 7.5.1988 (Towards Research Base Education – The yearbook of Professor Veikko Heinonen 7.5.1988)*, University of Jyväskylä, Jyväskylä, Finland, pp. 54–80.
- Leino, K., Nissinen, K., Puhakka, E. and Rautopuro, J. (2017), *Lukutaito luodaan yhdessä: kansainvälinen lasten lukutaitotutkimus (PIRLS 2016)*, Finnish Institute for Education, University of Jyväskylä, Jyväskylä, Finland (<http://urn.fi/URN:ISBN:978-951-39-7292-9>).
- Leino, K., Ahonen, A., Hienonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S. et al. (2019), *PISA 2018 ensituloksia – Suomi parhaiden joukossa (PISA 2018 First Results – Finland among the best)*, No 2019:40, Ministry of Education and Culture, Helsinki.
- Literacy Movement (2018), 'Literacy Movement promotes the literacy and multiliteracy of Finnish children and youth' (<https://lukuliike.fi/the-literacy-movement/>).
- Literacy Movement at School (2019), *Lukuliike koulussa -kampanjan loppuraportti (Final report of the Literacy Movement at School campaign)* (<https://lukuliike.fi/wp-content/uploads/2019/12/Lukuliike-koulussa-kampanjan-loppuraportti-2019.pdf>).
- Määttä, K., Uusiautti, S. and Paksuniemi, M. (2013), *What are Finnish teachers made of? A glance at teacher education in Finland formerly and today*, Nova Science Publishers, Hauppauge, New York.
- Mäkinen, M. (2021), 'Perustaidot osallisuuteen kompleksisessa maailmassa' ('Basic skills to participation in a complex world'), *Finnish Journal of Education Kasvatus*, Vol. 52, No 1, pp. 116–121.
- Ministry of Education and Culture (2000), *Finland Can Read – Literacy in the information age*, No 4:2000, Ministry of Education and Culture, Helsinki.
- Ministry of Education and Culture (2017), 'The Minister of Education and Culture is worried about the weakening of literacy – a National Literacy Forum is set to challenge the declining of literacy and people's willingness to read', press release, 31 October (https://minedu.fi/-/kansallinen-lukutaitofoorumi-vastaamaan-lukutaidon-ja-lukuinnon-haasteeseen?languageId=en_US).

- Ministry of Education and Culture (2020) 'Extension of compulsory education' (<https://minedu.fi/en/extension-of-compulsory-education>).
- Ministry of Education and Culture (2021), 'The Finnish model' (<https://minedu.fi/en/the-finnish-model>).
- Mullis, I. V. S., Martin, M. O., Foy, P. and Hooper, M. (2017), *PIRLS 2016 International Results in Reading*, TIMSS & PIRLS International Study Center, Boston College, Boston, MA (<http://timssandpirls.bc.edu/pirls2016/international-results/>).
- Ouakrim-Soivio, N. (2013), *Toimivatko päättöarviointin kriteerit? Oppilaiden saamat arvosanat ja Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiseröjen mittareina (Do the national criteria for students' final grades work? Teachers' grading and national assessments as indicators for between-school differences)*, Finnish National Agency for Education, Helsinki.
- Ouakrim-Soivio, N. and Kuusela, J. (2012), *Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011 (Learning outcomes of history and social studies in the final stage of basic education in 2011)*, Finnish National Agency for Education, Helsinki.
- OECD (Organisation for Economic Co-operation and Development) (2010), *PISA 2009 Results (Volume IV): What makes a school successful? Resources, policies and practices*, OECD Publishing, Paris.
- OECD (2013), *PISA 2012 Results (Volume II): Excellence through equity: Giving every student the chance to succeed*, OECD Publishing, Paris.
- OECD (2016), *PISA 2015 Results (Volume I): Excellence and equity in education*, OECD Publishing, Paris, doi:10.1787/9789264266490-en.
- OECD (2018), *Education at a Glance 2018: OECD indicators*, OECD Publishing, Paris, doi:10.1787/eag-2018-en.
- OECD (2019a), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (<https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>).
- OECD (2019b), *PISA 2018 Results (Volume II): Where all students can succeed*, OECD Publishing, Paris (<https://www.oecd.org/publications/pisa-2018-results-volume-ii-b5fd1b8f-en.htm>).
- Saarinen, J., Siekkinen, K., Laimi, T., Ahonen, A., Bemelius, V., Brunila, K. et al. (2021), *Tasa-arvo teoiksi ja todeksi – Sukupuolten sisäisten ja välisten oppimistuloseröjen syitä ja taustoja perusopetuksessa*, No 19:2021, Finnish Education Evaluation Centre, Helsinki. (https://karvi.fi/wp-content/uploads/2021/08/KARVI_1921.pdf).
- Seppänen, P., Rinne, R., Kauko, J. and Kosunen, S. (2019), 'The use of PISA results in education policy-making in Finland', in Waldow, F. and Steiner-Khamsi, G. (eds), *Understanding PISA's Attractiveness: Critical analyses in comparative policy studies*, Bloomsbury, London, pp. 137–159.
- Sulkunen, S. and Kauppinen, M. (2018), "Ainakin muutaman pojan luokallani tiedän lukevan": poikien (osin piiloiset) luku- ja kirjoitustaidot' ('Boys' partly hidden literacies'), in Kivijärvi, A., Huuki, T. and Lunabba, H. (eds), *Poikatutkimus*, Vastapaino, Tampere, Finland, pp. 146–173.
- Telhaug, A. O., Mediås, O. A. and 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*, Vol. 50, No 3, pp. 245–283.

- Thuneberg, H., Vainikainen, M.-P., Ahtiainen, R., Lintuvuori, M., Salo, K. and Hautamäki, J. (2013), 'Education is special for all – the Finnish support model', *Gemeinsam leben*, Vol. 21, No 2, pp. 67–78.
- Vainikainen, M.-P. and Harju-Luukkainen, H. (2020), 'Educational assessment in Finland', in Harju-Luukkainen, H., McElvany, N. and Strang, J. (eds), *Monitoring of Student Achievement in the 21st Century – European policy perspectives and assessment strategies*, Springer, Cham, Switzerland, pp. 131–142.
- Vainikainen, M.-P., Thuneberg, H., Marjanen, J., Hautamäki, J., Kupiainen, S. and Hotulainen, R. (2017), 'How do Finns know? Educational monitoring without inspection and standard-setting', in Blömeke, S. and Gustafsson, J.-E. (eds), *Standard Setting: International state of research and practices in the Nordic countries*, Springer, Cham, Switzerland, pp. 243–259.
- Väljärvi, J. and Sulkunen, S. (2016), 'Finnish school in international comparison', in Niemi, H., Toom, A. and Kallioniemi, A. (eds), *Miracle of Education: The principles and practices of teaching and learning in Finnish schools*, Sense Publishers, Rotterdam, pp. 3–21.
- Vettenranta, J., Väljärvi, J., Ahonen, A., Hautamäki, J., Hiltunen, J., Leino, K. et al. (2016), *PISA 2015 ensituloksia – Huipulla pudotuksesta huolimatta (PISA 2015 First Results – At the top despite the drop)*, No 2016:41, Finnish National Agency for Education, Helsinki (<https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/79052/okm41.pdf>).
- Vettenranta, J., Hiltunen, J., Kotila, J., Lehtola, P., Nissinen, K., Puhakka, E. et al. (2020a), *Perustaidoista vauhtia koulutielle – Neljännen luokan oppilaiden matematiikan ja luonnontieteiden osaaminen – Kansainvälinen TIMSS-tutkimus Suomessa (Fourth-grade Students' TIMSS Results in 2019)*, Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland.
- Vettenranta, J., Hiltunen, J., Kotila, J., Lehtola, P., Nissinen, K., Puhakka, E. et al. (2020b), *Tulevaisuuden avaintaidot puntarissa – Kahdeksannen luokan oppilaiden matematiikan ja luonnontieteiden osaaminen – Kansainvälinen TIMSS-tutkimus Suomessa (Eighth-grade Students' TIMSS Results in 2019)*, Finnish Institute for Educational Research, University of Jyväskylä, Jyväskylä, Finland.
- Vitikka, E., Krokfors, L. and Hurmerinta, E. (2012), 'The Finnish national core curriculum: structure and development', in Niemi, H., Toom, A. and Kallioniemi, A. (eds), *Miracle of Education: The principles and practices of teaching and learning in Finnish schools*, Sense Publishers, Rotterdam, pp. 83–96.
- Volante, L., DeLuca, C., Baker, E., Heritage, M., Schneider, C., Stobart, G. et al. (2020), 'Synergy and tension between large-scale and classroom assessment: international trends', *Educational Measurement: Issues and Practice*, Vol. 39, No 4, pp. 21–29.
- White, B. (2007), 'Are girls better readers than boys? Which boys? Which girls?', *Canadian Journal of Education*, Vol. 30, No 2, pp. 554–581.
- Willms, J. D. (2010), 'School composition and contextual effects on student outcomes', *Teachers College Record*, Vol. 112, No 4, pp. 1007–1037.

CHAPTER 8

Cross-national achievement surveys and educational monitoring in Ireland

Anne Looney, Michael O’Leary, Darina Scully and Gerry Shiel

Abstract

This chapter begins with an overview of how compulsory schooling in Ireland is organised. Ireland’s patterns of performance across various international surveys of achievement at primary and post-primary levels as well as adult literacy studies are then described and evaluated. Although the analysis makes it clear that, in general, between 1995 and 2018, Irish students performed well in comparison with their international counterparts, Programme for International Student Assessment (PISA) 2009 was an exception. At the time, the term ‘PISA shock’ was used to describe a set of poor results that challenged prior notions about high standards of achievement in Ireland and accelerated policy change. As we discuss, the most prominent example of political reform in response to international survey results in Ireland was the introduction of the 2011–2020 national literacy and numeracy strategy by the Department of Education and Skills in 2011. The strategy set out a number of planned actions, and the outcomes of these actions are evaluated towards the end of the chapter. The chapter concludes by considering issues pertinent to the usefulness of cross-national achievement surveys in supporting the process of educational monitoring in Ireland.

Introduction

In Ireland, the period of compulsory education spans ages 6 to 16 or until students have completed 3 years of post-primary education (see Table 8.1). However, almost all children begin primary school in the September following their fourth or fifth birthday, and most (90 % and over) complete their secondary education by 18 years of age. The Department of Education (DES), led by the Minister for Education, is responsible for overall educational policy and funding for the system. Readers should note that, in Ireland, the DES was known as the Department of Education and Science between October 1997 and May 2010, and the Department of Education and Skills between May 2010 and October 2020. The abbreviation DES has been used for both.

Table 8.1: The Irish educational system

Doctoral degree (3+ years) (level 10)		
Master's degree (1 or 2 years) (level 9)	Postgraduate diploma (1 or 2 years) (level 9)	
Third-level professional courses (including integrated bachelor's and master's) (4+ years) (level 8)	Post-degree higher diploma (level 8)	Honours bachelor's degree (4 years) (level 8)
Post-leaving certificate course (1 or 2 years) (levels 5–6)	Advanced or higher certificate (1–2 years) (levels 5–6), leading, in some cases, to a bachelor's degree	Third-level bachelor's degree (3 years) (level 7)
Upper secondary (senior cycle) (grades 10 (optional), 11 and 12; established, applied and vocational options) (levels 3–4)		
Lower secondary education (junior cycle, grades 7–9) (level 2)		
Primary education (infant classes, ages 4–6; grades 1–6, ages 7–12) (level 1)		
Early years education 0–6 years (including government-funded early childhood care and education programme (ages 3–6))		

NB: Levels refer to the national framework of qualifications (<https://nfq.qqi.ie/>).

Primary education

The majority of primary schools in Ireland are owned and managed by churches or religious communities, with a small minority under direct state ownership through a network of local education and training boards. A further and growing sector of primary education is under the patronage of voluntary trustee groups, particularly those promoting multidenominational schools. The state pays most of the running costs, including upkeep of buildings and teachers' salaries, irrespective of ownership. The state-funded schools include religious schools, non-denominational schools, multidenominational schools, Irish language-medium schools (called *Gaelscoileanna* outside predominantly Irish-speaking areas) and special schools. There is also a very small number of private schools that do not get any state funding.

Primary education consists of an 8-year cycle spanning junior infants, senior infants and first class to sixth class. The stated aims for primary education include nurturing the spiritual, moral, cognitive, emotional, imaginative, aesthetic, social and physical dimensions of every child's life in a way that develops their potential to the full, encourages a love of learning and helps them to contribute to the good of society. All schools follow the primary curriculum (1999), which is designed to cater for the different needs of individual children by providing a broad learning experience and a rich variety of approaches to teaching and learning.

Approximately one in five primary schools in Ireland has disadvantaged status (under the Delivering Equality of Opportunity in Schools (DEIS) initiative). These schools benefit from additional state funding supporting measures to address educational disadvantage (e.g. smaller class sizes and additional teachers).

The primary curriculum, including individual subjects, is structured as follows:

- languages: Irish (*Gaeilge*) and English;
- mathematics;
- social, environmental and scientific education: history, geography and science;
- arts education: visual arts, music and drama;
- physical education;
- social, personal and health education.

Religious or ethical education is also part of the curriculum but is the responsibility of the different school owners.

Primary schools are required to be open for a minimum of 183 days per year. The period of instruction per day for classes 1–6, excluding religious/ethical education, must be at least 4 hours and 10 minutes. The ratio of teachers to students is 15:1, and average class size is 25 (DES, 2020a, p. 9).

The National Council for Curriculum and Assessment, a statutory body of the DES that provides advice to the Minister for Education on curriculum and assessment matters, is currently leading the process of reviewing and revising the current primary curriculum (see <https://ncca.ie/en/>).

Post-primary education

Students in Ireland transfer from primary to post-primary education when they are 12 or 13 years of age, and most attend one of three types of school: a voluntary secondary school (approximately 60 % of students), an education and training board school (approximately 25 % of students) or a community/comprehensive school (approximately 15 % of students). Although all voluntary secondary schools are privately owned and managed by churches or religious communities or private voluntary organisations, the state funds most of the running costs including upkeep of buildings and teachers' salaries. A small proportion of voluntary secondary schools (approximately 7 % of all schools) are fee charging and do not receive state subvention other than for teachers' salaries. A small percentage of students (approximately 3 %) attend *Gaelcholáistí* – schools in which all students study all subjects (except English and foreign languages) through the medium of Irish. Some students attend fee-charging private 'grind' schools that operate outside the state sector and are focused predominantly on preparing students for the state examinations. Approximately one in four post-primary schools in Ireland has DEIS status.

Post-primary schools are required to be open for a minimum of 167 days per year. A typical school day comprises 6 hours of instruction. The ratio of teachers to students is 14:1 (DES, 2020a, p.9).

Post-primary education in Ireland is divided into two distinct phases, known as the junior cycle (the lower secondary phase lasting 3 years) and the senior cycle (the upper secondary phase lasting 2 or 3 years, depending on whether or not students opt to participate in a transition year programme).

Junior cycle

The main objective of the junior cycle is for students to complete a broad and balanced curriculum that meets their individual needs and supports them in developing the knowledge and skills necessary to progress to senior cycle education. A process to reform the junior cycle programme, led by the National Council for Curriculum and Assessment (NCCA) and begun in 2014, features newly developed subjects and short courses, a focus on literacy, numeracy and other key skills, and new approaches to assessment and reporting. This work is ongoing.

The full range of subjects available in the junior cycle can be viewed on the NCCA's website (<https://curriculumonline.ie/Junior-cycle/>). Students typically study between 8 and 10 subjects, as well as some short courses. All students study English, Irish and mathematics, and these subjects are assessed at ordinary and higher levels. For other subjects, there is just one common level. Junior cycle achievement is assessed formally using a mixture of school-based tasks (10 %) and a state-certified examination (90 %) at the end of year 3. Outcomes are recorded separately for the two modes of assessment in a profile of achievement. The State Examinations Commission (SEC), a non-departmental public body under the aegis of the DES, is responsible for all aspects of the junior cycle examination process.

Senior cycle

Students in the senior cycle span the ages of 15 to 18. Following completion of the junior cycle programme, approximately two of every three students opt to participate in the transition year programme. Ireland is one of the few countries in the world to offer such a programme (South Korea being another). The programme is intended to be non-academic and designed to provide students with the freedom to participate in a wide range of extracurricular activities, including work experience. During their final 2 years in post-primary education, students take one of three programmes of study.

- The leaving certificate established is a broad and balanced programme of study usually comprising at least eight subjects over 2 years, with some specialisation offered depending on the range of subject choices offered.
- The leaving certificate vocational programme has a stronger vocational dimension and combines leaving certificate subjects with some vocational and enterprise education modules.
- The leaving certificate applied is taken by students wishing to follow a programme with a strong practical and vocational emphasis. It is not recognised for direct entry to third-level courses, but it can enable students to engage in further education and training courses.

Each year in June, about 55 000 students sit a minimum of six leaving certificate established examination subjects, taken at either higher or ordinary level (Irish and mathematics are also available at foundation level). The number of students taking the leaving certificate applied examinations is less than 3 000. The process is overseen by the State Examinations Commission and results are issued in mid August. If students have applied for a place in higher education, their grades on the leaving certificate examination are converted into a points system, processed by the Central Applications Office (CAO) and used as the criterion for entry to undergraduate programmes. CAO points are awarded on a set of sliding scales. For example, a grade H1 (the highest grade) on a higher-level paper converts to 100 points, and a grade H5 at higher level and grade O1 at ordinary level both convert to 56 points. Scores are aggregated across a student's six best subjects. A 25-point bonus is applied in the case of grades H6 or above on higher-level mathematics, meaning that, for that subject, the maximum number of points available is 125. Hence the maximum points score for entry into higher education is 625 points. The powerful influence of the leaving certificate examination on teaching and learning in post-primary education was highlighted by an Organisation for Economic Co-operation and Development (OECD) review of the Irish education system in 1991 (Coolahan, 2017; OECD, 1991) and continues to be a concern even now (OECD, 2021). In 2020, due to the COVID-19 pandemic, students' exam grades were calculated on the basis of their teachers' estimated marks and class ranks, followed by national standardisation to ensure comparability across schools (DES, 2020b).

Ireland's participation and general performance patterns in international achievement surveys

Ireland's participation in international comparative studies of achievement during compulsory schooling began in 1971 with involvement in a civics education survey across 10 countries (Torney et al., 1975). Following that, Ireland participated in five more studies up to the early 1990s: the Six Subject Study (Passow et al. 1976); the Second International Mathematics Study (Travers and Westbury, 1989); the First International Assessment of Educational Progress in Mathematics and Science (Lapointe et al., 1989); the Second International Assessment of Educational Progress in Mathematics and Science (Lapointe et al., 1992a,b); and the International Association for the Evaluation of Educational Achievement Study of Reading Literacy (Elley, 1992). According to Coolahan (2017), the OECD's 1991 review of the Irish system revitalised debate on a range of educational issues (e.g. teacher education, role of the inspectorate) and was instrumental in decisions ensuring Ireland's continuing participation in large-scale studies featured in this chapter.

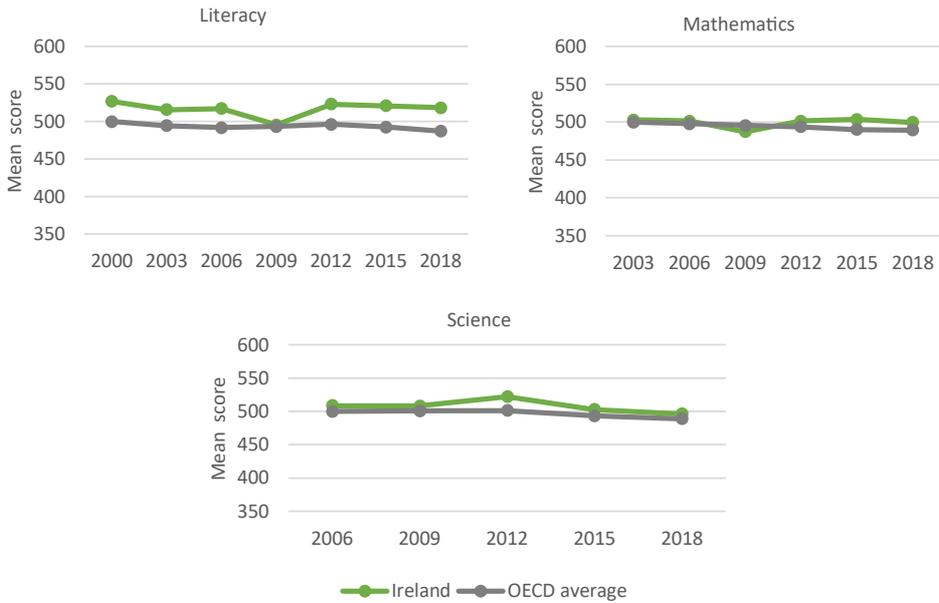
Programme for International Student Assessment

Ireland has participated in all seven cycles of the Programme for International Student Assessment (PISA) since its inception in 2000. Over 5 000 students across 157 schools participated in the most recent iteration in 2018, and the country is also due to take part in 2022. The implementation of PISA in Ireland is overseen by the Educational Research Centre (ERC), on behalf of the DES.

Aside from notable anomalies observed with respect to literacy and mathematics in 2009, Ireland's performance in PISA over the years can be described as relatively stable (Figure 8.1) ⁽¹⁴⁾. Performance in literacy and science has been strong, with mean scores significantly above the OECD average observed in both domains in every cycle (with the exception of 2009, when Ireland's literacy score did not differ from the OECD average). Mathematics performance was generally moderate in the earlier years; however, since 2012, mathematics scores have also significantly exceeded the OECD average on a consistent basis.

⁽¹⁴⁾ Accurate trend data cannot be established until after a domain has been the major domain; hence trends in mathematics and science performance can be examined only from 2003 onwards and 2006 onwards, respectively.

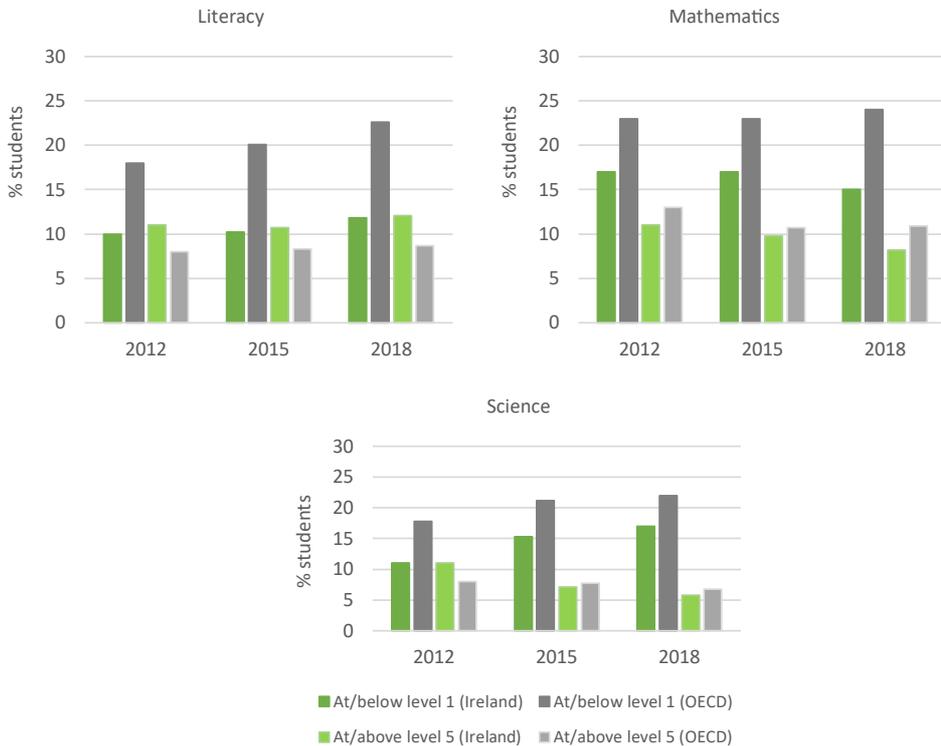
Figure 8.1: Trends in Ireland's performance in literacy, mathematics and science in PISA, 2000–2018



The decline in literacy and mathematics performance in 2009 raised concerns about the national-level performance at the time. However, performance returned to previous levels in 2012. In 2018, Ireland achieved mean scores of 518.1 in literacy, 499.6 in mathematics and 496.1 in science (ranking 8th, 21st and 22nd, respectively, out of 78 participating countries). A noteworthy characteristic of Ireland's performance in PISA over the years is that there has tended to be a relatively narrow spread of achievement. In 2018, for example, the standard deviation of Ireland's performance in literacy was just 90.7, compared with the OECD average of 99.4. In mathematics, it was 77.8, compared with the OECD average of 90.6. Similar differences have been observed in all three domains across the past few cycles.

This narrow spread is primarily due to the relatively strong performance of Ireland's lower-achieving students. This is evident when considering the proportions of students performing at particularly low and high levels of proficiency. In all three domains, there is a clear pattern over the past few cycles whereby the percentages of students in Ireland achieving scores at or below proficiency level 1 have been significantly lower than those on average across all OECD countries (Figure 8.2). The percentages of those scoring at or above level 5 have also generally been lower than the OECD averages in the case of mathematics, despite overall average performance in recent cycles. In literacy, however, the opposite has been the case: the proportion of high achievers in Ireland has tended to be higher. In science, no differences between the proportions of high achievers in Ireland and those across all OECD countries have been observed in recent years.

Figure 8.2: Proportions of students achieving at the lowest and highest levels of proficiency in literacy, mathematics and science, compared with OECD averages, in the 2012, 2015 and 2018 cycles of PISA



In terms of gender differences, girls have significantly outperformed boys in literacy on a consistent basis (Figure 8.3). In mathematics, the contrary has been the case, with boys tending to outperform girls; however, in 2018, this gap had narrowed to the point that it was no longer significant. No gender differences have been observed in science, with the exception of 2015, when boys significantly outperformed girls by 11 points.

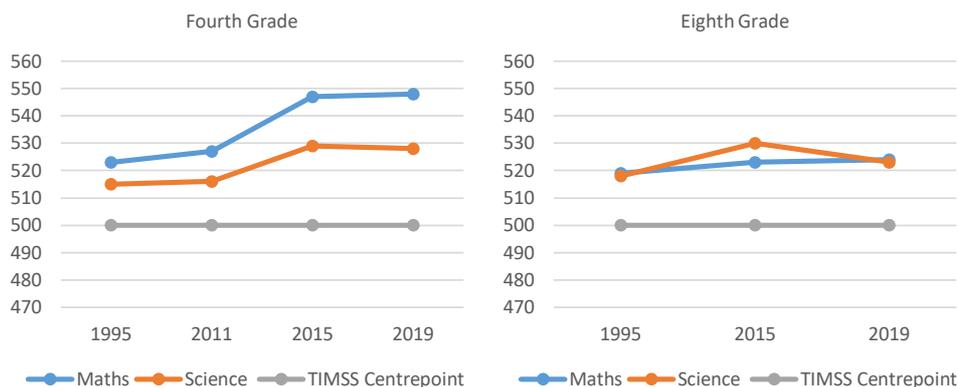
Figure 8.3: Gender differences in Ireland's performance in reading, mathematics and science in PISA, 2000–2018



Trends in International Mathematics and Science Study

Ireland has participated in four of the seven cycles of the Trends in International Mathematics and Science Study (TIMSS), in 1995, 2011 (fourth-grade level only), 2015 and 2019. As with PISA, its implementation is overseen by the ERC. Overall, performance in TIMSS has been strong, with average scores significantly above the centre point observed in both mathematics and science at the relevant grade levels in every instance (Figure 8.4). At fourth-grade level, a significant increase in scores in both domains occurred between 2011 and 2015, and these levels of achievement were maintained in the most recent cycle in 2019. Of the 58 countries participating in TIMSS at fourth-grade level in 2019, Ireland's score of 547 in mathematics was significantly higher than that of 46 countries, whereas in science its score of 528 was significantly higher than that of 33 countries. At eighth-grade level, performance has been stable, aside from a significant improvement in science scores observed between 1995 and 2015. In 2019, eighth-grade students in Ireland obtained a mean score of 524 in mathematics, significantly outperforming those in 26 of 37 other participating countries, and 523 in science, significantly outperforming those in 23 of those countries. No gender differences have been observed with regard to Ireland's performance in either domain at either grade level over the years.

Figure 8.4: Trends in Ireland's performance in mathematics and science in TIMSS at fourth-grade and eighth-grade levels, c1995–2019



Progress in International Reading Literacy Study

Ireland participated in two of the four cycles of the Progress in International Reading Literacy Study (PIRLS), in 2011 and 2016, obtaining mean scores of 552 and 567, respectively. In both instances, these scores were significantly higher than the PIRLS centre point. The 15-point improvement between the two cycles was also statistically significant. Ireland was among the top performers in PIRLS 2016, scoring significantly higher than 43 of the other 49 participating countries.

The 2016 cycle included the introduction of ePIRLS, an additional measure of online reading skills and digital literacy, in which Ireland also participated. No difference was observed between Ireland's overall PIRLS and ePIRLS scores. In terms of gender differences, girls significantly outperformed boys, by 15 points in 2011 and by 12 points in 2016. Girls also outperformed boys in ePIRLS, by 11 points. These differences reflect both the size and the direction of the average gender gap observed across all participating countries.

Adult skills surveys

Ireland participated in the International Adult Literacy Survey (IALS) in 1995 (overseen by the ERC) and in the first cycle of the Programme for the International Assessment of Adult Competencies (PIAAC) in 2012 (overseen by the Central Statistics Office). Almost 6 000 adults aged 16–65 years participated in PIAAC, and the country is also due to participate in the second cycle in 2022. Ireland's performance in these adult skills surveys has been weaker than that in PISA, TIMSS and PIRLS (Table 8.2). A mean literacy score of 266 in PIAAC placed Ireland 17th out of 24 participating countries. This did not differ significantly from the study average of 270, or from Ireland's score of 264 in IALS⁽¹⁵⁾. A drop was

⁽¹⁵⁾ Despite some difference between IALS and PIAAC data, it is possible to compare them in a meaningful way, bearing in mind that parameters for the linking items have been refined since they first appeared in IALS, such that IALS estimates used in comparisons are slightly different from those originally reported in 1995.

observed, however, with respect to the proportion of adults scoring at or below the lowest level of literacy (from 22.1 % in 1995 to 17.9 % in 2012).

PIAAC also included measures of numeracy and problem-solving in technology-rich environments. In numeracy, Ireland performed significantly below average, with a mean score of 255, and one quarter of adults scoring at or below level 1. In problem-solving, for which means are not reported and proficiency is divided into just three levels, 42 % scored at or below level 1, similar to the study average. No gender differences were observed with respect to literacy, but men significantly outperformed women, by 12 points, in numeracy.

Table 8.2: Summary statistics for Ireland's performance and proportions of adults performing at various levels of proficiency in literacy in IALS (1995) and PIAAC (2012) and in numeracy and problem-solving (PIAAC only)

	Mean score	At or below level 1	Level 2	Level 3	At or above level 4
		%	%	%	%
Literacy					
Ireland (IALS)	264	22.1	32.2	34.5	11.2
Ireland (PIAAC)	266	17.9	37.6	36.0	8.5
International average (PIAAC)	270	16.7	33.3	38.2	11.8
Numeracy					
Ireland	255	25.6	38.0	28.8	7.6
International average	266	20.2	33.0	34.4	12.1
Problem-solving					
Ireland (*)	—	42.0	22.1	3.1	—
International average	—	41.7	28.2	5.8	—

(*) Percentages do not sum to 100 %, as data pertaining to this scale were available only for those who opted for a computer-based assessment.

When the PIAAC results for adults who could have been in one of the preceding PISA cohorts are isolated and compared with those of PISA, it seems that a deterioration in literacy and numeracy skills occurs as individuals move into adulthood in Ireland (OECD, 2013). Such comparisons should be interpreted with caution, however, as there are some differences with respect to the cohorts involved, and no common items were used across the two studies. Possible differences in the motivation of test-takers across the two contexts should also be considered. (For comparisons between PISA and PIAAC, see CSO, 2013, p. 90; and OECD, 2016, Chapter 6.)

Education policy monitoring

School-based literacy and mathematics

The term 'PISA shock' is often used to describe the effect of unexpectedly poor PISA results on a country. In December 2010, when the results of PISA 2009 were announced by the OECD, Ireland experienced something akin to a PISA shock. A number of issues were highlighted by the ERC and others at the time that were likely to have contributed to the reported declines in performance. These included an increase in the proportion of non-English speakers in post-primary schools, poor student engagement during the PISA 2009 assessment, as evidenced by a greater-than-expected deterioration in performance as students progressed through their booklets and skipped more items (LaRoche and Cartwright, 2010), and difficulties with the approach taken by the OECD and its contractors to linking performance across PISA cycles, such as too few link items and large link errors (Cartwright, 2011; Cosgrove, 2011; Sachse et al., 2019).

Although the then Minister for Education described the PISA 2009 results in a newspaper interview as a 'black mark for the education system' (Flynn, 2010), in general the response was relatively measured. Reflecting on the 2009 results, the Chief Inspector noted that the results challenged some complacency about standards in Ireland, while at the same time reminding stakeholders of the need to interrogate the results of international comparative tests of this kind and recognise their limitations (Hislop, 2012).

The DES had already begun working on a draft strategy for literacy and numeracy in 2010, and accelerated the development of this in response to the PISA 2009 results. The first draft of the strategy, launched in November 2010, contained no reference to international test data (DES, 2010). The Irish government launched the national literacy and numeracy strategy (DES, 2011) in July 2011. That final version, post 2009, referred to the data from PISA and, for the first time, included improving performance in an international test as a policy goal.

The strategy set out a number of actions that would be taken across the educational system in order to raise standards. Surprisingly, these actions covered early childhood and primary levels, as well as post-primary level, even though there was no evidence of a decline in standards before post-primary level. A greater focus on literacy rather than on numeracy was also evident, mainly because a new mathematics curriculum, Project Maths, which was intended to improve conceptual understanding and real-life problem-solving, had been disseminated, on a phased basis, at post-primary level since September 2010 (with full implementation to be completed by 2015) (Shiel and Kelleher, 2017). Furthermore, independent of the strategy, the bonus CAO points for students taking leaving certificate mathematics (as described previously) were due to be introduced from 2012 onwards. The following were among the targets outlined in the 2011 strategy document.

- Establish specific targets for reading literacy and mathematics, based on national assessments at primary level ⁽¹⁶⁾, and on PISA at post-primary level. The targets for PISA reading literacy and mathematics included increasing the proportion of 15-year-olds performing at or above level 4 by at least 5 percentage points by 2020, and reducing the proportion performing at or below level 1 by half in the same time frame ⁽¹⁷⁾.
- Increase the percentage of students taking the higher-level junior certificate examination (end of grade 9) to 60 % by 2020, and increase the proportion taking the higher-level leaving certificate examination (end of grade 12) to 30 % by 2020.

Several actions were also outlined in the strategy in relation to early childhood education, school leadership, teacher education, curriculum content, learners with additional needs, and assessing and reporting on students' progress. The actions included:

- the intention to introduce standardised tests on an annual basis at primary level (grades 2, 4 and 6) and at post-primary level (grade 8), with schools required to submit aggregated results to school boards of management and to the DES ⁽¹⁸⁾;
- increasing awareness of digital literacy by including tests of digital literacy in national assessments;
- increasing the duration of teacher education for teachers intending to teach at primary and post-primary levels and introducing mandatory courses on literacy and numeracy for prospective teachers;
- increasing instructional time for the teaching of literacy and numeracy in primary schools;
- placing a greater emphasis on learning outcomes in revised curricula in English (and Irish) at primary level, and prioritising the English (and Irish) curricula at post-primary level, with an emphasis on literacy.

There was surprise at the time at the broad range of actions in the strategy, since the country was in economic recession, and cutbacks were being enacted across a broad range of areas in the education system, leading, for example, to increases in class size at primary level (Kelleher and Weir, 2016).

Although the original literacy and numeracy strategy was intended to extend to 2020, a review of the strategy was published by the DES in 2017 (DES, 2017). The review

⁽¹⁶⁾ The targets at primary level did not encompass PIRLS or TIMSS because of the prevalence of the national assessments at the time. Ireland administered PIRLS for the first time in 2011, and administered TIMSS (at grade 4 only) in the same year for the first time since 1995. Ireland also administered TIMSS at grade 8 in 2015, for the first time since 1995. It may be noted that effects of the strategy can be inferred by looking at changes in performance between PIRLS 2011 and PIRLS 2016, between TIMSS 2011 and subsequent TIMSS studies in 2015 and 2019 at grade 4, and between TIMSS 2015 and TIMSS 2019 at grade 8.

⁽¹⁷⁾ It might be noted that these targets were presented in absolute terms; measurement error around the proportions achieving various proficiency levels was not referred to.

⁽¹⁸⁾ Some of the actions outlined in the original strategy document were never implemented, including the administration of standardised tests to students in grade 8 at post-primary level.

examined progress in implementing actions set out in the original strategy, and proposed revised targets for 2020. This was necessary at primary level in particular, because the targets set in 2011 had been achieved in the 2014 national assessments. In addition to extending existing targets for PISA, the review issued new PISA targets for two additional groups – those performing at levels 5–6 in reading literacy and mathematics in PISA, and students in disadvantaged (DEIS) post-primary schools (despite the fact that PISA 2018 would be administered just 12 months after the interim review was published).

There is also evidence of a link between international assessments and curriculum development, which occurs at national level in Ireland. For example, there are clear commonalities between the framework for the 2009 national assessment of reading (ERC, 2008) and the PIRLS framework, with respect to definitions of reading literacy, and explication of key content (text types) and processes. At post-primary level, PISA mathematics had a strong impact on a review of mathematics education prior to the development of Project Maths (Conway and Sloane, 2006), and performance in PISA was used as a basis for evaluating the implementation of Project Maths (Shiel and Kelleher, 2017). In addition, the current specification for junior cycle science (NCCA, 2015) includes PISA's definition of scientific literacy and has added an additional science content area also found in PISA – earth and space science.

In her analysis of the growing influence of PISA in a range of education policy developments, Looney (2016) takes mathematics as a case study and notes that the initial consultation paper on mathematics reform published in 2002 did not refer at all to Ireland's performance in PISA 2000. The report on the consultation was published in 2005; in the interim, PISA 2003 had pointed to some particular weaknesses in Ireland's performance in mathematics (17th out of 29). These weaknesses are explicitly referenced in the proposals for change in upper secondary education, particularly the need to respond to the relatively low performance of high-achieving students. Looney notes that PISA is absent from the initial rationale for the reform of mathematics. As the reform progressed, the influence of PISA became more obvious in the details of the proposals. In their final form, the reform proposals specifically referenced addressing a system weakness identified in the analysis of PISA data. PISA is becoming not just a contextual factor in reforms but a rationale for change.

Notably, the TIMSS and PIRLS data are almost invisible in Irish policy texts. For example, two recent initiatives by the NCCA – a new integrated language curriculum for primary schools and a new framework for the primary curriculum – make no reference to international test data in discussions of the context for the reforms or the measures of success (see <https://ncca.ie/en/primary/primary-developments/>). The influence of the OECD in Irish policymaking more broadly (e.g. economic policy) and in education policy particularly is undoubtedly a contributing factor. The focus of TIMSS and PIRLS on younger children, who are further from the labour market or tertiary education, may be another.

Adult literacy and numeracy

IALS was administered in Ireland in 1997, and results were published by the OECD in 1997. The conclusion that 500 000 Irish adults (22.6 % of those aged 16–64 years) ⁽¹⁹⁾ had literacy difficulties provided an impetus for significant advances in adult education in subsequent years. For example, financing for adult education, including courses for adults with literacy and numeracy difficulties, increased from EUR 1 million to EUR 10 million, and had surpassed EUR 40 million by 2007 (Holden, 2007). The outcomes of IALS also gave impetus to the development of the White Paper on adult education (Government of Ireland, 2000), which included outcomes of the IALS study and set out priorities in adult education for the future. A new national adult literacy, numeracy and digital literacy strategy is currently being developed by SOLAS, the national agency for further education and training, and the consultation paper on the strategy (SOLAS, 2020) contains an analysis of PIAAC outcomes from 2013, including an examination of factors associated with low levels of literacy and numeracy. The consultation paper also draws on information about Irish adults' digital literacy skills from PIAAC. Hence, although the IALS study had a significant impact on policy in terms of expanding access to programmes, the 2012 PIAAC study provided data to support the continuation of existing programmes, albeit with the addition of a stronger focus on digital literacy.

Analysis of policies, practices and political reforms

As outlined in the previous section, the most prominent example of political reform in response to international survey results in Ireland has undoubtedly been the introduction of the national literacy and numeracy strategy (DES, 2011). A series of outcomes deriving from various aspects of this strategy will now be discussed. In terms of evaluating the impact of these reforms, however, it is important to bear in mind that the improvement in Ireland's PISA performance observed between 2009 and 2012 should not be attributed to the strategy, as PISA 2012 occurred before its implementation began in earnest. Indeed, as highlighted previously, questions have been raised surrounding the validity of the apparent 'shock decline' in Ireland's performance in 2009 (see, for example, Cosgrove and Cartwright, 2014; Sachse et al., 2019).

With respect to literacy, Table 8.3 shows that the target of 8.5 % at or below level 1 set in 2011 had not been achieved by PISA 2018 (standard errors are not referred to in the strategy documents). Although the proportion at or below level 1 dropped from 17 % in PISA 2009 to 10 % in both PISA 2012 and PISA 2015 (indicating fewer lower achievers), and 12 % in 2018, this, in fact, represents a return to historical levels (11 % in 2000). Notwithstanding the transition to computer-based assessment, this does not indicate a long-term reduction in the proportion of lower-achieving students in the system. Although a target of 34 % was set for students at or above level 4 in 2011, and this was achieved by PISA 2012 (37 %), the revised target of 40 % had not

⁽¹⁹⁾ This was typically referred to as 25 % or one quarter in media reports after results were released.

been achieved by 2018. In fact, that target was similar to the proportion in Ireland who performed at or above level 4 in PISA 2000 (41 %). The target of 12 % established for at or above level 5 in 2017 was reached by 2018 and is remarkably stable across PISA cycles, with the exception of 2009.

Table 8.3: Performance of students in Ireland at key proficiency levels in PISA reading literacy and mathematics, and initial and revised targets in national literacy and numeracy strategy documents

PISA proficiency levels	PISA 2009	Target for 2020 set in the 2011 strategy	PISA 2012	PISA 2015 (*)	Revised target for 2020 set in 2017 interim report	Position as of PISA 2018	Was revised target achieved by 2020?
	%	%	%	%	%	%	
Reading literacy							
At or below level 1	17	8.5	10	10	8.5	12	Not yet achieved
At or above level 4	29	34	37	37	40	36	Not yet achieved
At or above level 5	7	No target in 2011	11	11	12	12	Achieved
Mathematics							
At or below level 1	21	10.5	17	15	10.5	15	Not yet achieved
At or above level 4	26	31	31	31	36	31	Not yet achieved
At or above level 5	7	No target in 2011	11	9	13	8	Not yet achieved

(*) PISA transitioned to computer-based assessment in most countries, including Ireland, in 2015.

In mathematics, the proportion of those achieving at or below level 1 had been declining (from 21 % in 2009 to 15 % in 2018), but the target of 10.5 % was not achieved by 2020. For those scoring at or above level 4, the initial target of 31 % was achieved well ahead of time (by PISA 2012), but the revised target of 36 % was not achieved. Similarly, only 8 % of students were at levels 5–6 in PISA 2018, below the target of 13 %. This reflects a persistent challenge in Ireland's performance patterns; indeed, for the past 25 years or so, the percentages of students scoring at the highest performance levels in both mathematics and science in Ireland have been significantly lower than those in countries with similar average performance (see Pitsia, 2020, for a detailed exposition of the problem).

Targets at primary level have been set with reference to national assessments in reading and mathematics, which are conducted every 5–6 years. Although the 2014 national assessments showed substantial improvement compared with the previous ones in 2009, and, as noted earlier, there is some evidence of carry-over to PIRLS at grade 4 in 2016 and TIMSS at grade 4 in 2015 and 2019, there has been no obvious transfer to post-primary level and to PISA in particular. Given the substantial gains reported for the 2014 national assessments at sixth grade, one might have expected a similar improvement in PISA 2018, although the interpretation of performance in PISA 2018 is complicated by the transition to computer-based assessment.

In terms of whether greater progress might have been made in relation to the targets, some aspects of the strategy are worth scrutinising. To begin with, no clear rationale has ever been given for the specific targets identified. With respect to literacy, the target of 8.5 % at or below level 1 could be regarded as overly ambitious. Indeed, the lowest proportion of students at or below level 1 in literacy in 2018 was 11 % (recorded by both Estonia and Singapore). Furthermore, a wealth of available evidence at the time indicated that low reading achievement in Ireland was concentrated in DEIS schools (see, for example, Eivers et al., 2010). With this in mind, it could be argued that more progress towards the target of 8.5 % at or below level 1 might have been made had efforts to improve literacy been focused more specifically on the most ‘at-risk’ pupils (e.g. those in DEIS schools), rather than being spread across all pupils.

The relatively greater emphasis on literacy at the expense of numeracy could also be called into question, particularly given that Ireland’s performance in literacy (with the exception of PISA 2009) has always been a relative strength. As noted previously, this decision was influenced by the fact that separate policy initiatives relating to numeracy, including the introduction of the Project Maths curriculum and the availability of bonus points for leaving certificate higher-level maths, were being (or were soon to be) implemented when the strategy was being devised. These initiatives, however, do not seem to have impacted at all on PISA mathematics. Moreover, Ireland’s SEC has drawn attention to a lack of basic skills among some students sitting higher-level mathematics, both in the junior certificate (SEC, 2015a) and in the leaving certificate (SEC, 2015b), and concerns have also been raised about the ‘mathematical underpreparedness’ of students entering higher education since the introduction of the bonus points (STEM Education Review Group, 2016, p. 8). Although this issue seemed to have been taken into account in the interim review, in which it was indicated that numeracy would ‘need to be prioritised over the second half of the lifetime of the Strategy’ (DES, 2017, p. 5), more challenging targets were set for both domains between 2017 and 2020. Questions remain as to whether or not it was necessary to revise literacy targets that had already been met, and, indeed, whether or not it was realistic to expect that the much-needed emphasis on numeracy could be realised under these conditions.

As noted previously, many aspects of the strategy were targeted at primary level, despite the fact that the apparent decline in performance was evident only at post-primary level. For example, additional instructional time was afforded to both literacy and numeracy at primary level, with 29 minutes and 23 minutes extra, on average, added to

English and mathematics classes, respectively, between 2009 and 2014 (Kavanagh et al., 2015). There is a question, however, as to whether this simple increase was the most appropriate action to take, and/or whether greater attention might have been paid to the time spent on particular aspects of mathematics. Content such as shape and space, and cognitive processes such as problem-solving, for example, have been identified as particular areas of weakness among Irish students (Close, 2013).

Compulsory standardised testing in English reading and mathematics was also introduced in primary schools, along with mandated reporting of aggregated data to boards of management and the DES on an annual basis (DES, 2011). In addition, with effect from September 2017, the results were to be used at national level as part of the process involved in determining the allocation of special educational teaching resources to schools. Although the stakes associated with these changes are relatively low, the increased profile that standardised tests now have has caused some concern about the possibility that some of the well-documented negative consequences associated with their higher-stakes use in other jurisdictions (see, for example, Abrams et al., 2003) might be coming into play in Ireland.

In May 2017, a survey involving a representative sample of 1,564 Irish primary teachers was conducted to explore, among other issues, what teachers felt was the impact of standardised testing on their professional practice and on student learning (O'Leary et al., 2019). In terms of student learning, just under two thirds of teachers reported that standardised test scores had either improved or remained constant over the 3 years preceding the survey. On the surface, this could be interpreted as evidence of positive outcomes deriving from the policy changes in the years prior. Indeed, many explicitly cited the introduction of the strategy as one of the key factors influencing test score changes, whereas others attributed improvements to changes in teaching strategies, teacher effectiveness and internal evaluation practices, all of which could also be linked to certain aspects of the strategy.

Some teachers, however, felt that a focus on test-taking skills, familiarity with test content, and changes in test preparation and administration practices had contributed to improvements in test scores. Of concern was that one in four indicated that they were aware of what might be described as low-level 'teaching to the test' by their colleagues, whereas approximately 1 in 10 said they were aware of inappropriate practices occurring in their schools, such as pupils being given more than the allowed time, having questions rephrased for them or having access to potentially helpful materials during testing. In addition, many teachers claimed that they felt under pressure to improve their pupils' standardised test scores. These findings suggest that key principles underpinning the use of standardised tests may be eroding as a result of the increased focus on them in Ireland in recent years – a concerning and unintended outcome deriving from policy changes.

Conclusion

The ongoing funding of Ireland's participation in international studies of educational achievement at primary, post-primary and adult levels by the DES suggests that such studies are viewed as being useful for purposes such as benchmarking the performance of students and adults in Ireland against their counterparts in other countries, informing curriculum development, identifying policy priorities and setting system-wide performance targets. However, questions must be raised about the ongoing usefulness of PISA for these purposes. PISA has now moved to computer-based assessment, whereas in many countries the predominant mode of assessment in high-stakes state examinations continues to be paper and pen. This is also the case in Ireland (with the exception of a small number of courses that incorporate an element of digital assessment, and computer science, which was assessed fully on computers for the first time in 40 schools in 2020). The new item types favoured by PISA, such as scenarios in PISA science in which students must perform and interpret the outcomes of virtual experiments, are quite different from the teaching and assessment activities currently found in post-primary classrooms in Ireland, and it remains to be seen how quickly they will be incorporated into teaching and learning.

Another issue concerning the value of international assessments in general, and PISA in particular, is that changes to the way in which items are selected and scaled may be contributing to the greater stability of scores over time. This is evident in, for example, the fact that 31 % of students in Ireland achieved level 4 or higher in mathematics in three consecutive PISA cycles (2012–2018). Recent changes to PISA that have contributed to greater stability include increases in the numbers of link items carried over from cycle to cycle, the inclusion of country-by-cycle-specific item parameters during scaling (prior to 2015, items showing differential item functioning would have been dropped), and the treatment of not-reached items as not administered (OECD, 2017). The greater stability evident in PISA scores across many participating countries means it is more difficult to detect improvements or declines in performance, especially over short periods of time.

A related issue is the extent to which schools view themselves as having ownership of targets based on international assessments. Since schools do not have access to the tests used in international assessments or to the performance profile of their students on such tests, it might be more useful to base targets on tests that are administered to all students in a school and perhaps make schools more responsible for achieving school-level and national targets. Such an approach might also be expected to raise performance in international assessments. However, there are also potential concerns associated with this approach, given that a key feature of the international assessments at present is that they are research projects without stakes for individuals and schools. Arguably, adding ownership will also add stakes.

Finally, as noted previously, in primary education in Ireland, data from international tests play a far less significant role in monitoring, evaluation and policy development than data from PISA and PIAAC. Less frequent participation in PIRLS and TIMSS alongside

strong performance in these tests focused on younger students may well be factors, but the overemphasis on and attention to PISA is a notable feature of the system. The historical role of the OECD in Irish education may also be relevant. The OECD report on Irish education in 1991, discussed earlier, was seminal in initiating a series of system reforms (Coolahan, 2017), and the organisation continues to play a significant role in policy deliberation across the system, from issues such as funding of higher education to issues such as early childhood education.

REFERENCES

- Abrams, L., Pedulla, J. and Madaus, G. (2003), 'Views from the classroom: teachers' opinions of state-wide testing programs', *Theory into Practice*, Vol. 32, No 1, pp. 18–29.
- Cartwright, F. (2011), *PISA in Ireland, 2000–2009: Factors affecting inferences about changes in student proficiency over time*, Educational Research Centre, Dublin (http://www.erc.ie/documents/pisanational-factors_affecting_inferences.pdf).
- Close, S. (2013), 'Mathematics items: context and curriculum', in Eivers, E. and Clerkin, A. (eds), *National Schools, International Contexts: Beyond the PIRLS and TIMSS test results*, Educational Research Centre, Dublin, pp. 152–175.
- Conway, P. and Sloane, F. (2006), *International Trends in Post-primary Mathematics Education – Perspectives on learning, teaching and assessment*, National Council for Curriculum and Assessment, Dublin.
- Coolahan, J. (2017), *Towards the Era of Lifelong Learning – A history of Irish education 1800–2016*, Institute of Public Administration, Dublin.
- Cosgrove, J. (2011), *Does Student Engagement Explain Performance on PISA? – Comparisons of response patterns on the PISA tests across time*, Educational Research Centre, Dublin (http://www.erc.ie/documents/engagement_and_performance_over_time.pdf).
- Cosgrove, L. and Cartwright, F. (2014), 'Changes in achievement on PISA: the case of Ireland and implications for international assessment practice', *Large-scale Assessments in Education*, Vol. 2, pp. 1–17 (<https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/2196-0739-2-2>).
- CSO (Central Statistics Office) (2013), *PIAAC 2012, Survey Results for Ireland*, Stationery Office, Dublin (<https://www.oecd.org/skills/piaac/Ireland.pdf>).
- DES (Department of Education and Skills) (2010), *Better literacy and numeracy for children and young people – A draft national plan to improve literacy and numeracy in schools*, Department of Education and Skills, Dublin.
- DES (2011), *National Strategy: Literacy and numeracy for learning and life 2011–20*, Department of Education and Skills, Dublin (<https://www.education.ie/en/Schools-Colleges/Information/Literacy-and-Numeracy/Literacy-and-Numeracy-Learning-For-Life.pdf>).
- DES (2017), *National Strategy: Literacy and numeracy for learning and life 2011–2020 – Interim review 2011–2016 – New targets: 2017–2020*, Department of Education and Skills, Dublin (<https://assets.gov.ie/24960/93c455d4440246cf8a701b9e0b0a2d65.pdf>).
- DES (2020a), *Education Indicators for Ireland*, Department of Education and Skills, Dublin (<https://assets.gov.ie/117558/ef43dc12-56c3-4f5a-9e78-bafce7ef9764.pdf>).
- DES (2020b), *Calculated Grades for Leaving Certificate 2020: Guide for schools on providing estimated percentage marks and class rank orderings*, Department of Education and Skills, Dublin (<https://assets.gov.ie/86709/aacaf08c-a32b-4e74-817b-5a3ba1e2a7e8.pdf>).

- Eivers, E., Close, S., Shiel, G., Millar, D., Clerkin, A., Gilleece, L. and Kiniry, J. (2010). *The 2009 national assessments of mathematics and English reading*, Stationery Office, Dublin (https://www.erc.ie/documents/na2009_report.pdf).
- Elley, W. B. (1992), *How in the World Do Students Read? IEA study of reading literacy*, International Association for the Evaluation of Educational Achievement, The Hague (<https://files.eric.ed.gov/fulltext/ED360613.pdf>).
- ERC (Educational Research Centre) (2008), *Framework for the 2009 national assessment of English reading (NAER)*, Educational Research Centre, Dublin (https://www.erc.ie/documents/reading_framework2009.pdf).
- Flynn, S. (2010), 'Results black mark for our education system says Quinn', *The Irish Times*, 8 December (<https://www.irishtimes.com/news/results-black-mark-for-our-system-says-quinn-1.685951>).
- Government of Ireland (2000), *Learning for Life: White Paper on adult education*, Government Publications Office, Dublin (<https://assets.gov.ie/24723/8eaa8222690f43279dd017f686427e9b.pdf>).
- Hislop, H. (2012), 'The quality assurance of Irish schools and the role of evaluation: current and future trends', The Professor Seamas Ó Súilleabháin memorial lecture 2012, National University of Ireland, Maynooth (<https://assets.gov.ie/25310/68cf8ab4f10142b6a6dd3c7c8a9c3f03.pdf>).
- Holden, L. (2007), 'A gradual learning curve: two decades of adult education', *The Irish Times*, 13 November (<https://www.irishtimes.com/news/education/a-gradual-learning-curve-two-decades-of-adult-education-1.981749>).
- Kavanagh, L., Shiel, G., Gilleece, L. and Kiniry, J. (2015), *The 2014 national assessments of English reading and mathematics – Volume II: Context report*, Educational Research Centre, Dublin (http://www.erc.ie/wp-content/uploads/2016/09/NA-Context-Report-Oct_2016.pdf).
- Kelleher, C. and Weir, S. (2016), 'Class size and student-teacher ratio at primary level in Ireland and other OECD countries', *Irish Journal of Education*, Vol. 41, No 1, pp. 39–60.
- Lapointe, A. E., Mead, N. A. and Phillips, G. W. (1989), *A World of Differences: An international assessment of mathematics and science*, Educational Testing Service, Princeton, NJ.
- Lapointe, A. E., Mead, N. A. and Askew, J. M (1992a), *Learning Mathematics*, Educational Testing Service, Princeton, NJ.
- Lapointe, A. E., Askew, J. M. and Mead, N. A. (1992b), *Learning Science*, Educational Testing Service, Princeton, NJ.
- LaRoche, S. and Cartwright, F. (2010), *Independent review of the PISA 2009 results for Ireland: Report prepared for the ERC at the request of the Department of Education and Skills*, Department of Education and Skills, Dublin (http://www.erc.ie/documents/statscan_pisa00to09_final_report.pdf).
- Looney, A. (2016), 'Ireland: the intersection of international achievement tests and education policy development', in Volante, L. (ed.), *The intersection of international achievement tests and education policy*, Routledge, London and New York, pp. 75–90.

- NCCA (National Council for Curriculum and Assessment) (2015), *Junior Certificate Science – Curriculum specification*, National Council for Curriculum and Assessment, Dublin.
- O’Leary, M., Lysaght, Z., Nic Criath, D. and Scully, D. (2019), *Standardised testing in English reading and mathematics in Irish primary schools: A survey of Irish primary teachers*, Centre for Assessment Research, Policy and Practice in Education, Dublin City University, and Irish National Teachers’ Organisation, Dublin (<https://www.dcu.ie/sites/default/files/carpe/carpe-into-standardised-testing-survey-2019.pdf>).
- OECD (Organisation for Economic Cooperation and Development) (1991), *Review of Irish Education*, OECD Publishing, Paris.
- OECD (2013), *OECD Skills Outlook 2013: First results from the survey of adult skills*, OECD Publishing, Paris ([https://www.oecd.org/skills/piaac/Skills volume 1 \(eng\)-full v12-eBook \(04 11 2013\).pdf](https://www.oecd.org/skills/piaac/Skills_volume_1_(eng)-full_v12-eBook_(04_11_2013).pdf)).
- OECD (2016), *The Survey of Adult Skills: Readers’ companion, second edition*, OECD Publishing, Paris ([https://www.oecd.org/skills/piaac/The_Survey %20of_Adult_Skills_Reader %27s_companion_Second_Edition.pdf](https://www.oecd.org/skills/piaac/The_Survey_%20of_Adult_Skills_Reader_%27s_companion_Second_Edition.pdf)).
- OECD (2017), *PISA 2015 Technical Report*, OECD Publishing, Paris (https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf).
- OECD (2021), *Education in Ireland – An OECD assessment of the senior cycle review*, OECD Publishing, Paris.
- Passow, A. H., Noah, H. F., Eckstein, M. A. and Aallea, J. R. (1976), *The National Case Study: An empirical comparative study of twenty one educational systems*, Almqvist and Wiksell, Stockholm.
- Pitsia, V. (2020), ‘Investigating high achievement in mathematics and science in Ireland: an in-depth analysis of national and international assessment data’, unpublished dissertation, Dublin City University, Dublin.
- Sachse, K. A., Mahler, N. and Pohl, S. (2019), ‘When nonresponse mechanisms change: effects on trends and group comparisons in international large-scale assessments’, *Educational and Psychological Measurement*, Vol. 79, No 4, pp. 699–726.
- SEC (State Examinations Commission) (2015a), *Junior Certificate Examination 2015 – Mathematics – Chief examiner’s report*, SEC, Athlone, Ireland (<https://www.examinations.ie/misc-doc/EN-EN-25073660.pdf>).
- SEC (2015b), *Leaving Certificate Examination – 2015 – Chief examiner’s report*, SEC, Athlone, Ireland (<https://www.examinations.ie/misc-doc/EN-EN-53913274.pdf>).
- Shiel, G. and Kelleher, C. (2017), *An evaluation of the impact of Project Maths on the performance of students in junior cycle mathematics*, Educational Research Centre, Dublin (https://www.erc.ie/wp-content/uploads/2018/07/PM_EvaluationStrand1.pdf).
- SOLAS (2020), *Adult Literacy, Numeracy and Digital Literacy Strategy: Consultation paper*, SOLAS, Dublin (<https://www.solas.ie/f/70398/x/61f272cf8c/consultation-paper-alnd-strategy.pdf>).

STEM Education Review Group (2016), *STEM Education in the Irish School System – A report on science, technology, engineering and mathematics education*, Department of Education and Skills, Dublin.

Torney, J. V., Farnen, R. F. and Oppenheim, A. N. (1975), *Civics Education in Ten Countries*, Wiley, New York.

Travers, K. and Westbury, I. (eds) (1989), *Second International Mathematics Study*, University of Illinois, Urbana, IL (<https://files.eric.ed.gov/fulltext/ED325360.pdf>).

CHAPTER 9

Cross-national achievement surveys and educational monitoring in Poland

Maciej Jakubowski and Tomasz Gajderowicz

Abstract

This chapter shows how the educational system in Poland went through a long and demanding process of reforms and structural changes in various directions. The first section describes the most impactful changes over the past 20 years. The second section describes Poland's participation in international assessments and provides an overview of the spectacular achievement progress visible in the Organisation for Economic Co-operation and Development's Programme for International Student Assessment. The third section examines the system structure and monitoring schemas, and discusses the extent to which evidence informs political decisions. The final section provides an assessment of the reforms conducted. It identifies recent critical challenges in the Polish educational system and discusses how they should be addressed using international evidence.

Introduction

The past 25 years have been a time of many changes and transformations in the Polish education system. Over these years, education has become a subject of lively public debate. The educational system in Poland, in the second wave of reforms after the post-communist transition, went through a long and demanding process of two fundamental structural reforms and several changes. As a result, Poland is one of the few countries worldwide that has significantly improved educational outcomes in primary and secondary education. This chapter presents the institutional changes, reforms and measures introduced in Poland, and how they correspond to student achievement observed in the international assessment data. In addition, this chapter aims to show how these educational reforms were informed by international data and the extent to which recently introduced policy measures that have reversed reforms from the 1990s are backed by evidence.

The turn of the 21st century is considered the beginning of the most substantial process of decentralisation of education in Poland. This process started with transferring responsibility for schools from the central government to local governments. This process forced the engagement of the local authorities and placed a greater level of autonomy on local governments, school principals and, finally, teachers. Decentralisation resulted in the responsibility for the realisation of subsequent education reforms planned by the central government falling on local authorities; in fulfilling this responsibility, they faced both organisational and financial challenges (Herbst et al., 2012).

Perhaps the most impactful reform occurred in 1999 – a restructuring of the organisation of the entire school system. A three-stage system was introduced in place of the two-stage system (before the reform, 8 years of the primary school plus 4 years of secondary school; after the reform, 6 years of primary school plus 3 years of lower secondary school (*gimnazjum*) and 3 years of upper secondary school). At the same time, the curriculum was fundamentally reshaped (Act of 25 July 1998 amending the Act on the Education System (*Journal of Laws*, item 759)). The 1999 reform changed the governance and financing of the education system to be more incentive compatible. With the ownership of schools transferred from the central government to local governments, a new spending formula to distribute resources was introduced. This formula assumed that funding should follow students. Thus, the system was shaped, to a greater extent, by demand-driven forces. This mechanism set incentives for local governments for the rationalisation of the size and structure of school networks and the promotion of high-quality educational services. Decentralisation led to increased organisational efficiency, as the local governments were able to manage schools in a more effective way than the ministry at national level. Paradoxically, despite local financing, most of the spending rules (e.g. teachers' salaries) were still regulated centrally.

The structural changes in 1999 met with strong resistance. Most education stakeholders, including teachers, underlined the shortcomings of this reform. Local governments were overloaded with infrastructural investments and organisational adjustments. Perhaps

the greatest mistake of this reform implementation was that key stakeholders were not informed about the changes until very late in the process. The consultation mechanism that could have convinced teachers and parents of the value of the reforms were, in reality, ignored. ‘The conflict of teachers with the opposition, although of low intensity, contributed to media criticism of the reform, giving rise to the “black legend of middle schools” as places full of threats and violence’ (Szelewa et al., 2018).

The longer compulsory education introduced by the 1999 reform was solidly backed by science (Kirdar et al., 2016). It resulted in the improvement of the performance of Polish students in the Programme for International Student Assessment (PISA) assessment (Jakubowski et al. 2010). Nevertheless, the strong evidence behind the measures introduced was not adequately communicated to society. This mistake resulted in increasing sentiment towards the old system and finally caused a cascade of consequences for further reforms. Thus, the desirability of a return to the old system became a populist postulate of the ruling party almost 20 years later.

Another critical reform, which lowered the school age, was introduced in 2008. This project was included first in the Law and Justice (PiS) party programme of 2005 (Law and Justice, 2005, p. 99) and then in the agenda of Civic Platform (Civic Platform, 2007, p. 53). The reform was finally announced in the 2008/2009 school year (Act of 19 March 2009, amending the Act on the Education System and some other acts (*Journal of Laws*, item 458)). The general motivation behind this reform was to minimise inequalities by expanding general education and strengthening the quality of pre-primary education. Again, the reform became a political focal point, regardless of the strong evidence supporting the changes. Lowering the school age provoked emotional reactions on the political right, where the changes were described as ‘stealing childhood’. Finally, the reform, as the subject of a political struggle, was withdrawn after the PiS party came to power. Prior to the introduction of the most recent education reforms (initiated in the 2017/2018 school year), the starting age for compulsory education was increased back to 7 years. The reversion of the school-age reform was just the beginning of a set of programmes designed to overturn previously introduced reforms.

Regardless of the strong improvement in the educational outcomes that are visible in international assessment data (to be presented below), the most recent reform moved the system structure back to 1999. This reform was established by the Act of 14 December 2016 – Law on School Education (*Journal of Laws*, item 59), which changed the structure of schools; the *gimnazjum* was liquidated. This reform was extraordinarily costly and demanding for local governments, principals and teachers. This time, however, the government invested substantial resources in carrying out and promoting the reform. After the announcement in February 2016 of the plan to introduce changes, a series of debates and consultations were initiated. According to the data of the Ministry of National Education, during public pre-consultation ‘teams of experts analyzed nearly 800 opinions submitted by social partners, non-governmental organizations, institutions, representatives of the educational environment and citizens’ (from 2 January 2017 to 14 March 2017, 815 meetings were held with over 35 000 participants; Ministry of National Education, 2017). Regardless of their scope, the consultations may have been a façade, as doubts were not carefully considered.

Szelewa et al. (2018) also noted the quick pace of introducing the reform, which raised objections. The fundamental reform entirely changed the educational system and was designed, consulted on and implemented over just a few months.

The changes made to the education system are best reflected in the following policy documents:

- Act of 8 January 1999 – Regulations introducing the reform of the school system (*Journal of Laws*, item 96);
- Regulation of the Minister of National Education of 20 February 2013, changing the regulation on the method of distribution of subsidies from the state budget for public and non-public universities (*Journal of Laws*, item 273);
- Act of 22 December 2015 on the Integrated Qualifications System (*Journal of Laws*, item 64);
- Act of 14 December 2016 – Law on School Education (*Journal of Laws*, item 59);
- Regulation of the Minister of National Education of 28 March 2017, on the framework teaching plans for public schools (*Journal of Laws*, item 703);
- Act of 20 July 2018 – Law on Higher School Education (*Journal of Laws*, item 1 668);
- Act of 27 October 2017 on financing educational tasks (*Journal of Laws*, item 2 203, as amended).

To summarise, Figure 9.1 illustrates the system changes 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 reforms reversed.

Figure 9.1: Changes in the provision of preschool education and compulsory education with the general curriculum (in green)

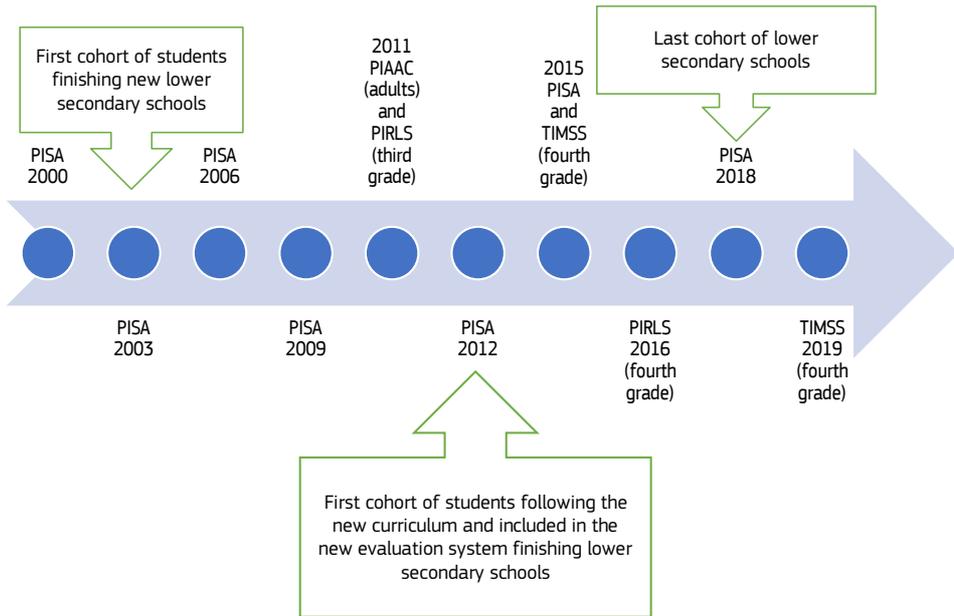
Age	3–4	5	6	7	8	9	10	11	12	13	14	15
Before 1999	Voluntary preschool with no government guarantees		Zero class	Primary								
After 1999			Zero class	Primary				Lower secondary				
2008–2015	Place guaranteed	Compulsory	Primary						Lower secondary	1 year of common curriculum		
Since 2017	Place guaranteed		Zero class	Primary school								

NB: Dark-green shading denotes compulsory education. Light-green shading denotes recommended but non-compulsory education with guaranteed school place.

The most fundamental change resulting from the 1999 reform was that it extended the period of compulsory education initially by 1 year and later by a further year with the introduction of the compulsory 'zero class' for 6-year-olds. Thus, from 2008, two additional years of compulsory education with the common curriculum were added. This decision was backed by international comparisons showing that Polish children began school relatively late and that too few children, especially in rural areas, participated in preschool education compared with top-performing countries. International assessments clearly show that children who attend preschool education perform better in their later educational careers (Barnett, 2008). Starting from 2015, the government introduced new regulations that extended access to preschool education for 3- and 4-year-olds. Unfortunately, since the PiS government came to power in 2017, lower secondary schools have been abolished, the upper secondary school curriculum has changed and compulsory preschool for 5-year-olds has also been abolished (see Figure 9.1). Access to preschool education for children has remained.

International achievement surveys

The disruptive and frequent reforms of the educational system in Poland were introduced at the time of Poland's first participation in most international assessments. Thus the Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS) and PISA data can be used to provide insights into the potential impacts on the learning outcomes for student cohorts affected by different education reforms. Here, some important caveats should be noted. First, international assessments show only the association between student performance and reforms, not the cause. Second, due to the complexity of the reforms, it is difficult to isolate particular policies and associate them with student achievement. Finally, in a rapidly developing economy such as the Polish economy, there are numerous factors that could also influence human capital formation and thus student achievement. Nevertheless, while bearing in mind these challenges, international assessments remain the only objective data that allow for international comparisons of the performance of the educational systems worldwide. International assessments provide opportunities to gain policy-relevant lessons and learn from other's failures and successes. Figure 9.2 presents Poland's participation in international assessments with notes on the waves of students affected by subsequent reforms.

Figure 9.2: Key international assessments and major reforms of the Polish school system

NB: PIAAC, Programme for the International Assessment of Adult Competencies.

An overview of the results from large-scale international assessments shows substantial improvements in Polish students' outcomes over the 20 years of Poland's participation. It should be noted that Poland was just one of the few countries that experienced significant progress in educational outcomes over this period. Poland's results combined with the evidence from other countries provide evidence that the 1999 reform had a substantial positive impact on student achievement and human capital formation.

Polish reforms in the light of evidence from international assessments

The reform in 1999 was designed because of common dissatisfaction in society with the post-communist system (Rura and Klichowski, 2011). Another force behind the reform was the desire to change the post-communist organisation of schooling. However, the communication and public relations around the reform were ineffective. The education reform was implemented as just one of four fundamental reforms in the county. The process of implementation left many people with the impression that the changes were in the wrong direction and chaotic. The sentiment towards the old system grew over the years and resulted in popular support for the populist PiS party's reversing of the reforms in 2016. This happened regardless of the successive improvements in performance in international assessments which were not widely known in society at the time.

Although the 1999 reform could not have been motivated by international assessment data on Poland (as shown in Figure 9.2, Poland participated in PISA for the first time in 2000), the performance of the pre-reform educational system was reflected in the first round of PISA. In general, Poland performed far below the Organisation for Economic Co-operation and Development (OECD) average. These results motivated the next large education reform, conducted between 2008 and 2009. The participation of Poland in PISA 2000, 2003 and 2006 raised discussions about modernising curricula and teaching methods. The PISA assessment framework became the benchmark driving policymakers to change the curriculum to be more skills oriented.

The modernised curricula had a fundamentally different construction in relation to the previous approach, changing the focus from descriptions of topics and content that teachers should cover to learning outcomes. This was linked to increased teacher autonomy over what the learning process should be and how it should progress. After the reform, it was the outcomes (not the process) that were most important. The new curriculum introduced an interdisciplinary way of thinking, and cross-subject topics were included. The principles emphasised the application of knowledge and allowed for teacher autonomy as to how content would be arranged over the school year. These changes also left more room for teachers to develop more individualised programmes and ways of teaching, and a greater variety of teaching materials.

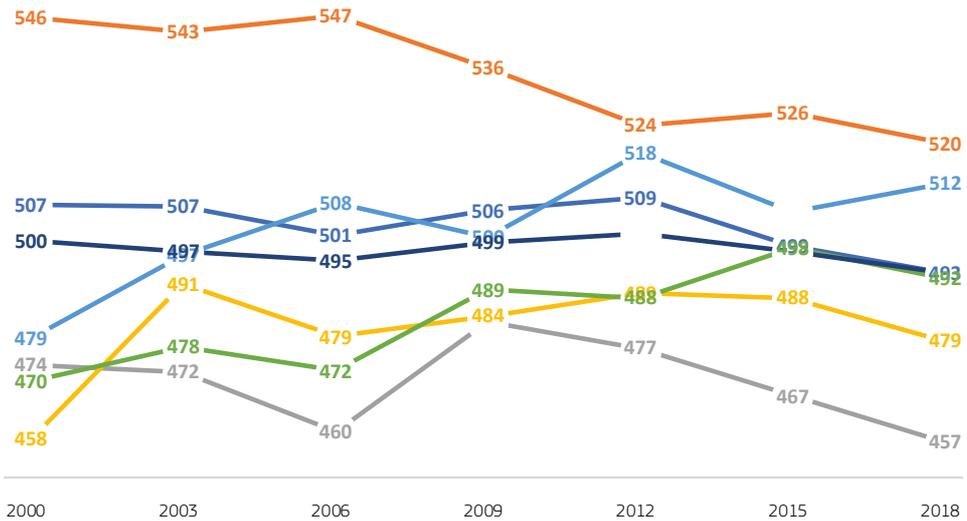
Below, an overview of the results of Polish students in international assessments is presented. Figure 9.3 presents the outcomes of Polish students in all three subjects covered by PISA: reading, mathematics and science. The figure compares average performance in Poland for each subject with the current overall OECD average, which is around 488 points (487 points in reading, 489 points in mathematics and 489 points in science). As the figure presents PISA achievement trends, it can be seen that Polish 15-year-olds improved their skills in all domains during the 20-year period. It is worth noting that, between 2000 and 2009, the impact of the 1999 reform can be observed. Figure 9.4 presents performance changes in the countries with the most significant upward or downward shifts over the past 20 years of international assessments. Poland experienced one of the most extensive improvements in the average scores in reading among participating countries. Moreover, Poland was the only country that improved from far below to above the OECD average. Even among the top-improving countries, the progress of Polish 15-year-olds looks impressive.

Figure 9.3: Achievement trends in reading, mathematics and science in Poland between 2000 and 2018



Source: OECD (2019b).

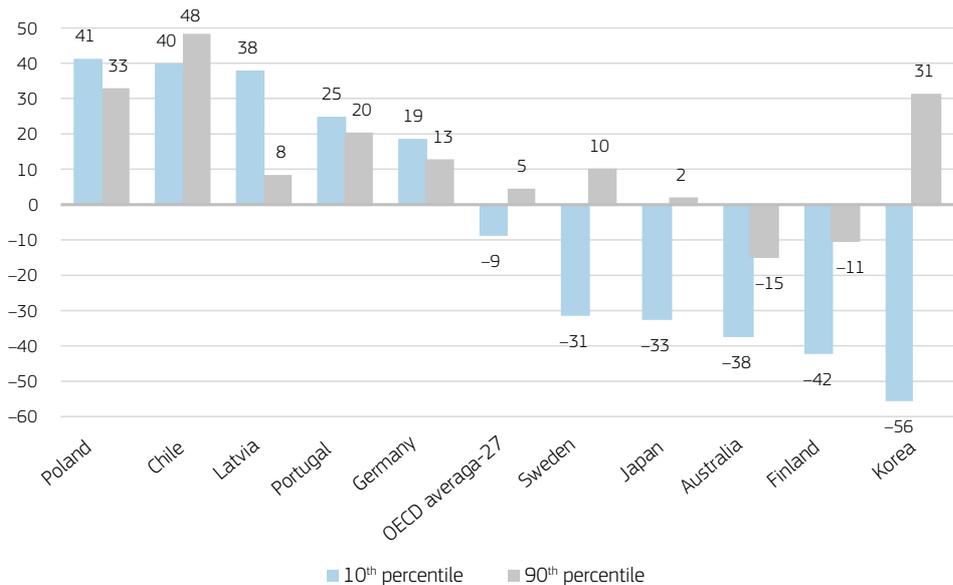
Figure 9.4: Achievement trends in reading for countries with the greatest changes in educational outcomes between 2000 and 2018



Source: OECD (2019a, 2019b, 2019c).

Along with its overall improvement in PISA, Poland also witnessed a reduction in education inequalities across the system. As an example, Poland experienced a massive decline in differences in reading performance between schools. In 2000, Poland was among the countries with the largest performance gaps across schools, whereas in the PISA assessment in 2009, Poland had one of the smallest within-country differences between schools. This happened despite the fact that the overall socioeconomic disparities in Poland remained above the European average. Figure 9.5 presents the change in the achievement of the 10th and 90th percentiles of students across selected countries between PISA 2000 and PISA 2018. These results suggest the progress on educational outcomes in Poland came from the improved achievement of both the highest-performing students and the lowest-performing students.

Figure 9.5: Change in the achievement of the 10th and 90th percentiles of students across selected countries between PISA 2000 and PISA 2018



Source: OECD (2019b).

Linking the reforms to educational achievement is always methodologically difficult, as, with so many changes happening in the economy and society, causality is not obvious. More in-depth causal inference studies (Jakubowski et al., 2010, 2016), based on econometric modelling, confirm that the positive changes in the educational outcomes of Polish students in large-scale international assessments could be associated with the 1999 reform, which extended compulsory education by 1 year. After the reform was introduced, former basic vocational 15-year-old students were still in a comprehensive programme of lower secondary schools. Data from PISA 2000 showed that, within cohorts from before the reform, most of the students in vocational schools were far below the mean for the country. Since the reform was introduced, the share of those who perform below the basic PISA proficiency level has become one of the smallest across participating countries.

There is further evidence beyond these achievement data that shows how unique and successful the Polish reforms at the beginning of the century were. When assessing the educational system's performance in the light of cross-country data, it is worth combining educational outcomes with data on expenditure on education. From this analysis, the Polish system is not only high performing but also cost effective. Estonia and Poland are the only top PISA countries that spend a smaller amount of money on education per student than the OECD average (in US dollars; see graph I.4.4. in OECD, 2019a, p. 66).

Education policy monitoring

In this section, descriptions of the monitoring processes that came with system reforms in Poland are provided. In particular, incentive compatibility mechanisms were established in the form of national examinations and communication of their results and a school evaluation system. Unfortunately, international cross-country data were not properly disseminated to serve as an important support in the data-driven decision-making processes. To summarise the extent to which the educational system in Poland is driven by international assessment, it should be noted that most educational reforms (excluding the latest one) were in fact inspired by international evidence and motivated by data. However, the international assessment outcomes analyses were limited to a group of experts and researchers closely connected to the Ministry of National Education at the time of reform implementation (Białecki et al., 2017).

The communication gap between educational researchers and the public became one of the most serious challenges in the system. Educational researchers often do not focus their attention on policy debates. This created a situation in Poland in which education stakeholders did not have a sound understanding of the educational reforms or the rationale for these reforms. This lack of understanding led to the educational system being in the middle of political battles for many years. Opinion polls in Poland showed that *gimnazja* were not very popular among older people or among those with lower levels of education. Among the younger cohorts who actually finished their education in these schools, opinion was more balanced, with a majority wanting to keep them. Urban legends about the low quality of education and behavioural problems spread across society. The lower secondary schools were blamed for youth behaviour problems, and it became a common belief that putting teenagers in separate schools created negative consequences, in spite of a lack of evidence that this was occurring. Contrary to these beliefs, international surveys indicated that behaviour issues (e.g. bullying) were not more common in these schools than before the reform implementation or in other educational systems (Jakubowski, 2021).

Apart from changes in the structure of the Polish educational system, the 1999 reform introduced other performance monitoring and evidence-informed mechanisms through the implementation of external national examinations. The exams monitored achievement at the end of every level of three-stage compulsory education. The first national exam was conducted at the end of sixth grade (primary school), the second was conducted after lower secondary (*gimnazja*) – ninth grade – and the final exam (*matura*) was conducted after 3 years of upper secondary school (however, this last exam was not part of the

general programme, as vocational education students did not take the exam). All of the exams mentioned were standardised and evaluated centrally to ensure comparability.

There were many benefits of the examination system. First, control and monitoring of performance and progress were possible at national level, and evidence regarding the impact of policy measures could be collected. Second, the examinations motivated students and teachers to prepare for the next round of education. The results were quite important for lower and upper secondary students, because they could enable students to access better successive secondary or tertiary educational programmes. Finally, the outcomes at school level were visible to the general public. Thus, Polish society was able to assess the quality of education in particular schools. Thanks to the three-stage examination system, it was possible to control for selection bias for lower- and upper-secondary schools – the net educational value added by each school was estimated. This measure of educational progress compares school-end exam results with the standardised results achieved by the students when entering school.

The visibility of results allowed for school and teacher self-evaluation and created a strong incentive to focus on preparation for the examination. The effectiveness of this system relied on social, local and political pressures to achieve good outcomes at school level. These pressures worked both between and within schools, creating a competitive environment. However, this competitiveness had many opponents among parents and educational experts in the system, who argued that children and the school environment were being placed in a 'rat race', which negatively affected children's and educators' psychological well-being.

The second wave of reforms in 2008 and 2009 supplemented the system with qualitative rather than quantitative self-evaluation mechanisms of evidence collection and use. This system put more focus on formative evaluation of soft skills, well-being and composition of the school environment. The new formative school evaluation system used examples from top-performing countries to build a national evaluation framework. The system was focused on the organisation of education, highlighting the diversity of pathways to success defined beyond educational outcomes. The reform had a positive impact on the organisation of evaluated schools, introducing modern principles such as teamwork and self-evaluation. Moreover, it positively affected the attitudes of important players in the education system in regard to the relevance of a data-driven way of thinking. The evaluation system had built a structural background for creating a culture of self-evaluation that had not been present in the system before (Mazurkiewicz et al., 2014).

Analysis of policies, practices and political reforms

To summarise, the reforms of 1999 and 2008/2009 were highly driven by evidence from several perspectives (both international assessments and monitoring and evaluation strategies). As a result, the educational system in Poland became more incentive compatible – all stakeholders' goals were consistent with an overall improvement in the quality of education. This effective environment resulted in significant progress

and visible achievement growth in international assessment data. The most important positive changes were:

- expanding compulsory education at both ends;
- introducing external assessments of learning outcomes at three stages, with publicly visible results at school level to provide a measure of value added;
- building a formative school evaluation framework;
- ensuring a large degree of autonomy for Polish teachers in terms of teaching and realising the curriculum.

All of the above seemed to have provided the appropriate balance between autonomy and external monitoring. However, evidence from international survey results has not been used enough to shape political reforms. The analyses of international cross-country data were limited to one group of experts. Although PISA and TIMSS results were covered in the media, their impact on policies was weak. Political parties in Poland do not have a history of using external or international measures to guide policy decisions. Rather, there is a tendency for ‘strong voices’ that have influence with voters to have a greater political influence on decision-making in the country.

Looking back at Figure 9.1, it is clear that one of the most important policies that affected educational outcomes was to consistently follow the path of expanding general education (up to the time of reform reversion). As mentioned, before 1999, students had only eight compulsory years (primary school) of the same general programme. Those who started education in the 2015 cohort have had guaranteed access to preschool education, and from the age of 5 until the age of 16 they have followed the general curriculum. Altogether, 11 years of common education (or even 13 years when including preschool) is obligatory for all students – 3 (or 6) years more than before 1999. This policy fundamentally improved the educational outcomes of students. Moreover, as presented in Figure 9.2, educational outcomes kept improving over the years after the reform implementation.

To summarise, the educational system in Poland has experienced several contradictory reforms, and many of them were driven mainly by politics. Thus, in recent years the Polish political reality related to education has been characterised by instability, uncertainty, populist movements and a multitude of events in various directions, depending on the driving force or party in power. Extreme impermanence of transformations became a reality in education (Zalewska-Bujak, 2010). In all the reforms (especially in the case of the latest one), communication and an evidence-oriented way of thinking were lacking. The reforms left the educational environment with a lack of mutual trust and divided by political views. Szelewa et al. (2018) diagnosed accurately that, in recent decades, the Polish education sector has been characterised by a patronising Enlightenment style of implementing reforms, reluctance, a lack of will to conduct professional and seriously treated social consultations, a tendency to manipulate opinion and a lack of monitoring of consequences of implemented solutions. As a result, Poland faces a lack of willingness from the main political actors to build broad cross-party and cross-class coalitions around evidence-based social policy reforms.

Regarding the dynamic period of reforms, particular attention should be paid to how the education reforms (the introduction of lower secondary schools with all other system changes, up to the re-reform of resetting the post-communist structure of the system) had an impact on teachers, in terms of both finances and their motivation. The reforms negatively influenced teachers' morale and working conditions. Each of the changes resulted in a need to adapt to new standards and adjust to an uncertain environment. Studies analysing the education reforms highlight the superficial nature of consultations and dialogue with teachers (see, for example, ETUCE, 2021).

At the beginning of the century, Poland had an educational system that prospered, and the quality of the reforms introduced was visible in international assessments. According to the PISA results, Poland was the only country from the group of countries with results below the OECD average to become a world educational leader and strengthen its position among European countries. Along with the results of PISA showing the high quality of education in Poland, the results of the TIMSS 2015 assessment provide similar evidence, indicating that fourth graders achieved above the international average. Admittedly, the TIMSS 2019 results show a decline, although it is too early to attribute this to the most recent education reforms. In addition, in PIRLS 2016, Polish fourth graders were ranked at the top of Europe. Finally, in the Programme for the International Assessment of Adult Competencies assessment, which examined Polish teenagers and adults, only the group aged 16–19 years had results higher than the OECD average. Combined, these data highlight the high quality of the educational system in Poland.

In the light of the evidence, the most recent reform, which reversed the entire process, cannot be understood ⁽²⁰⁾. Both abolishing lower secondary schools and reducing the period of general education seem to have been serious mistakes. In the next round of PISA (which was planned for 2021 but has been postponed to 2022 due to the COVID-19 pandemic), it will be possible to investigate to what extent these changes have affected student outcomes. Based on international evidence, one might expect inequalities to be greater. This especially concerns those students who, instead of continuing their general education in secondary schools, decide to continue their education in vocational schools at the age of 15.

Finally, there are other issues of concern that have been revealed by international assessment data and have not yet been addressed by policy measures in Poland. As can be seen in both the PISA and the TIMSS assessment results, Polish students, relative to students in other countries, feel less connected with their schools. Polish teenagers are more likely to think that intelligence is something innate that cannot be changed than teenagers from other countries. These are areas of concern, and policymakers should carefully analyse these results in order to find solutions that can help to maintain the success of Polish education and at the same time build more positive school relationships and appropriate attitudes about education. Currently, the only source for such critical information is the international cross-country assessments.

⁽²⁰⁾ At the moment, there are no data to show the consequences of this structural change, as PISA 2018 captured the last cohort under the old system, whereas TIMSS in 2019 was implemented only with fourth graders in Poland.

REFERENCES

- Barnett, W. S. (2008), 'Preschool education and its lasting effects: research and policy implications', National Institute for Early Education Research, Rutgers, The State University of New Jersey, New Brunswick, NJ.
- Białecki, I., Jakubowski, M. and Wiśniewski, J. (2017), 'Education policy in Poland: the impact of PISA (and other international studies)', *European Journal of Education*, Vol. 52, No 2, pp. 167–174.
- Civic Platform (2007), *Program PO, Polska zasługuje na cud gospodarczy*, Platforma Obywatelska, Warsaw (<https://mamprawowiedziec.pl/file/14512>).
- ETUCE (European Trade Union Committee for Education) (2021), *Rhetoric or Game Changer: Social dialogue and industrial relations in education midst EU governance and privatisation in Europe*, ETUCE, Brussels.
- Herbst, M., Herczyński, J. and Levitas, A. (2012), *Finansowanie oświaty*, Wydawnictwo ICM, University of Warsaw, Warsaw.
- Jakubowski, M. (2021), 'Poland: Polish education reforms and evidence from international assessments', in, Crato, N. (ed.), *Improving a Country's Education*, Springer, Cham, Switzerland, pp. 137–158.
- Jakubowski, M., Porta, E. E., Wisniewski, J. and Patrinos, H. A. (2010), 'The impact of the 1999 education reform in Poland', *Policy Research Working Paper Series*, No WPS 5263, World Bank, Washington DC.
- Jakubowski, M., Patrinos, H. A., Porta, E. E. and Wiśniewski, J. (2016), 'The effects of delaying tracking in secondary school: evidence from the 1999 education reform in Poland', *Education Economics*, Vol. 24, No 6, pp. 557–572.
- Kirdar, M. G., Dayoğlu, M. and Koc, I. (2016), 'Does longer compulsory education equalize schooling by gender and rural/urban residence?' , *The World Bank Economic Review*, Vol. 30, No 3, pp. 549–579.
- Law and Justice (2005), *Program of the Law and Justice 2005*, Prawo i Sprawiedliwość, Warsaw.
- Mazurkiewicz, G., Walczak, B. and Jewdokimow, M. (2014), 'Implementation of a new school supervision system in Poland', *OECD Education Working Papers*, No 111, OECD Publishing, Paris.
- Ministry of National Education (2017), *Informacja Ministra Edukacji Narodowej dla Komisji Edukacji, Nauki i Młodzieży oraz Komisji Samorządu Terytorialnego i Polityki Regionalnej Sejmu Rzeczypospolitej Polskiej Przygotowanie i wdrażanie reformy edukacji na poszczególnych poziomach: Ministerstwo, samorzady, kuratoria oświaty – Informacja MEN dla komisji edukacji*, Ministry of National Education, Warsaw.
- OECD (Organisation for Economic Co-operation and Development) (2019a), *Education at a Glance: OECD and G20 indicators*, OECD Publishing, Paris. OECD (2019b), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-i_5f07c754-en).

- Rura, G. and Klichowski, M. (2011), *Założenia programowo-organizacyjne reformy oświaty z 1999 roku w zakresie edukacji elementarnej*, Wydawnictwo Naukowe UAM, Poznan, Poland.
- Szelewa, D., Polakowski, M., Sadura, P. and Obidniak D. (2018), *Polski system edukacji – czego możemy się nauczyć*, Friedrich-Ebert-Stiftung, Warsaw.
- Zalewska-Bujak, M. (2010), *Udział nauczycieli w przemianach edukacyjnych przełomu XX i XXI stulecia w Polsce*, University of Silesia Press, Katowice, Poland.

CHAPTER 10

Cross-national achievement surveys and educational monitoring in Estonia

Gunda Tire and Einar Värä

Abstract

In recent years, the Estonian education system has received considerable international attention due to the country's high rankings in the PISA assessment. The system is characterised as high performing, with the lowest proportions of low-performing students among the Organisation for Economic Co-operation and Development countries and European Union Member States. The Estonian education system has a high degree of equity, as students from disadvantaged families have access to high-quality education and achieve strong results. The state has set guidelines for national standards and system monitoring; at the same time, local governments and schools enjoy high levels of autonomy. Current education policies were adopted after the re-establishment of Estonian independence in the 1990s. They have been regularly updated, with the prime objective being to provide and guarantee the best education possible for every child. The results from international assessments are carefully considered and analysed when setting educational policy priorities. This chapter takes a closer look at the Estonian education system, its outcomes in international assessments and the policies behind it.

Introduction

The Estonian education system is considered to have a good combination of quality and equity (OECD, 2019a,b,c). It has decentralised governance and allows schools and their owners, which in most cases are local governments, a high degree of autonomy. The state has set the national standards alongside principles of educational funding, system monitoring and quality assessment (Lees, 2016). The state is also responsible for developing the strategic framework for the Estonian education system, and it defines the required professional skills levels for educational staff and sets the minimum wage for teachers (Santiago et al., 2016). Local governments or municipalities are responsible for monitoring and implementing the national guidelines at local level. They also maintain the local school network of general education schools and pre-primary institutions. Operating within the guidelines provided by the state, schools and local governments have high levels of autonomy in terms of resource allocation, the applied curriculum and school development in general. School principals make decisions about hiring (and firing) teaching staff. They are responsible for developing a school curriculum based on the national curriculum framework and, together with teachers, decide on teaching methodology and materials. The division of responsibilities between the state, local government or municipality and schools is clearly defined (Basic Schools and Upper Secondary Schools Act, 2010⁽²¹⁾).

Educational institutes in Estonia can be state, municipal or private. State schools are financed by the state budget, whereas municipal schools receive funding from the local municipality budget. Private education is regulated by the Private Schools Act (1998)⁽²²⁾, and it has become slightly more popular in recent years. Currently, it comprises 11 % of schools and is funded by private organisations (Haridussilm, 2021). Private schools often cover niche forms of education that are not covered by the municipal or state schools, such as alternative or religious education. These schools can operate only with a valid licence granted by the Ministry of Education and Research.

In order to guarantee the constitutional right to free general education for all students, support from the state budget on equal grounds is allocated to both municipal schools and private schools to cover the expenses related to teachers' and head teachers' salaries and in-service training, textbooks and study materials, as well as school lunches for all students.

Estonia has an egalitarian education system. It follows the comprehensive school system, which aims to provide all students with the best education regardless of their background (Basic Schools and Upper Secondary Schools Act, 2010). According to PISA data, the quality of schools does not differ to a great extent, and this allows students to receive an equally good education regardless of their family background and the place where they live. Students most commonly attend the school closest to their home; however, they can also apply to attend private schools or other schools in another neighbourhood if they choose to and if there are vacant spots available.

⁽²¹⁾ <https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/519042021001/consolide>

⁽²²⁾ <https://www.riigiteataja.ee/en/eli/503062019009/consolide>

The Estonian education system comprises four levels.

- Preschool education (International Standard Classification of Education (ISCED) level 0) is for children up to 7 years of age. Attending a preschool is not compulsory; however, around 96 % of children attend kindergarten.
- Basic or integrated primary and lower secondary education (ISCED levels 1 and 2) is 9-year compulsory general education for all. Children start school at the age of 7 and usually attend the school closest to their home. The first streaming into academic or vocational tracks takes place after grade 9, when students are 15–16 years old. Successfully completing the curriculum and passing final examinations at the end of grade 9 is the condition for acquiring basic education.
- Upper secondary education (ISCED level 4) is based on basic education and divided into general upper secondary education and vocational secondary education. Around 25 % of students opt for the vocational track.
- Higher education (ISCED levels 6, 7 and 8) is provided by universities and applied higher education institutions.

The structure of the Estonian education system (see Table 10.1) provides opportunities for everyone to move from one level of education to the next.

Table 10.1: Education system in Estonia

Doctoral degrees (4 years)	
Master's degrees (2 years)	Integrated bachelor's and master's studies (5 years)
Bachelor's degrees or applied higher education (3 years)	
General upper secondary education Ages 16–19 (3 years)	Vocational secondary education (3 years)
Basic education (primary and lower secondary education) Ages 7–16 (9 years), compulsory	
Early childhood education (ages 3–7)	

Estonia has a population of 1.3 million, of whom approximately 69 % are Estonian, 25 % are Russian and 6 % are of other ethnic origins (Statistics Estonia, 2021). This division is also reflected in the education system, with the main language of instruction being either Estonian or Russian. In the 2020/2021 academic year, there were 521 general education schools attended by just over 155 000 students (Haridussilm, 2021). In Estonia, as in many other countries, the population is ageing, and this is reflected in the decreasing numbers of students. Optimisation of the school network has been part of educational processes in recent years – many schools have been closed, and others have been restructured or merged.

Compulsory education is organised with a view to supporting the development of every student, and schools are given flexibility in organising their educational processes. They

can decide themselves in which order certain topics are taught. For this reason, the state assesses the acquired competencies and the national curriculum outcomes of basic education acquisition by the stage of study. Grades 1–3 (ages 7–9) are in the first stage of study, grades 4–6 (ages 10–12) are in the second stage of study and grades 7–9 (ages 13–15) are in the third stage of study. This structure gives schools significant autonomy in organising studies depending on the needs of students and capacities of schools.

According to the national curriculum, assessment is an integral part of teaching and learning. Schools use both formative and summative assessments. During formative assessment, the teacher gives students oral or written feedback on achieved results and suggestions for improvement. The teacher inspires and directs students in setting their further goals and learning, involving parents, support specialists and other teachers. At least once a year the school organises an individual progress interview with every student, at which the student, the class teacher and a parent are present. During the interview, the participants agree on the aims of further learning and development (Basic Schools and Upper Secondary Schools Act, 2010). In the summative assessment, each student's skills are assessed on a scale from one to five, one being the lowest grade and five the highest, with three being the passing grade. Schools are also allowed to use assessment systems that differ from this one, and only verbal descriptive assessment may be used in grades 1–6.

The standard of basic education is determined by the national curriculum for basic schools ⁽²³⁾. It establishes basic values, learning outcomes, learning approaches, requirements for the learning environment, assessment principles and requirements for finishing school, as well as the syllabi of subjects.

The simplified national curriculum ⁽²⁴⁾ is meant for students with various levels of learning difficulties. According to the law, general education should follow the principles of inclusive education and be accessible to all (Basic Schools and Upper Secondary Schools Act, 2010). Inclusive education is understood as a person's basic right to high-quality education. The principles of inclusive education were defined at legislative level in 2013. Individual curricula and the study process are adapted in compliance with the students' needs.

What are the goals for basic education schools in Estonia? What qualities do we want to see in young people when they complete compulsory education? Estonian education strategy documents recognise the value of a self-managing lifelong learner. The basic education schools should create an age-appropriate secure learning environment that helps students in their development. They should develop students' intellectual curiosity and learning skills, self-reflection and critical-thinking abilities, and basic values. Students should understand the values underlying their actions and feel responsible for their consequences, and prepare themselves for further studies and lifelong learning (2021–2035 education development plan; Ministry of Education and Research, 2020).

⁽²³⁾ The 2014 national curricula are available on the website of the Ministry of Education and Research (<https://www.hm.ee/en/national-curricula-2014>). See also the 2021 edition of the national curriculum for basic schools on the government's website (<https://www.riigiteataja.ee/akt/123042021010>).

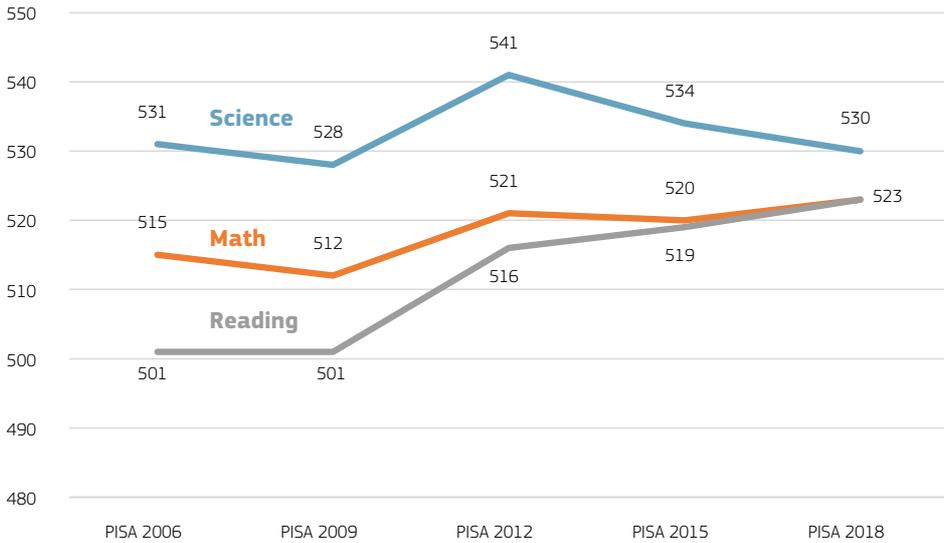
⁽²⁴⁾ <https://www.riigiteataja.ee/akt/120092011008>

International achievement surveys

Slightly more than a decade after the break away from the Soviet Union and a complete makeover of its education system, Estonia started to participate in international assessment programmes. The first international assessment in education in which Estonia participated was the Trends in International Mathematics and Science Study (TIMSS) in 2003. The results were unexpected by the Estonian public and education community. Estonian students ranked fifth in science and eighth in mathematics, which was the best result of the countries with a similar historical background. In science, Estonia ranked first in Europe, following the high-performing Asian countries. Estonia ranked third in Europe after Belgium (Flanders) and the Netherlands in mathematics. An interesting paradox was noted in the TIMSS 2003 data – even though Estonian students ranked 5th in science, their science teachers were very dissatisfied with their job and their students. The situation was rather similar with respect to mathematics teachers (Mere, 2004). Since Estonians by nature are never satisfied with their results (Kööts-Ausmees, 2018), the high performance in TIMSS was considered a ‘happy accident’. In addition, this survey was never repeated.

Estonia joined the Programme for International Student Assessment (PISA) in 2006. Since then, it has participated in all consecutive PISA cycles and showed an above-average performance in all domains. The high rankings in this survey have made the survey very popular in the country, and public interest and expectations are high prior to the release of each cycle’s PISA results. Estonians have experienced a positive ‘PISA shock’, and the participation in PISA has boosted their self-confidence and belief in their education system.

The Organisation for Economic Co-operation and Development (OECD) has noted that, in general, there has been very little or no improvement in student performance since the beginning of PISA among the OECD countries (OECD, 2019). Similar conclusions are stated about the European Union Member States, where the results have deteriorated in reading and science but have remained stable in mathematics (European Commission, 2019). Estonia has been a positive example, and as illustrated in Figure 10.1 the performance trend since 2006 has been positive in reading and mathematics, and stable (and high) in science. The observed decrease in science performance is not statistically significant.

Figure 10.1: Performance trends for Estonia

Source: OECD (2019a,b,c).

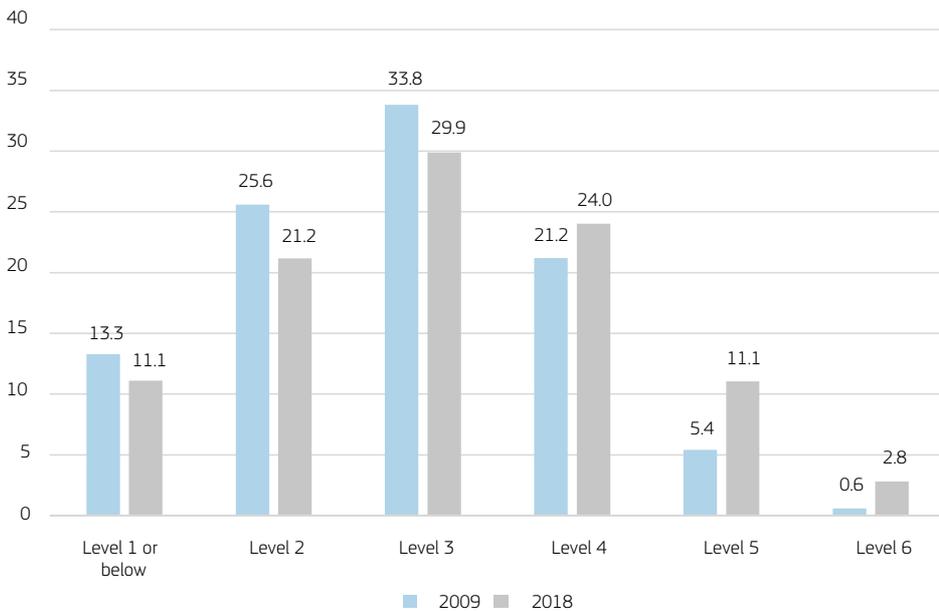
In PISA 2018, Estonia ranked first among the OECD countries in reading and science and was third in mathematics. The results noted small proportions of underachieving students and an increase in top performers. The Estonian education system also ranked high on equity, which means that students from different socioeconomic backgrounds have access to a good education and achieve strong results. The domain in which Estonian students have excelled in all PISA assessments is science. In PISA 2018, Estonia ranked fourth after the Chinese provinces, Singapore and Macau (China), with a score of 530 points, showing similar results to Japan (529 points). When looking at the distribution of student achievement according to levels of proficiency, we see that there is a relatively small proportion of underachieving students. In total, 8.8 % of Estonian students did not reach the baseline level of proficiency in science, and this is the lowest percentage among the EU Member States (the EU average is 21.6 %). The EU benchmark for low-achieving students is 15 %; however, only Estonia (8.8 %), Finland (12.9 %), Poland (13.8 %) and Slovenia (14.6 %) have achieved this goal.

There has never been a gender gap in science in Estonia; however, in 2018 we noted for the first time a statistically significant 5-point difference in favour of girls. If we look at the other end of performance, we see that the share of top performers (proficiency levels 5 and 6) was 14.2 % (the OECD mean is 7.6 %), and again for the first time we noted that the share of girls in comparison with boys at level 5 had increased.

The main domain of assessment in PISA 2018 was reading, and we can analyse trends by looking at PISA 2009, which also had reading literacy as the main domain. Estonia has shown a positive trend, and the mean score in reading increased from 502 points in PISA 2009 to 523 points in PISA 2018. The results were statistically similar to the scores

of Macau (China) (534 points), Hong Kong (China) (524 points), Canada (520 points), Finland (520 points) and Ireland (518 points). The improvement in reading in Estonia is mostly connected with the decrease in the number of low-performing students and the increase in the number of high-performing students. The data show that 89 % of Estonian students reached the baseline level of proficiency in reading (the OECD mean is 77 %). The corresponding number for the EU Member States was 78 %. The EU benchmark of 15 % for low-achieving students was achieved by Estonia (11.1 %), Ireland (11.8 %), Finland (13.3 %) and Poland (14.7 %). Between PISA 2009 and PISA 2018, the proportion of high-performing students in Estonia increased by almost 8 percentage points (see Figure 10.2). There was a significant gender gap in reading of 31 points in favour of girls (the OECD mean is 30 points). The gap was 44 points in PISA 2009, but it was reduced after the test moved to a computer-based mode of assessment in 2015.

Figure 10.2: Percentage of Estonian students at different proficiency levels in reading in PISA 2009 and PISA 2018



Source: OECD (2019a).

There has been a slight improvement in mathematics. With a mean score of 523 points, Estonian students ranked directly below high-performing Asian countries. This score was statistically similar to Japan, the Netherlands and South Korea. In Estonia, 10.2 % of students did not reach the baseline level of proficiency and have difficulties in solving the easiest mathematic items. There are only four countries among the EU Member States that have been able to keep the proportion of low performers at or under the benchmark of 15 % – Estonia (10.2 %), Denmark (14.6 %), Poland (14.7 %) and Finland (15 %) – with the EU mean being 22.4 %. The proportion of top performers in Estonia increased to 15.5 %, and boys outperformed girls by 9 points, which was a statistically significant difference.

Another aspect of education quality is equity, and PISA results repeatedly show that the Estonian education system manages to provide education for all students regardless of their socioeconomic background. In PISA 2018, 15.4 % of students from disadvantaged backgrounds belonged to the best-performing 25 % of students. In addition, 7.4 % of students with disadvantaged backgrounds (2.9 % in OECD countries) reached the top levels of performance. PISA classifies these students as ‘resilient students’. In Estonia, the performance gap in reading between students from high socioeconomic backgrounds and students from low socioeconomic backgrounds is 66 points, which is considerable but smaller than in other OECD countries (99 points). The mean score for Estonian students in the lowest socioeconomic quartile was 497 points, which is above the OECD mean of 487 points. This means that the poorest students in Estonia perform better than the top quarter in some other well-off OECD countries.

Along with PISA, the OECD also administers the Teaching and Learning International Survey (TALIS), which examines teachers’ and school principals’ perspectives. TALIS is administered every 5 years, and Estonia has participated three times. The survey tracks teacher education and professional development, and provides an insight into practised teaching methods, classroom climate and job satisfaction. Since the Estonian teacher population is ageing (the mean age is 49 years, and 54 % of teachers are aged 50 years and above), the sustainability of the Estonian education system is an important issue (OECD, 2019d). Within the next decade, there will be an urgent need to renew the workforce. In general, Estonian teachers are well trained, and almost all of them participate in professional development activities. They spend their time mostly on teaching (86 % versus 78 % in OECD countries). A further 84 % of teachers (42 % in OECD countries) say they have the freedom to decide on teaching methods and materials, and they report having a positive classroom environment. The data also indicate that Estonian teachers’ job satisfaction has increased over time.

PISA measures 15-year-olds, who are about to finish compulsory education in Estonia, while the Programme for the International Assessment of Adult Competences (PIAAC) explores 16- to 65-year-old adult skills in literacy, numeracy and problem-solving in technology-rich environments. PIAAC data were collected in Estonia between August 2011 and April 2012. It was a unique data collection, as it was the first time adult skills used at work and home were measured using computers (OECD, 2013). Estonian results in literacy and numeracy were above average among the participating 24 countries. However, the outcomes in problem-solving in technology-rich environments were below average. It was also surprising to see the lack of confidence in computer use shown by many Estonian adults, including those with a higher level of education (Valk, 2013). The younger generation (aged 20–24 years) clearly managed the problem-solving tasks better than 60- to 65-year-olds, with a gap of 48 points. People from more densely populated areas and with higher education levels had better information-processing skills. More skilled adults were concentrated in the north and south of the country, and the least skilled were found in the north-eastern parts of Estonia.

In summary, PIAAC revealed that the information-processing skills of Estonian adults are competitive and they are ready to perform more complex work. There was a relatively small difference in skills among people with different educational and socioeconomic

backgrounds, as well as between genders. The survey showed that the labour market valued education more than skills, which is important for people with incomplete levels of education to know. Table 10.2 lists the main international surveys in education in which Estonia has participated.

Table 10.2: International surveys in which Estonia has participated

Year	Survey	Survey organised by
2003	TIMSS	IEA
2006	PISA	OECD
2008	TALIS	OECD
2009	PISA	OECD
2009	ICCS	IEA
2012	PISA	OECD
2012	PIAAC	OECD
2013	TALIS	OECD
2015	PISA	OECD
2016	ICCS	IEA
2018	PISA	OECD
2018	TALIS	OECD
2018	IELS	OECD

NB: ICCS, International Civic and Citizenship Study; IEA, International Association for the Evaluation of Educational Achievement; IELS, International Early Learning Study.

Education policy monitoring

The current legal framework for the Estonian education system was created and established in the 1990s (Lees, 2016), after the country regained independence from the Soviet Union in 1991. During the following decade, the system was completely reconstructed; education reforms were implemented with the aim of liberating education from the leftovers of Soviet ideology and establishing a modern education system. In 1987, the Estonian Teachers' Congress criticised the existing system and demanded an independent Estonian education system. This led to the development of a new curriculum free from Marxist-Leninist ideology. Its implementation started in 1988/1989, even before the regaining of independence. This was followed by the rewriting of textbooks and retraining of teachers, and the restructuring (and renaming) of educational institutions.

In the 1990s, the following main laws on education and national standards were adopted: the Republic of Estonia Education Act (1992) ⁽²⁵⁾, which provided the legal basis for the formation, functioning and development of the education system and outlined the

⁽²⁵⁾ <https://www.riigiteataja.ee/en/eli/ee/Riigikogu/act/517072020006/consolide>

principles and rights of equal opportunities for everybody; the Basic Schools and Upper Secondary Schools Act (1993); the national curricula for basic and upper secondary education (1996); and the Private Schools Act (1998). The establishment of an external evaluation system took place in 1997.

The national curriculum adopted in 1996 was an important milestone. Until 1996, teachers were provided with fairly detailed content to be taught in the subjects of the curriculum. After the reform, attention was turned towards learning outcomes: what students should know and do with what they have learned. The curriculum introduced the notion of competencies; it described knowledge, skills, attitudes and values that were to be mastered during the learning process. More attention was paid to the development of cross-curricular competencies, subject-based competencies and general competencies. The new approach received some resistance at first, but, in retrospect, we can say that this was a positive change. The next curriculum reform was implemented in 2002, and then in 2011. The reform in 2011 aimed to increase choice for schools and students. Curriculum development is a constant process that follows the demands of the time and is updated on a regular basis. The curriculum states the learning outcomes and what students should master at different stages of their education. At the same time, the national curriculum is just a framework for schools; it leaves space for schools to modify the list of compulsory subjects and decide when and in what sequence they are taught.

Apart from the legal framework, there were several initiatives in Estonian society that contributed to the modernisation of education. In the 1990s, the Open Estonian Foundation (allied with the Soros Foundation) played a role in promoting free and independent schools. It promoted the idea that every school should 'have their own face' and that education reforms could be implemented only if schools accepted them and actively participated (Voolma, 1995).

Another project that left a strong impact on the education system as well as society in general was a 1996 state-funded initiative by the former president of Estonia, Toomas Hendrik Ilves, called the 'Tiger Leap' (Roonemaa, 2017). The goal of this initiative was to provide schools and public libraries with internet access and computer classrooms, teach citizens of all ages about technology and its use, and digitalise society. In 1997, almost 4 000 teachers completed courses in basic computer training, followed by many more in subsequent years. This overarching project was very successful and managed to achieve remarkable results. By 2001, all schools were equipped with computers and connected to the internet. Now, Estonia is among the most digitalised societies in the world, with numerous online services available to its citizens, including digital contracts, online voting and filing tax returns online.

The demand for digitally educated citizens is also reflected in the education system. Conscious investment in the development of digital competencies and infrastructure has been a high priority in state policies (Ministry of Education and Research, 2014). All schools have access to online platforms for school communication and administration, online digital materials, and testing systems. There are numerous state-funded training programmes and activities for teachers to learn how to apply different digital solutions in their lessons.

Improving the quality of education requires the use of relevant evidence and data. The state has developed a public education database called the Estonian Education Information System. It was started in 2004 and it is compulsory for all schools to use and update it. It contains data on all educational levels, and allows monitoring of many local and institutional processes such as student performance, funding and human resource management (Santiago et al., 2016). This database provides information about education establishments, students, teachers, curricula, learning outcomes, etc. It is an important tool that can be used with different access options for a wide variety of audiences. To ensure the security of the system, users must identify themselves through individual identity cards or in some other secure way. The Ministry of Education and Research has also made aggregated educational indicators available to the public on the Education Eye (Haridussilm) website ⁽²⁶⁾. These data are mainly used by parents, students, schools and local municipalities. They are also used for monitoring purposes, as school performance information can be found among them.

Most students go to state-funded public schools that are autonomous and enjoy a high degree of freedom in terms of how they organise their school culture and what learning methods they apply. To get feedback and monitor how well students have mastered the learning outcomes set in the national curricula, certain accountability mechanisms have been in use since 1997. Standard-determining national tests are administered at the beginning of years 4 and 7. They used to be paper-based tests, but with the digitalisation programme they have been implemented using a computer-based system in recent years. The tests are in Estonian (or Estonian as a second language) and mathematics, as well as in other subjects, such as science, foreign languages and social sciences. The tests are sample based and compulsory for 10 % of schools, but they are very popular, and schools find them a valuable tool of quality measurement and participate voluntarily. The results are used in improving the quality of teaching in the school. For example, if a school sees that its students perform less well than expected in mathematics, it will look for reasons, adjust the teaching and seek ways to improve. This is an important accountability tool that schools distribute widely to their stakeholders. These tests are low stakes for the students and are not graded (Education and Youth Board, 2021).

At the end of compulsory education in grade 9, students take three exams: one in Estonian (or Estonian as a second language), one in mathematics and one in a subject of their choice from a list of 10 subjects. Students are also expected to finish an independent research project. These are high-stakes exams and determine the completion of basic school. However, due to the COVID-19 disruptions in education, the laws have been temporarily changed, and the exam results will not affect the completion of basic school.

At the end of secondary education, students take three national examinations that are centrally provided and marked. These exams are also valid as entrance exams for higher educational establishments. Students must pass Estonian or Estonian as a second language, mathematics (two levels of exams are provided depending on the number of hours and depth of curriculum studied during upper secondary) and a foreign language. In addition, students are required to pass a school exam and do an independent research project. The aggregated school results in the national examinations can be seen in the

⁽²⁶⁾ <https://www.haridussilm.ee/ee>

Education Information System and are available to the public. The publication of the results creates yearly media interest, and the media ranks the schools according to the examination results. To show that, apart from national examinations, schools also have other quality indicators, the state has provided schools with value-added measurements determining how much each school has contributed to the development of its students (Education and Youth Board, 2021). This is a positive effort, but these measurements are more difficult for the public to understand; therefore, the interest in the national exam rankings and results still prevails.

In 2014, the Estonian government adopted the 2020 Estonian lifelong learning strategy, which has since become the main guiding document on the prevailing directions in education (Ministry of Education and Research, 2014). The document sets five areas of priority.

1. Change the approach to learning or focus on student-centred learning. Each learner should be provided with education that supports their individual and social development. This goal underlines the need to master appropriate learning skills, and foster creativity and entrepreneurship.
2. Empower competent and motivated teachers and school leadership by providing them with knowledge and skills on modern approaches and practices.
3. Align lifelong learning opportunities with the needs of the labour market.
4. Ensure a digital focus in lifelong learning.
5. Provide equal opportunities and increase participation in lifelong learning.

The strategy document describes the goals in more detail, and they are linked to indicators from national and international assessments. Substantial funding, provided jointly with the European Structural Funds, was granted to implement the strategy goals. As highlighted below, the international assessments PIAAC, TALIS and PISA were used in determining the indicators and setting the goals of the strategy.

In 2020, a new strategic document – the 2021–2035 education development plan (Ministry of Education and Research, 2020) – was developed and adopted, with the goals and priorities set until 2035. The new strategy is a continuation of the 2020 Estonian lifelong learning strategy (Ministry of Education and Research, 2014).

Analysis of policies, practices and political reforms

International surveys have affected Estonian education in a positive way. The results have repeatedly confirmed that the education policies implemented in the 1990s and updated along the way have been effective. Estonian policymakers have used international assessment data for benchmarking purposes but have not changed policies with the aim of improving PISA results. All changes are aimed at improving and advancing student learning following the demands of a constantly challenging and changing world. The Estonian education system has strongly followed the principles of equity by trying to provide the best learning conditions for each student regardless of their socioeconomic background.

What policy implications arise from the international assessments? As one example, PIAAC data were closely examined after data collection in 2012, and various national reports were compiled by the Ministry of Education and Research. The input was used in refining the Estonian lifelong learning strategy in 2014, which subsequently became the main guiding document in education. The goal regarding motivated teachers and school principals referred to analyses from PIAAC data. In addition, the goal of aligning learning with the needs of the labour market was based on PIAAC, as the data showed that Estonian adults use less skills at work than they have on average. PIAAC also referred to the most vulnerable groups of the population with respect to lifelong learning participation (Valk, 2014).

Teachers are key actors in sustaining the education system, and the data from TALIS have been used to hear the needs and voices of teachers and school principals. TALIS 2018 data point out that 86 % of Estonian teachers are female, with an average age of 49. As already mentioned, the Estonian teacher population is ageing; there is an urgent need to renew the workforce, and the situation for some subjects is already quite critical. Estonian teachers are well qualified: 81 % of them have obtained all required qualifications. Through TALIS, Estonian teachers identified the need to receive additional training related to ICT skills, teaching in multicultural or multilingual settings, and teaching students with special educational needs. The need for ICT skills training is part of the educational strategy goals; it was noted well before the COVID-19 crisis, and teachers were supported through training provided by the state. Although the school lockdowns in March 2020 turned teaching and learning upside down, the common conclusion was that teachers in Estonia were more prepared for the new reality of online teaching and learning than their counterparts in some other countries where digital teaching had not been practised to the same extent (Education Estonia, 2020).

The five PISA cycles in which Estonia has participated have confirmed the positive impact of national curriculum reforms. The national curriculum determined core and cross-curricular competencies in 1996. Amendments in 2002 introduced problem-solving skills and socioemotional skills. Further updates in 2011 focused on teaching core competencies in all subjects. Since PISA assesses the application of competencies in real-life situations, students are probably required to think more broadly and reflect on what they can do with their knowledge.

Although overall performance of Estonian students in PISA is very positive, PISA results have identified a performance gap between students in Estonian and Russian language schools. The gap is 42 points in reading and science and 29 points in mathematics, favouring students in the Estonian language schools. The two groups are treated equally – they receive equal funding – but the differences persist. Although the average results of students in the Russian language schools are lower than those of their Estonian peers, they are above the OECD average. This PISA finding has created much discussion among policymakers and politicians about whether Estonia should maintain the education system with two languages or create a unilingual system to better integrate Estonian society.

As already mentioned, Estonia introduced a national external evaluation system in 1997 for state monitoring purposes. Since then, centrally administered assessments and examinations have been in place at several stages of schooling. Students are assessed in

grades 4 and 7 with low-stakes tests in mathematics, literacy and other subjects. The tests have always included open-response items and have been computer based since 2017. At the end of grade 9, students take high-stakes national examinations to finish basic school. The national model grade for PISA is grade 9, and the time of assessment is 1 month before the basic school final examinations.

One of the main Estonian education policy priorities has been to guarantee equal access to education for all students. This dates back to the 1990s when principles of social equity were implemented. All students are provided with free school meals and free textbooks. Schools must ensure the best learning environment for all their students. Students are supposed to get help if they experience learning difficulties, and grade repetition is rarely practised.

In 2013, several new amendments to the Basic Schools and Upper Secondary Schools Act were adopted. One of the goals was to improve the availability of support services in schools, such as social pedagogy, psychology and speech therapy. In 2014, the state established a nationwide specialist network called Pathfinder (*Rajaleidja*) from which students with learning, psychological or behavioural difficulties could get help. All schools can receive additional support from these regional services for their students who need them. These support systems for struggling students may have contributed to the decrease in the proportion of low-performing students, as observed in PISA.

PISA data have been used in setting national key indicators for the general education development plans. The 2007–2013 education development plan included the goals of decreasing the shares of low-performing students and increasing the shares of top performers in all three assessment domains of PISA. PISA data were also used in developing the 2020 lifelong learning strategy. For example, the key indicators on top performers were taken from PISA. In 2014, the goal of having 10 % of students among the top performers in reading was set; this goal was achieved in PISA 2018 when 13.9 % of students reached the top levels of proficiency. The goals for maths and science were more ambitious but were not quite achieved. The goal in maths was to have 16 % of students among the top performers, and 15.5 % achieved this standard. In science, the goal was to aim for 14.4 %, and 12.2 % of students achieved the highest levels of proficiency. The goal of increasing the share of top-performing students has also been added to the new education strategy for 2035.

Changes in education policies and practices are often determined by unexpected circumstances. The COVID-19 pandemic at the beginning of 2020 completely changed methods of education and schooling. Although the 2020 Estonian lifelong learning strategy had a gradual plan for the implementation of digital learning tools and assessments, in reality schools and teachers were forced to start remote learning and become digital overnight. What are the short- and long-term consequences of this world crisis? Nobody knows. Nevertheless, in response to the COVID-19 crisis, the Estonian government has just announced a new priority in education. For the next 4 years, funds and effort will be invested to address and minimise the learning gaps caused by school closures. More digital tools to diagnose the learning gaps and support teaching and learning will be developed, and more attention will be paid to every student to monitor their knowledge as well as their socioemotional well-being. Policymakers have announced their readiness to support anyone who needs support as a result of the unprecedented situation in the world.

REFERENCES

- Education and Youth Board (2021), 'Tasemetööd' (<https://www.harno.ee/tasemetood>).
- Education Estonia (2020), 'What helped Estonian schools successfully handle the corona crisis?', 25 August (<https://www.educationestonia.org/what-helped-estonian-schools-successfully-handle-the-corona-crisis/>).
- European Commission (2019), *PISA 2018 and the EU: Striving for social fairness through education*, European Commission, Brussels (https://ec.europa.eu/education/sites/default/files/document-library-docs/pisa-2018-eu_1.pdf).
- Haridussilm (2021), 'Educational Statistics Database' (<https://www.haridussilm.ee/ee>).
- Kööts-Ausmees, L. (2018), 'Eestlase rahulolematusest ja rahust' (<https://www.sirp.ee/s1-artiklid/c21-teadus/eestlase-rahulolematusest-ja-rahust/>).
- Lees, M. (2016), 'Estonian education system 1990–2016: reforms and their impact' (http://4liberty.eu/wp-content/uploads/2016/08/Estonian-Education-System_1990-2016.pdf).
- Mere, K. (2004), *Rahvusvaheline matemaatika ja loodusainete võrdlusuring TIMSS 2003 ja Eesti tulemused* (<https://dspace.ut.ee/handle/10062/40798>).
- Ministry of Education and Research (2014), *The Estonian Lifelong Learning Strategy 2020*, Ministry of Education and Research, Tartu (https://www.hm.ee/sites/default/files/estonian_lifelong_strategy.pdf).
- Ministry of Education and Research (2020), *Haridusvaldkonna arengukava 2021–2035 (Education Development Plan 2021–2035)*, Ministry of Education and Research, Tartu (https://www.hm.ee/sites/default/files/eesti_haridusvaldkonna_arengukava_2035_seisuga_2020.03.27.pdf).
- OECD (2013), *OECD Skills Outlook 2013: First results from the survey of adult skills*, OECD Publishing, Paris ([https://www.oecd.org/skills/piaac/Skills volume 1 \(eng\)-full v12-eBook \(04 11 2013\).pdf](https://www.oecd.org/skills/piaac/Skills volume 1 (eng)-full v12-eBook (04 11 2013).pdf)).
- OECD (2019a), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (<https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>).
- OECD (2019b), *PISA 2018 Results (Volume II): Where all students can succeed*, OECD Publishing, Paris (<https://www.oecd.org/publications/pisa-2018-results-volume-ii-b5fd1b8f-en.htm>).
- OECD (2019c), *PISA 2018 Results (Volume III): What school life means for students' lives*, OECD Publishing, Paris (<https://www.oecd.org/publications/pisa-2018-results-volume-iii-acd78851-en.htm>).
- OECD (2019d), *TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners*, OECD Publishing, Paris, doi:10.1787/1d0bc92a-en.
- Roonemaa, M. (2017), 'Global lessons from Estonia's tech-savvy government', United Nations Educational, Scientific and Cultural Organization (<https://en.unesco.org/courier/2017-april-june/global-lessons-estonia-s-tech-savvy-government>).

- Santiago, P., Levitas, A., Radó, P. and Shewbridge, C. (2016), *OECD Reviews of School Resources: Estonia 2016*, OECD Publishing, Paris, doi:10.1787/9789264251731-en.
- Statistics Estonia (2021), 'Population' (<https://www.stat.ee/en/find-statistics/statistics-theme/population>).
- Valk, A. (2013), 'PIAAC and its meaning in Estonia' (https://www.hm.ee/sites/default/files/piaac_and_its_meaning_in_estonia.pdf).
- Voolma, H. (1995), *Omanäoline kool – Arenduse Kaudu Avatusele*, Open Estonia Foundation, Tallinn (<https://www.digar.ee/arhiiv/nlib-digar:130187>).

CHAPTER 11

Cross-national achievement surveys and educational monitoring in Slovakia

Katarína Zamborová and Zlatica Jursová Zacharová

Abstract

This chapter presents information on cross-national achievement surveys, along with their implications and recommendations for the transformation of the Slovak education system. The Slovak education system has been seeking reform that would enhance not only students' content knowledge but also their critical-thinking skills. The Slovak system has undergone several changes over the years, but the biggest one will likely happen in the upcoming years due to the new plans and intentions of the Ministry of Education, Science, Research and Sport. The chapter provides a brief overview of the structure of compulsory state education and basic information on the curricula and the Slovak national monitoring system, and compares this system with international assessments that have influenced educational policy and research in the past 15 years. We believe that changes in the next few years will improve public perceptions of Slovak education.

Insights into the Slovak compulsory school system

Despite the fact that Slovakia was established only on 1 January 1993, its history goes back to the 5th century when the ancestors of Slovaks started to gather in the middle Danube area. Historically, the Slovak educational system was connected to the Austro-Hungarian educational system and later to the Czechoslovakian educational system. The school system during the communist regime was strongly centralised; all schools had the same curriculum and performed the same teaching procedures; they had a normative and uniform character that did not respect the individuality of pupils and did not tolerate difference. Learning was based on memorisation, respecting authority and an emphasis on content knowledge. Despite that, the educational level of the nation was good, and there were few behavioural problems among pupils (Spilková et al., 2005). Teacher education was also at a high level.

It was not an easy path to democracy and humanism. However, a major transformation in the 1990s was humanistic ideas penetrating the Czechoslovakian school system, which opened the way to alternative schools (Bagalová, et. al., 2015; Matulčíková, 2007; Zelina, 2000), as well as private and church schools (Petlák, 2004). Later, under Act No 171/1990 Coll. ⁽²⁷⁾, the system of primary and secondary schools was formed, whose aim was decentralisation and the creation of new content meeting the needs of the 21st century.

According to the results of the Organisation for Economic Co-operation and Development (OECD) Programme for International Student Assessment (PISA) in 2015, the index of autonomy of Slovak education is above the OECD average, which suggests schools have relatively sufficient flexibility and autonomy (Hall et al. 2019). In just the past 10 years, Slovakia has become interested in inclusion (Vančíková, 2015), but it is still dealing with stereotypes and prejudice against difference among teachers (Hall et al., 2019; Jursová Zacharová et al., 2019), parents and the public.

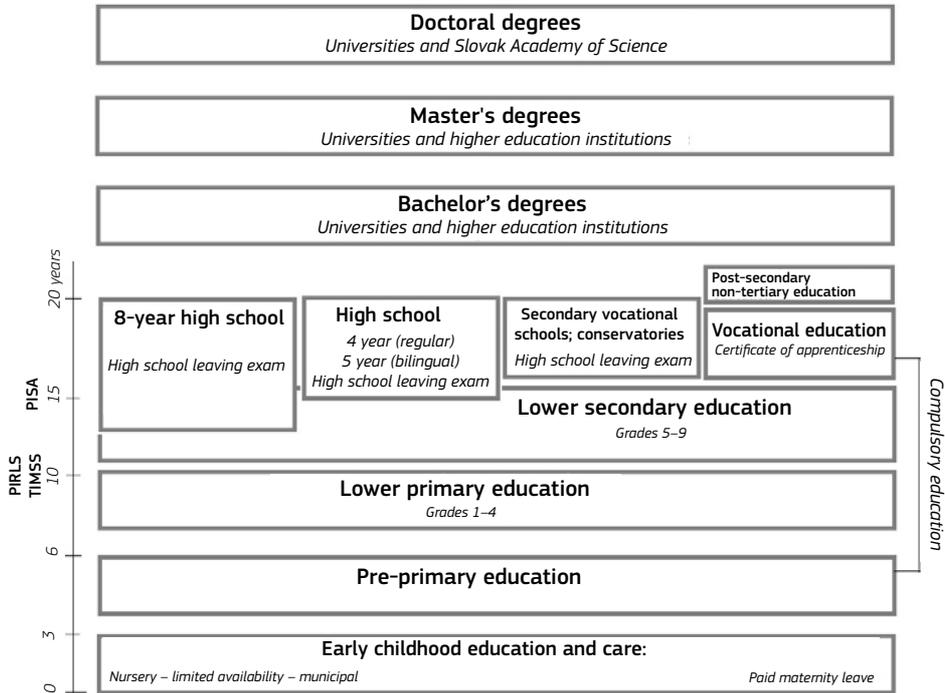
The Slovak compulsory educational system is guaranteed and free up to the age of 26, including higher education; however, parents need to cover extra expenses for food, before-/after-school clubs and/or school trips. The state provides textbooks for pupils. Part of teachers' dissatisfaction arises from the fact that they cannot work with textbooks they consider relevant. Often parents need to pay extra for alternative textbooks, especially for languages.

Generally speaking, the Slovak compulsory school system is split into two components: (i) the compulsory regional school system and (ii) higher education. Currently, the Slovak compulsory (regional) school system is divided into three levels. Pre-primary education occurs between 3 and 6 years of age, with the last year for children aged 5 years becoming compulsory in September 2021. Primary education (grades 1–4) and lower secondary education (grades 5–9) both belong under the umbrella of primary school education. The subjects in the first four grades are taught by class teachers, and in lower secondary education subjects are taught by a specialised teacher who has expertise in the

⁽²⁷⁾ Act modifying Act. No 29/1984 on the scheme for elementary and high schools (School Act).

given subject. After 9 years of education, students continue with 1 year of compulsory upper secondary education or until they reach the age of 16 years (high school, with specialised teachers) (see Figure 11.1).

Figure 11.1: The Slovak education system



Regional schools consist of state schools (without tuition fees), and private schools and church schools (which may have tuition fees). The system of schools also includes art schools, language schools and special schools for pupils with disabilities. Since 2008, based on Section 24 of Act No 245/2008 Coll. ⁽²⁸⁾, there has been the option of individual education (including homeschooling). Sometimes children are situated in special centres for individual education voluntarily or if they demonstrate inappropriate behaviour in a regular classroom, especially children with attention deficit hyperactivity disorder and similar diagnoses, including above-average intelligence. At the end of the school year, pupils who are being homeschooled have to pass specific tests at an equivalent level to each grade in regional schools. Pupils in the traditional school system do not need to pass such tests.

Every school in Slovakia is obliged to use the state educational programme, which is a binding document for the first stages (grades 1–4) of the elementary school system, combining specific requirements with the content of education for a given school, based on the regional and local conditions/needs of the students and parents. Based on the

⁽²⁸⁾ Act No 245/2008 Coll. on education and training (School Act) and on amendment of certain acts.

democratic and humanistic values of the society, the educational standards are targeted and performance related, and require high-level content. Through the use of educational fields and cross-sectional topics, there is an interdisciplinary approach to learning. The programme supports cognitive activities, such as active discovery, searching, investigating, finding out new things (such as outcomes of experiments), and interpreting or creating verbal, pictorial and graphic texts that reflect the social and cultural world. These activities are intended to:

- develop pupils' cognitive ability and motivation to study;
- strengthen pupils' self-reliance in solving problems connected with learning;
- involve pupils in active citizenship;
- develop moral values (ŠPÚ, 2015a).

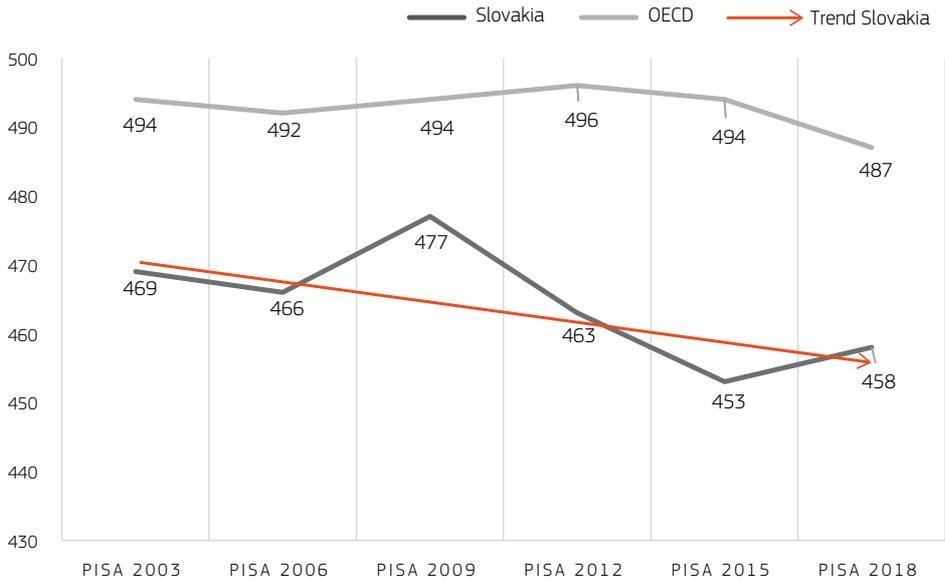
The main goals of elementary education (grades 1–4) are to develop literacy and key competencies. This happens through self-discovery and exploration of the immediate social, natural and cultural environments. The pupils are encouraged to find effective ways of self-teaching. The emphasis is on developing pupils' cognitive abilities through problem-solving independently as well as group collaborations. Social competencies with a responsible and active approach are built into the lessons (ŠPÚ, 2015a). The principles of grades 5–9 of lower secondary education repeat those of grades 1–4; however, the goals are aimed at functional literacy and critical thinking. Discovery and creating new meaning are supported, as well as students' sense of responsibility, interests and active involvement in the learning process. Effective strategies for learning, communication and cooperation are strongly emphasised (ŠPÚ, 2015b). Classes in lower elementary schools are divided into the following educational areas: (i) language and communication (Slovak language and literature, Slovak literature, minority languages and literature, and English language), in which students take four to five classes weekly; (ii) maths and work with information (i.e. informatics); (iii) man and nature (nature studies and natural science); (iv) man and society (natural sciences and homeland studies); (v) man and values (ethical and religious education – pupils can choose which they prefer); (vi) man and the world of work (technical education); (vii) art and culture (music and art education); and (viii) health and exercise (physical education; ŠPÚ, 2015a). Natural sciences, homeland studies and technical education become physics, chemistry, biology, history, geography and civics in lower secondary education (ŠPÚ, 2015b). However, this system should be replaced by 2024. In preparation for this change, the creation of new curriculum documents began in 2019 (ŠPÚ, n.d.).

The school year starts on 2 September and lasts until 30 June. It is divided into two terms; the first term is from 2 September until 31 January, and the second term is from 1 February until 30 June. Pupils are off during autumn break, Christmas, midterm break, spring break, Easter and summer break, and the dates are specified by the Ministry of Education, Science, Research and Sport (MŠVVaŠ SR) each year. The teaching activities generally start at 8 a.m. and the lessons last 45 minutes.

International achievement surveys

This section describes participation in and general performance patterns concerning international achievement surveys such as PISA, the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) since their initial implementation in Slovakia. In addition, the section offers explanations for Slovakia being below the OECD average across these assessments. The main organisation that ensures the testing of pupils' knowledge and the quality of education in Slovakia is the National Institute for Certified Education Measurements (NÚCEM), which is under the MŠVVaŠ SR. NÚCEM oversees the national assessments (Testing 5 and Testing 9) as well as the international assessments (i.e. TIMSS, PIRLS and, since 2003, mainly PISA, which concentrates on 15-year-olds attending elementary and secondary schools). Since 2013, International Computer and Information Literacy Study and International Civic and Citizenship Education Study assessments have been carried out. Both of these concentrate on pupils in the eighth grade.

In terms of the media, politics and research, the most observed international test of pupils during compulsory school education is PISA. This assessment has been carried out in Slovakia since 2003, and its results have provided important indicators regarding the need for a deep reform of Slovak education. Since the initial testing, Slovak students have been below the OECD average, especially in the field of reading literacy (Miklovičová and Valovič, 2019), in which Slovak students averaged 469 points in 2003. The trend in literacy scores in PISA has been declining in Slovakia since 2003 (see Figure 11.2); however, average literacy scores in PISA were slightly higher when literacy was the prime focus of PISA in 2009 and 2018, which may be related to teacher preparation that was more focused on reading literacy. PISA has also expanded the definition of reading literacy over time, and in 2009 the definition was updated to cover readers' use of reading and metacognitive strategies (Zápotočná et al., 2021). In 2018, the predominant tasks were assessing credibility and searching for information in several texts, and linking information from various sources (Miklovičová and Valovič, 2019; Zápotočná et al., 2021).

Figure 11.2: Reading literacy scores in PISA assessments

In general, the 2018 PISA report (Miklovičová and Valovič, 2019; OECD, 2018) highlights the poor results of students in the tested areas of reading, mathematics and science literacy, compared with the OECD average. Comparing the results of Slovak pupils, there has been a slight improvement in all years except 2009 and 2018. The largest decreases in the results of Slovak pupils compared with the OECD average were in 2015 in all areas: 41 points in reading literacy, 16 points in maths literacy and 32 points in science literacy. Trends in all areas show a decrease, although it is most significant in reading literacy and science literacy. Several professionals (see, for example, Zapotočná et al., 2015) have stated that the problem arose in the ability to find and obtain the necessary information from the text and the ability to integrate and interpret data.

In 2018, Slovak pupils attained better results in all areas than in 2015. Unfortunately, these results are still lower than in other years. Reading literacy and science literacy results have not returned to the levels seen before 2012. The differences between Slovak pupils and 15-year-old pupils from other countries in reading and science literacy are almost as much as one cognitive level (a difference of more than 50 points); in maths literacy, Slovak students' results are similar to the OECD average (except in 2015). Zápotočná et al. (2021) point out that, although the OECD average in individual areas is approximately balanced, in the case of Slovak students there is significant disproportion in favour of information retrieval and against the ability to think about and evaluate information obtained from the text. This problem may be caused by a reduced level of critical thinking on the part of students and their insufficient ability to synthesise information and use flexible, creative thought processes (Štefková et al., 2018). The problematic level of understanding arises mainly when students have to search for and combine information from text, pictures and tables.

An analysis of reading literacy predictors performed by Zápotočná et al. (2021) highlighted that, in interpreting the results of PISA, it is necessary to look at not only the predictors recommended by the OECD but also control variables describing school intent and pupil characteristics. These might include measures of pupils' economic and cultural backgrounds (such as access to books), pupils' 10-year expected employment status and motivation to read for pleasure. Last but not least, the differences between students are also explained by the year the student attends school, which is often reflected in the type of school the student attends – that is, whether he or she is in grade 9 at elementary school, or in the first year of high school or in the eighth grade at an 8-year high school. The biggest difference in PISA results observed by Zápotočná et al. (2021) and also Štefková et al. (2018) was between pupils attending primary school and pupils attending high school. There was also a clear points difference between students from secondary vocational schools without a high school diploma and students from some high schools. The difference between the students from vocational schools without a high school diploma and the students with a high school diploma equates to a difference of almost two cognitive levels, as defined by the PISA reference framework (Zápotočná et al., 2021).

Research (Štefková et al., 2018) suggests that pupils with a lower ability to comprehend a text have a problem with the elimination of irrelevant information in tasks concentrating on the short-term or working memory. The opposite effect can also happen. Children with d comprehension may fail in tests because they remember the words they are supposed to ignore. Another factor that influences reading comprehension is planning (the ability to understand which of the tasks are the most important to reach a goal). Accordingly, children with reading comprehension problems get lower results when it comes to planning and organising an answer. Furthermore, research suggests that cognitive flexibility, the ability to choose adaptively between representations of objects, perspectives or strategies, is a significant predictor of successful reading comprehension; however, it is not as strong as working memory, inhibition control or planning (Štefková et al., 2018).

Problems with comprehension of the text and lower reading literacy also affected mathematical and natural science literacy tasks. Although Slovak students' performance in mathematical literacy has oscillated around the average for OECD countries since the first assessment in 2003, there has been a noticeable decline in mathematical literacy. In 2003, 2009 and 2018, Slovak students reached the level of the OECD average, but a significant difference was observed in 2012 and 2015 compared with the OECD average (Plavčan, 2018). When monitoring the distribution of students according to the achieved PISA reference level, up to 25 % of students are at the lowest proficiency level (Slovak documents consider this level 'at risk'). A slight decrease in the number of Slovak pupils in the risk group over time or compared with the OECD average is not statistically significant, but in 2018 the percentage of top-level pupils increased significantly to 2.3 %. Most top-level mathematics students study at 8-year high schools (10.6 % compared with 0.8 % in primary schools), and most students belonging to the lowest group attend secondary vocational schools without a diploma (63 % in comparison with 34.6 % in primary schools and 1.6 % in 8-year high schools) (Miklovičová and Valovič, 2019). From 2003 to 2018, noticeably significant differences between the results of boys and the results of girls in mathematical literacy tests (with boys performing better) reduced to the level of an insignificant difference of only 4 points between boys and girls. Slovak students lag significantly behind the average for

OECD countries in terms of science literacy. Science literacy declined sharply after 2009, although, in the last assessment in 2018, the results increased slightly but insignificantly, while the average across OECD countries decreased (Miklovičová and Valovič, 2019; OECD, 2019a). Nevertheless, the difference in pupils' results in science literacy between the OECD average and Slovakia is 25 points, and up to 29 % of pupils found themselves in the risk group (compared with the OECD average of 22 %).

There is a huge difference in PISA scores between students in selected 8-year high schools and students in the other types of schools, who are most often in the top group in all areas of literacy. There is a large difference in the results based on socioeconomics, and this affects students in Slovakia significantly more than in the OECD countries on average (Hall et al., 2019). The socioeconomic impact in Slovakia is strongly connected with repetition of the school year. According to the findings of Hall et al. (2019), in Slovakia every seventh pupil from marginalised Roma communities repeats a year, and pupils from socially disadvantaged backgrounds have a repetition rate that is eight times higher than other pupils. It is interesting that the results for 4-year-old students (PIRLS and TIMSS) are more favourable for Slovakia, but a deterioration in results is noticeable among older pupils, especially 15-year-old pupils. It is possible to assume that the selection of pupils for selective schools may lead to a reduction in the overall level of literacy in the population.

Slovakia's involvement in PIRLS, which measures the literacy of 9- to 10-year-olds, showed that, in 2001, Slovak students' results were 18 points above the international average, but they were well below the level of the best countries (43 points). On average, Slovak students achieved the second cognitive level of assessment (Obrancová et al., 2004). In PIRLS 2006, the average 10-year-old student in Slovakia achieved a score of 531 points (31 points above the international average) – a statistically significant improvement on previous results (Mullis et al., 2007). However, the results were 2 points below the average for the European Union Member States and OECD countries. In Slovakia, only 8 % of students were included in the group of students with the best results, and 14 % of students were included in the group with the weakest results. Slovak students' reading for literary experience improved significantly, but there was no improvement in reading to obtain information (Ladányiová, 2007). Subsequently, in 2011 there was no statistically significant improvement in Slovak students' PIRLS results. An increase of 2 points was recorded, whereas the average for OECD countries increased by 5 points and the average for EU Member States increased by 1 point (Galádová et al., 2013). The percentage of students reaching the highest level remained at 8 %, which is the level of the international median but lower than percentages for OECD countries (10 %) and the EU (9 %). Testing shows that 4 % of pupils did not reach the lowest level of reading literacy – an increase of 2 % compared with 2006. A stagnant trend in PIRLS results was also found in 2016 (NÚCEM, n.d. a). As in other countries, girls performed significantly better than boys in all PIRLS tests (Galádová et al., 2013; Ladánziová, 2007; Obrancová et al., 2004).

In terms of the mathematical literacy of students, Slovakia wavers around the average of the TIMSS scale, whereas the success rate of 11-year-old Slovak students shows a slowly rising trend, although the results are uneven and fluctuate. In 2019, the number of points obtained in the field of mathematical literacy increased by 14 points (from 496 points to 510 points; NÚCEM, n.d. b). In the field of science literacy, the opposite occurred, although

in 2007 and 2011 the results of Slovak students (526 points and 532 points, respectively) were above the results of the OECD countries and EU Member States. In 2015, there was a drop of 12 points, and in 2019 the negative trend in science literacy did not reverse, which contrasts sharply with the growing trend in the OECD countries and EU Member States. In both areas, there is a significant difference in success between boys and girls, with boys achieving 10 points more than girls. Gradually, however, the number of students at the top level increased slightly, but it did not reach even half of the OECD average. In 2019, about 20 % of pupils reached only the lowest level, whereas in the OECD countries 24 % of pupils were at this lowest level. In the field of natural sciences, 7 % of pupils were placed in the top level, which is the same as the EU average and slightly below the OECD average (9 %), and 24 % of Slovak pupils were included in the lowest proficiency level, which is the same as the OECD average (NÚCEM, n.d. b).

Due to the PIRLS and PISA assessment results of Slovak students, conferences are regularly organised to improve the literacy of Slovak students, and teachers have organised training to improve reading and maths literacy. Despite intensive efforts, it has not been possible to significantly increase the students' success and reverse the negative trend (Zápotočná et al., 2021). An in-depth analysis of education problems has shown that one of the possible problems is the lack of and selection in early childhood care, and the relationship between early childhood education and care and success in international measurements (Hall et al., 2019). Other causes may be related to the traditional and long-established attitudes that persist in educational practice (Zápotočná, 2021) as well as in the social discourse on the need for education. Although the scientific community has long drawn attention to the need to develop competencies in pre-primary education (Zápotočná, 2007, 2010), in practice negative attitude in this regard (Petrová and Zápotočná, 2018). Another reason for the low level of reading literacy is rigid insistence on the correct technique of reading, speed and fluency of reading, and a mechanical approach to reading, which neglects comprehension of the text (Zápotočná, 2021). One of the significant problems is that any problems that arise in teaching reading are diagnosed in terms of learning disabilities (dyslexia) or attention deficit disorders, which justify a lower level of student competence and thus relieve the teacher of responsibility.

Pupala (2007) points out that, in education, we often encourage children to read literature for pleasure, but reading for information is not supported in the curriculum. When comparing the results of PIRLS, it is clear that Slovak students have a problem finding and sorting information, as these competencies are not explicitly supported in the curriculum (Pupala, 2007). Another possible reason for the overall reduction in reading literacy among students may be online reading, which has different requirements from reading books. Petrová (2021) points out that, when reading in the online space, patience for longer texts is lost and there is a tendency towards superficial fast scrolling of text, with people focusing more on images and videos than on text. At this point, it is possible to ask if the use of online reading literacy testing confuses students who quickly search and scroll through text in the digital world looking for fun elements, as they use the online world as a place of entertainment rather than a place to gather information (Petrová, 2021). Scrolling, actively searching for information, and analysing and comparing it requires students to have a good operational memory and the ability to inhibit distractors. According to Štefková et al. (2018), a lack of these two factors is key to student failure in the PISA tests.

Several studies on the results of PISA, PIRLS and TIMSS (Hall et al., 2019; NÚCEM, n.d. b; Miklovičová and Valovič, 2019; OECD 2019a,b,c; Zápotočná et al., 2021) point to a large difference in the results with respect to socioeconomic background. Significant differences in success rates between children with a high socioeconomic index score and children with a low socioeconomic index score appeared in all results, despite efforts to stop the expansion of the socioeconomic gap in Slovakia. This has led to differences among regions and children with lower performance remaining at elementary school (Habánik et al., 2016; Hall et al., 2019; Koršňáková, 2012).

Other reasons for the low level of reading literacy may be underpaid teachers and a lack of motivation (Hall et al. 2019), tests that do not correspond to the Slovak curriculum, and a lack of application of tasks encouraging higher levels of thinking (Zelina, 2019). In addition, Slovak students rank fifth from the bottom among observed countries on the desire for success in testing (NÚCEM, 2017; OECD, 2017). Furthermore, in the elementary school curriculum there is no meta-interpretation and evaluation of texts; this may explain why only high school students get into the top category of high proficiency (Somorová and Píšová, 2017).

Education policy monitoring

This section provides a description of reforms in Slovakia that have been planned but are yet to be implemented, and sheds light on the monitoring processes within the country. It includes a description of how achievement results are used to inform education policy development and reform for data-driven decision-making processes.

Education in Slovakia has undergone several transformational changes since 1989. One of the first steps was to establish private and church schools, the legal subjectivity of schools, and ways to finance primary and secondary schools (Act No 29/1984 Coll.). One of the first documents based on teacher innovation and the work of expert groups was *The Spirit of School* (Turek, 1990). Its aim was to create principles for the future Czechoslovak education system, expressing a new democratic and humanistic-oriented spirit, contrary to the previous totalitarian, administrative and bureaucratic-oriented spirit. The main position it took was to assume a creative climate. Its aim was also to revitalise lifelong education and compulsory secondary education, individualise the educational process and the need for self-teaching, ensure greater access to higher education and address the need to evaluate the quality of the educational process. It paved the way for the humanistic pedagogy developed following a model of creative and humanistic education in the work *Strategies and Methods of Development of Child Personality* (Zelina, 1996). Another strategic document was *Constantin* (Komárik et al., 1994), which promoted internationalisation and a liberal approach to education. Later, the Millennium project (Slovak government, 2001) was introduced by Vladislav Rosa, Ivan Turek and Miron Zelina. They proposed the concept of developing education for the next 15–20 years and considered humanism to be a way to improve the educational system in Slovakia. The outstanding principles of the programme are implementing a theory of creative and humanistic education, curricular transformation, professional development of the teacher,

raising financial subsidies and lifelong learning. This project is considered to be the most complex vision of school reform (Porubský, 2012).

In contrast to the humanistic approach, there is the conservative approach of the group from Trnava University that has devoted its time to educational reforms (Kaščák and Pupala, 2011; Kaščák and Žoldošová, 2007). This approach was first adopted by Pupala (2004) in *Narcissus in Education – Pedagogical contexts of individualism*. He highlighted the sociological view of pedagogy as a scientific theory and educative practice dependent on ideologies and policies as formative factors influencing social reality. The approach contrasted with the humanistic emphasis on a nurturing education. Furthermore, the Trnava University group has been discussing the reform of pre-primary education and the development of early literacy.

Later, as explained by Vladimír Crmoman (2018), a conceptual document was created – *Report on the state of education and on systematic steps to support its further development*. It was prepared under the pressure of teachers' strikes in 2013, when Dušan Čaplovič was Minister for Education, Science, Research and Sport. However, nothing came of the report after its submission (Crmoman, 2018).

One of the latest documents to reflect on the current situation and propose measures in education in Slovakia was *Učiace sa Slovensko (Learning Slovakia)*; Burjan et al., 2017). The document highlights the results of PISA 2015, in which the proportion of the weakest and most at-risk students increased. One of the proposed measures was a lifelong learning process with an emphasis on individualism and inclusive education and attention given to socially disadvantaged pupils, whose numbers had been on the rise. At the same time, as there were more students in the risk group (the lowest level of knowledge), there was a tendency for the number of pupils at the highest knowledge level to decline in comparison with the PISA 2012 results. Furthermore, the document reacts to the unsatisfying results in maths (PISA, TIMSS), and the authors suggest innovative curricular documents (such as the Hein method in Czechia), textbooks and methodological handbooks, as well as ensuring the adequate preparation of future teachers of maths and their participation in professional development. Looking at the PISA 2015 results, there are signs of failing to fulfil more of the educational goals. A potential reason behind this decrease in performance is of of interest, due to a reduction in inner motivation, a decrease in education quality and in the quality of the procedures teachers use, including for assessment. Here, the document suggests written evaluations that are more complex and reflect the whole picture of the students' work. Furthermore, the document, in general, welcomes international assessments and considers them a worthwhile financial investment through which Slovakia gains an unbiased external view of its school system compared with other developed countries. However, the PISA findings are not sufficiently analysed by the MŠVVaŠ SR, thus highlighting a need for secondary analysis. The new government did not accept the suggestions for reform, and, based on Hall et al.'s (2019) research on national testing, the current MŠVVaŠ SR suggests changes in education such as an open market for textbooks and increased digitalisation, which in fact occurred very quickly due to the COVID-19 pandemic situation (MŠVVaŠ SR, 2020). Several elementary schools have been experimenting with a new system of curricular documents since 2019 (ŠPÚ, n.d.). According to the plan of the new Minister for Education, Science, Research and Sport (MŠVVaŠ SR, 2020), there should be a reduction in

the amount of bureaucracy that teachers have to deal with, and there should be changes in the financing of schools. In addition, there was a proposal to create 'The Model School' as an innovative and evidence-based approach to education based on teamwork, solving problems in a group and discussions. The Model School supports a relaxing atmosphere for lectures and discussions. The MŠVVaŠ SR released EUR 3 million in an initial phase for the improvement of school education. The Model School will also function as a centre for teacher development, where teachers can acquire experience with innovative methods and approaches to teaching (MŠVVaŠ SR, 2021a).

All the changes mentioned paved the way for an evidence-based approach in education. One example is the use of national testing during compulsory school education, which takes place in the fifth grade (Testing 5) and the ninth grade (Testing 9) yearly (during the pandemic it took place online). The testing covers reading and maths literacy. For pupils educated in schools using a language other than Slovak, there are tests in English, Hungarian and Rusyn. Testing 5 has occurred since the 2012/2013 school year and Testing 9 since 2004/2005. Since 2008, the content knowledge and amount of information that the given pupils studied in the given year have varied in accordance with the state school programme. Testing 9 was extended in terms of reading competence by including the ability to gain information from the text to develop interpretation and evaluation as key competencies of pupils, as opposed to the previous testing of knowledge by memorisation. Maths tasks were extended through the use and application of maths knowledge in common life situations, as monitored by OECD PISA (Kanovská, 2009). Testing 9 showed that 80 % of pupils were capable of solving questions on maths and language requiring a very basic level knowledge. The ability to connect the learned knowledge and integrate it or interconnect more thought processes was mastered by 52 % of pupils; reading literacy from the perspective of information interpretation was mastered by 65 %. The highest level of maths literacy, and thus the ability to reflect on (speculate about) the processes and planning involved in solving the problem, was mastered by 47 % of pupils. Reading competence decreased significantly in the school year 2008/2009, and the highest level of reading competence, requiring the ability to ruminate and assess (the ability to think critically, reason using facts and find evidence outside the text) was reached by 29.5 % (Košinárová et al., 2009). Overall, pupils achieved a success rate of 55 % in mathematics and 59 % in reading literacy. A year later (2009/2010), the results were worse; 52 % of pupils achieved success in mathematical literacy and 57.7 % of pupils were successful in reading literacy (Košinárová, et al, 2010). Based on these tests, Testing 9 was introduced to all schools. Since 2011, the number of students tested in the last year of elementary education has been around 40 000. From 2010 until 2013, there was a slight growth in the ability of pupils in maths competence (up to 60 % in 2013) as well as in reading competence (67.5 %).

Since 2014, the tests have been adjusted and termed norm-referenced tests; these tests rank students according to performance relative to other students. The tests distinguish prepared pupils: those who achieve an average success rate of 50–60 % are considered to be prepared based on the norm-referenced tests. Only 54.7 % of pupils were successful in maths literacy tests and 62 % were successful in tests of reading and language literacy (Polgáryová et al., 2014). Currently, the aim of Testing 9 is to compare the performance of pupils and provide schools with feedback on pupils' success. The results also serve as a guideline for the entrance exam to secondary school. This leads to both schools and parents

putting pressure on pupils to perform well in Testing 9. Part of the final year of the ninth grade in primary school is devoted to test preparation. In 2019, there was an average success rate among pupils of 63.1 % in maths and 62.3 % in language literacy. Based on the results, it is clear that the biggest problems students had were with linear equation and fraction tasks. In the same year, some of the pupils took Testing 9 in online form, and data were compared. The following tasks were the most difficult for students: correct writing of numbers, writing of proper nouns and identification of nouns that appear only in the plural form (Khermová, 2019). Because of the pandemic, testing in 2020 was voluntary, and the option of an online system was provided.

In Slovakia, Testing 5 has sought to determine the performance of pupils entering the second level of elementary school, to verify their knowledge and ability to apply it in practice (Alföldyová et al., 2015). Similar to Testing 9, Testing 5 is compulsory. As Testing 5 is a norm-referenced test, it can compare pupils' performance. The expected success rate is 50–60 % in the test, and its reliability is higher than 0.8 (NÚCEM, n.d. c). When it comes to the content, the test is interconnected with the required knowledge following an innovative educational programme. Around 40 000–46 000 pupils take the test, and the test is provided in either Hungarian or Slovak in accordance with the teaching language of the school. In 2015, the average success rate in maths was 62 %, and in the Slovak language it was 66.6 % (NÚCEM, 2016). The results correlated significantly with the grades pupils were given in the end-of-school reports in grade 4. Differences between boys and girls were not significant, although boys had a higher success rate in maths but a lower success rate in the Slovak language. Over time, the success rate oscillated moderately, hovering around 63 %, with a considerable decrease to below 60 % in 2019 when pupils obtained similar average results to those in 2015. In spring 2021, pupils were educated remotely, and Testing 5 was implemented online. A total of 35 900 pupils voluntarily underwent the e-test (MŠVVaŠ SR, 2021b) to verify the knowledge they had gained at school.

School-level results for Testing 5 are publicly available, and parents are able to choose the best schools for their children based on the average success rates of pupils in individually tested competencies. Directors of schools and teachers consider testing an important means of strengthening the reputation of a school among the public. However, obtaining higher average results in the test does not necessarily mean that pupils can understand, evaluate and interpret a text on a meta level, as it is possible that they have been taught to memorise facts. Critics of school education (Pupala, 2007; Zelina, 2019) point out that memorising facts is less important than spending time analysing and interpreting text, which means using a higher level of critical thinking, such as analysis or interpretation.

Analysis of policies, practices and political reforms

This section provides information on why Slovakia is unable to sustain and create political reforms, and about future policy changes in Slovak education. Besides international monitoring, there has also been national testing in Slovakia. The results are analysed, described and evaluated; however, it seems that this has not helped to improve the below-average scores of Slovak pupils in international tests. Unfinished reforms and those proposed within a short

time frame may be connected with the fact that Slovakia has had 20 different ministers for education, science, research and sport since 1989, and only one of them stayed in the position for as long as 4 years. One can imagine that it has not been possible to sustain and create salient political reforms during this time of change. There have been several attempts, as seen above; however, real reform has never happened (Zajac, 2018). Zajac (2018) states that Slovakia is unable to do things that require persistence. Crmoman (2018) discusses and criticises the changes that have been made, such as the accepting of the legal subjectivity of schools in 1990, the multifunding of schools based on the number of students in 2003, the two-level model of education based on setting standards for each school in 2008, and the credit system for further professional development of teachers in 2009. In all these changes, we can find some positives (Maňčíková, 2007). However, as they were put into practice, these changes were widely criticised for several reasons, such as the underfunding of schools, the overloading of teachers, a lack of teaching positions, the need to chase for sufficient numbers of students for a school, excessive bureaucracy and a redundant credit system for the further professional development of teachers (Crmoman, 2018; Hall et al., 2019; MŠVVaŠ SR, 2019a). Therefore, policymaking has definitely had an impact on the results of international assessments through the instability of ministerial posts, the unsustainability of the government's acts, the underfunding of schools and, most importantly, the perception of teaching as an unattractive and underappreciated job (Hall et al., 2019).

However, when the results of international assessments are published, they catch the attention of the public through media channels and politicians' comparisons of Slovak pupils with other countries (mainly the results of PISA; other educational assessments are not covered by the media). Jurášek's 2020 analysis of PISA 2018 shows that the media depicted the results more negatively than positively, and they did not state the possible reasons or wider context (Jurášek, 2020). In addition, the assessment results came out before the parliamentary elections; therefore, they were closely monitored by politicians (Jurášek, 2020). Again, politicians did not consider measures for change; they simply compared the results with those of the OECD countries. Only the Minister for Education, Science, Research and Sport gave a few positive statements regarding significant improvement in maths, as Slovak pupils reached the OECD average. However, the opposition criticised the results and called for reforms, citing demotivated and underpaid teachers, insufficient numbers of quality textbooks, missing content, and unsuitable elementary and secondary schools. In addition, the Slovak research centre NÚCEM organises conferences and workshops on the results, which provide a better overview to teaching professionals; however, the general public is still not informed about these types of tests. The only reforming document that has taken into consideration the results of international assessments was *Učiace sa Slovensko (Learning Slovakia)* (Crmoman, 2018), as mentioned previously. According to Michalička (2016), the optimal outcome of the realisation of school reform is sufficient funding, which is emphasised by economists as the most effective investment. There has been inspiration from the Finnish education system and, recently, the Polish education system; however, we have to respect the historical context of the Slovak education system and its peculiarities and values in pursuit of humanism, democracy and progress based on science.

In addition, there has been a lively discussion in the past few years regarding compulsory foreign language study. With Act No 245/2008 Coll., the study of the English language starting in the third grade became the first compulsory foreign language learning

since 2011. It has led to many discussions among professionals (Lojová and Straková, 2011) and the public regarding the preparedness of English teachers who do not have the qualifications to teach English at primary level. Therefore, there have been many national projects and seminars for teachers to gain qualifications to teach English. Subsequently, due to a liberalisation of language learning, pupils could choose the foreign language they wanted to study first. However, English became compulsory for those who did not make it their first choice, starting in the seventh grade. English still prevails as the first choice of foreign language (ISTP, 2019; MŠVVaŠ SR, 2019b; Zamborová, 2021).

Another policy change occurred in September 2021 – Act No 209/2019 Coll.)²⁹ (completing Act No 245/2008 Coll. introduced compulsory pre-primary education for children aged 5 years. This act is in favour of developing pre-primary competencies in children and increasing their readiness for school. Thus, we may see a tendency towards prolonging compulsory school education, meaning that pupils will have more time to prepare for work after graduating from school. Therefore, they will be better equipped for the future. In 1989, there was a compulsory state nursery system for children from the age of 3 years; however, the system was changed, and now mothers can stay on paid maternity leave. Nursery school is considered a social service and is usually private.

Finally, based on the results of the international assessment surveys PISA, TIMSS and PIRLS, there has been discussion on approaches to educating children with educational disorders who are from backgrounds with low levels of stimulation, as the number of these children has been growing recently in Slovakia (Hall et al., 2019; Miklovičová and Valovič, 2018; Štefková et al. 2018; Zápotočná et al., 2021). This growth may have been caused by traditional teaching in elementary schools (Pupala, 2007). In several parts of Slovakia, pupils from minority groups as well as socially disadvantaged and culturally different groups attend schools but are not competent in the language of instruction (Jursová Zacharová et al., 2019). These children often need a specialised approach, such as special intentional language education. This is unfortunately not always provided because of a lack of teaching materials or a lack of teacher training. A recent research project (*To dá rozum (That makes sense)*; Hall et al., 2019; Jursová Zacharová et al., 2019; Zelina, 2020) found that teachers have a negative attitude towards difference in children. Currently, the Ministry of Education, Science, Research and Sport, and the National Institute for Education are trying to introduce requirements for inclusion in new policies that should be implemented in 2023/2024. We believe not only that these changes will help to improve Slovak pupils' reading literacy, maths literacy and natural science literacy but that they will ensure that they are better able to apply their knowledge in practice.

⁽²⁹⁾ Act of 27 June 2019 amending Act No 245/2008 Coll. on Education (School Act) and on amendments to certain acts, as amended, and amending certain acts.

REFERENCES

- Alföldyová, I., Khernová, V., Timárová, L. and Polgáryová, E. (eds) (2015), *Testovanie 5 – 2015 – Priebeh, výsledky a analýzy (Testing 5 – 2015 – Progress, results and analysis)*, National Institute for Certified Educational Measurements, Bratislava (https://www.nucem.sk/dl/2878/SPRAVA_T5_2015_na_zverejnenie.pdf).
- Bagalová, L., Bizíková, L., and Fatulová, Z. (2015), *Metodika podporujúca inkluzívne vzdelávanie v školách (Methodology Supporting Inclusive Education in Schools)*, National Institute for Education, Bratislava (https://www.statpedu.sk/files/articles/nove_dokumenty/projekty/eea_grants/metodika-podporujuca-inkluzivne-vzdelavanie-skolach.pdf).
- Burjan, V., Ftáčnik, M., Juráš, I., Vantuch, J., Višnovský E. and Vozár, L. (2017), *Učiace sa Slovensko (Learning Slovakia)* (https://www.minedu.sk/data/files/6987_uciace_sa_slovensko.pdf).
- Crnoman, V. (2018), '25 rokov slovenského školstva: reforma strieda reformu' ('25 years of Slovak education: reform alternates with reform'), *Denník*, 12 January (<https://dennikn.sk/996606/25-rokov-slovenskeho-skolstva-reforma-strieda-reformu/>).
- Galádová, A., Gallová, S., Katreniaková, E., Kelemen, Z. and Stovičková, J. (2013), *Trendy úrovne kľúčových kompetencií žiakov 4. ročníka základných škôl (Trends in the level of key competencies of pupils in the 4th grade of primary schools)*, National Institute for Certified Educational Measurements, Bratislava (https://www.nucem.sk/dl/3436/20131015_Klucove_kompetencie_web.pdf).
- Habánik J., Kordoš, M. and Hošťák, P. (2016), 'Competitiveness of Slovak economy and regional development policies', *Journal of International Studies*, Vol. 9, No 1, pp. 144–155, doi:10.14254/2071-8330.2016/9-1/10.
- Hall, R., Drál, P., Friedrichová, P., Hapalová, M., Lukáč, S., Miškolci J. and Vančíková, K. (2019), *Analýza zistení o stave školstva na Slovensku: To dá rozum (Analysis of findings on the state of education in Slovakia: That makes sense)*, MESA10, Bratislava (<https://analiza.todarozum.sk/analiza-zisteni-o-stave-skolstva-na-slovensku.pdf>).
- ISTP (Internetový sprievodca trhom práce) (2019), 'Od školského roka 2019/2021 dôjde k liberalizácii výučby jazykov' ('From the school year 2019/2021, language teaching will be liberalised'), 6 May (<https://www.istp.sk/clanok/14455/od-skolskeho-roka-2019-2020-d-jde-k-liberalizacii-vyucby-jazykov>).
- Jurášek, D. (2020), 'Výsledky testovania PISA 2018 – prehľad mediálneho a politického obrazu' ('PISA 2018 testing results – an overview of the media and political image'), *Juvenilia Paedagogica*, Faculty of Education, Trnava University, Trnava, pp. 58–65.
- Jursová Zacharová, Z., Lemešová, M., Miškolci, J., Cabanová, K., Horváthová, L. and Sokolová, L. (2019), *Postoje, inklúzia a predsudky v slovenských školách (Attitudes, Inclusion and Prejudice in Slovak Schools)*, Comenius University, Bratislava (https://www.havava.eu/accounts/Zlatica/Predsudky_inkluzia_monog_JZ_Final.pdf).

- Kanovská, R. (2009), 'Príhovor riaditeľky Národného ústavu certifikovaných meraní vzdelávania PhDr. Romany Kanovskej' ('Speech by the Director of the National Institute for Certified Educational Measurements, Dr. Romana Kanovská'), in Nogová, M. and Kvapil, R. (eds), *Stav a rozvoj funkčnej gramotnosti matematická a čitateľská gramotnosť – Zborník z konferencie (Status and development of functional literacy mathematical and reading literacy – Proceedings of the conference)*, pp. 10–11 (https://www.nucem.sk/dl/3213/zbornik%3C3%ADk_mcg_2009.pdf).
- Kaščák, O. and Pupala, B. (eds) (2011), *Školy v prúde reforiem (Schools in the Stream of Reforms)*, Renesans, Bratislava.
- Kaščák, O. and Žoldošová, K. (eds) (2007), *Námety na reformu počiatočného vzdelávania (Suggestions for the Reform of Initial Education)*, Renesans, Bratislava.
- Khernová, V. (ed.) (2019), *Testovanie 9 2019 – Priebeh, výsledky a analýzy (Testing 9 – 2019 – Progress, results and analysis)*, National Institute for Certified Educational Measurements, Bratislava (https://www.nucem.sk/dl/4558/Sprava_T9-2019_FINAL.pdf).
- Komárik, E., Rosa, V., Christenko, S., Ondrejko, P., Zelina, M., Majtán, M., et al. (1994), *Projekt Konštantín Národný program výchovy a vzdelávania – Štátna politika výchovy a vzdelávania v Slovenskej republike na obdobie 1995–2015 – 1 a 2 diel (Project Konstantin - National Education Programme – State policy of education in the Slovak Republic for the period 1995–2015 – 1 and 2 parts)*, MINEDU, Bratislava (<http://www.skvelarodina.sk/konstantin/>).
- Koršňáková, P. (2012), 'Výsledky medzinárodných štúdií ako spätná väzba vzdelávania na Slovensku' ('Results of international studies as feedback on education in Slovakia'), *Slovak Journal for Educational Sciences*, Vol. 3, No 2, pp. 72–99.
- Košinárová, T., Ringlerová, V., Laššová, E., Juščáková, Z., Hrbáček, M. and Pokrývková, K. (2009), *Matematická a čitateľská gramotnosť žiakov 9. ročníka základných škôl v školskom roku 2008/2009 – Výskumná správa (Mathematical and reading literacy of pupils in the 9th year of primary schools in the school year 2008/2009 – Research report)*, National Institute for Certified Educational Measurements, Bratislava (<https://www.nucem.sk/dl/3212/final.pdf>).
- Košinárová, T., Ringlerová, V., Kurajová, J., Laššová, E., Juščáková, Z., and Polgáryová, E. (2010), *Testovanie matematickej a čitateľskej gramotnosti – Akí úspešní boli žiaci 9. ročníka ZŠ v školskom roku 2009/2010? (Mathematical and Reading Literacy Testing – How successful were the pupils of the 9th grade of elementary school in the school year 2009/2010?)*, National Institute for Certified Educational Measurements, Bratislava (https://www.nucem.sk/dl/3211/Testovanie_M%3C4%8CG_2010.pdf).
- Ladányiová, E. (2007), *Čitateľská gramotnosť žiakov 4. ročníka ZŠ – Národná správa zo štúdie PIRLS 2006*, National Institute for Education, Bratislava (https://www.nucem.sk/dl/3443/%3C4%8Dit_gram_n%3C3%A1r_spr%3C3%A1v_PIRLS2006a.pdf).
- Lojová, G. and Straková, Z. (2011), *Teoretické východiská vyučovania angličtiny v primárnom vzdelávaní (Theoretical Basis for Teaching English in Primary Education)*, Comenius University, Bratislava.
- Mančíková, B. (2007), 'Reforma školstva alebo kozmetické zmeny' ('Education reform or cosmetic changes'), *TREND*, 4 December (<https://www.trend.sk/spravy/reforma-skolstva-alebo-kozmeticke-zmeny>).

- Matulčíková, M. (2007), *Reformnopedagogické a alternatívne školy a ich prínos pre reformu školy (Progressive and alternative schools and their contribution to the school reform)*, AG Musica Liturgica, Bratislava.
- Michalička, V. (2016), 'V očakávaní reformy školstva' ('In anticipation of education reform'), 15 August (<https://www.direktor.sk/sk/aktuality/v-ocakavani-reformy-skolstva.a-3346.html>).
- Miklovičová, J. and Valovič, J. (2019), Národná správa PISA 2018 (*PISA 2018 National Report*) (https://www.nucem.sk/dl/4636/Narodna_sprava_PISA_2018.pdf).
- MŠVVaŠ SR (Ministry of Education, Science, Research and Sport) (2019a), *Atraktivita učiteľského povolania (Attractiveness of the Teaching Profession)* (<https://www.minedu.sk/data/att/15376.pdf>).
- MŠVVaŠ SR (2019b), 'Najväčšou zmenou vo vzdelávaní bude v novom školskom roku uvoľnenie výučby cudzieho jazyka' ('The biggest change in education in the new school year will be loosening of foreign language teaching'), 26 August (<https://www.minedu.sk/najvacsou-zmenou-vo-vzdelavani-bude-v-novom-skolskom-roku-uvolnenie-vyucby-cudziedho-jazyka>).
- MŠVVaŠ SR (2020), 'B. Gröhling: Dost' bolo písania stostranových dokumentov, pracujeme na modernom školstve' ('B. Gröhling: There was enough writing of one hundred page documents, we are working on modern education'), 13 November (<https://www.minedu.sk/b-groehling-dost-bolo-pisania-stostranovych-dokumentov-pracujeme-na-modernom-skolstve/>).
- MŠVVaŠ SR (2021a), 'B. Gröhling: Modelová škola ako inšpirácia pre inovácie vo vzdelávaní' ('B. Gröhling: The Model School as an inspiration for innovation in education'), 17 March (<https://www.minedu.sk/b-groehling-modelova-skola-ako-inspiracia-pre-inovacie-vo-vzdelavani/>).
- MŠVVaŠ SR (2021b), 'V druhom polroku sa zapojilo do e-testovania vedomostí a zručností až 35 900 žiakov základných a stredných škôl' ('In the second half of the year, up to 35,900 primary and secondary school students took part in e-testing of knowledge and skills'), 4 June (<https://www.minedu.sk/v-druhom-polroku-sa-zapojilo-do-e-testovania-vedomosti-a-zrucnosti-az-35-900-ziakov-zakladnych-a-strednych-skol/>).
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L. A. and Sainsbury, M. (2007), *PIRLS 2006 Assessment Framework and Specifications – Progress in International Reading Literacy Study* (2nd ed.), Boston College, Chestnut Hill, MA (https://timssandpirls.bc.edu/PDF/PIRLS2006_international_report.pdf).
- NÚCEM (National Institute for Certified Educational Measurements) (2016), *Testovanie 5 – 2015: Priebeh, výsledky a analýzy. (Testing 5 – 2015: Progress, results and analysis)* (https://www.nucem.sk/dl/2878/SPRAVA_T5_2015_na_zverejnenie.pdf).
- NÚCEM (2017), *Životná pohoda a prosperita žiaka v štúdii PISA 2015 (Pupils' well-being and prosperity in the PISA 2015 study)* (https://www.nucem.sk/dl/3476/Pohoda_a_prosperita_%C5%BEiaka.pdf).
- NÚCEM (n.d. a), *PIRLS 2016 – Prvé výsledky medzinárodného výskum čitateľskej gramotnosti žiakov 4. ročníka ZŠ (PIRLS 2016 – The first results of the international research of reading literacy of pupils of the 4th grade of elementary school)* (https://www.nucem.sk/dl/3442/Prv%C3%A9_v%C3%BDsledky_Slovenska_v_%C5%A1t%C3%BAdii_IEA_PIRLS_2016.pdf).

- NÚCEM (n.d.b), *TIMSS 2019 – Prvé výsledky medzinárodného výskumu vedomostí a zručností žiakov 4. ročníka ZŠ v matematike a prírodných vedách (TIMSS 2019 – First results of international research of knowledge and skills of 4th grade elementary school students in mathematics and natural sciences)* (<https://www.nucem.sk/dl/4840/Prv%C3%A9%20v%C3%BDsledky%20Slovenska%20v%20C5%A1t%C3%BAadii%20TIMSS%202019.pdf>).
- NÚCEM (n.d. c), 'O Testovaní 5 – Základné informácie' ('About Testing 5 – basic information') (<https://www.nucem.sk/sk/merania/narodne-merania/testovanie-5/o-testovani-5>).
- Obrancková, E., Heldová, D., Lukačová, Z. and Sklenárová, I. (2004), *Čitateľská gramotnosť žiakov 4. ročníka ZŠ – Výsledky medzinárodnej štúdie PIRLS 2001 (Reading Literacy of 4th Grade Elementary School Students – Results of the international study PIRLS 2001)*, National Institute for Education, Bratislava (<https://www.nucem.sk/dl/3440/Publikacia.pdf>).
- OECD (Organisation for Economic Co-operation and Development) (2017), *PISA 2015 Results (Volume III): Students' well-being*, OECD Publishing, Paris, doi:10.1787/9789264273856-en.
- OECD (2018), *Education at a Glance 2018: OECD indicators*, OECD Publishing, Paris (<https://doi.org/10.1787/eag-2018-en>).
- OECD (2019a), *PISA 2018 Results (Volume I): What students know and can do*, OECD Publishing, Paris (<https://www.oecd.org/education/pisa-2018-results-volume-i-5f07c754-en.htm>).
- OECD (2019b), *PISA 2018 Results (Volume II): Where all students can succeed*, OECD Publishing, Paris (<https://www.oecd.org/publications/pisa-2018-results-volume-ii-b5fd1b8f-en.htm>).
- OECD (2019c), *PISA 2018 Results (Volume III): What school life means for students' lives*, OECD Publishing, Paris (<https://www.oecd.org/publications/pisa-2018-results-volume-iii-acd78851-en.htm>).
- Petlák, E. (2004), *Všeobecná didaktika (General Didactics)*, IRIS, Bratislava.
- Petrová, Z. (2021), 'Zmeny v čítaní v digitálnej dobe a ich dopad na testovanie čitateľskej gramotnosti' ('Changes in reading in the digital age and their impact on reading literacy testing'), in Zápotočná, O., Petrová, Z. and Urban, K. (eds), *Čitateľská gramotnosť – Výsledky slovenských žiakov v štúdiu PISA 2018 – Tematická správa (Reading Literacy – Results of Slovak pupils in the PISA 2018 study – Thematic report)*, National Institute for Certified Educational Measurements, Bratislava, pp. 62–70 (https://www.nucem.sk/dl/4867/tematicka_sprava_citatelska_gramotnost_PISA_2018.pdf).
- Petrová, Z. and Zápotočná, O. (2018), 'Early literacy education in preschool curriculum reforms: the case of post-communist Slovakia', *Global Education Review*, Vol. 5, No 2, pp. s145–s159.
- Plavčan, P. (2018). 'Matematická gramotnosť PISA 2006–2015' ('Literacy in mathematics PISA 2006–2015'), *Pedagogická revue*, No 3/2018, pp. 5–28.
- Polgáryová, E., Košinárová, T. and Mizerová, B. (eds) (2014), *Testovanie 9 – 2014: Priebeh, výsledky a analýzy (Testing 9 – 2014: Progress, results, and analysis)*, National Institute for Certified Educational Measurements, Bratislava (https://www.nucem.sk/dl/931/Sprava_T9_FINAL_2.pdf).
- Porubský, Š. (2012), *Školské reformy na Slovensku v medzinárodnom historickom kontexte (School reforms in Slovakia in the international historic context)*, Faculty of Education, Matej Bel University, Banská Bystrica, Slovakia.

- Pupala, B. (2004), *Narcis vo výchove: Pedagogické súvislosti individualizmu (Narcissus in Education: Pedagogical contexts of individualism)*, Typi Universitatis Tymaviensis a Veda, Bratislava.
- Pupala, B. (2007), 'Čítanie sa učí ako spartakiáda' ('Reading is taught like a spartakiad'), *Nové školstvo*, 29 January (<http://www.noveskolstvo.sk/article.php?86>).
- Slovak government (2001), *Národný program výchovy a vzdelávania v Slovenskej republike na najbližších 10–15 rokov (National programme of education in the Slovak Republic for the next 10–15 years)*, IRIS, Bratislava.
- Somorová, R. and Pišová, J. (2017), 'Čitateľská gramotnosť v pedagogických dokumentoch základnej a strednej školy – Štátny vzdelávací program – Vzdelávacie štandardy pre predmet slovenský jazyk a literatúra' ('Reading literacy in pedagogical documents of primary and secondary school – state educational program – educational standards for the subject Slovak language and literature'), In conference contribution 'PISA 2015 - Výzvy pre slovenské školstvo' ('PISA 2015 - Challenges for Slovak education') National Institute for Education, Bratislava (https://www.nucem.sk/dl/4522/02_%C5%A0P%C3%9A_-_%C4%8Citate%C4%BEsk%C3%A1_gramotnos%C5%A5_v_pedagogick%C3%BDch_dokumentoch_z%C3%A1kladnej_a_strednej_%C5%A1koly.pdf).
- Splíková, V., Dvořáková-Bloudková, M., Dvořák, D., Hausenblas, O., Havlíňová, M., Hejlová, H. and Tomková, A. (2005), *Proměny primárního vzdělávání v ČR (Transformations of Primary Education in Czechia)*, Portál, Prague.
- ŠPÚ (National Institute for Education) (2015a), *Štátny vzdelávací program Primárne vzdelávanie – 1. stupeň ZŠ (State Educational Program – Primary education – Lower elementary school)* (https://www.statpedu.sk/files/articles/dokumenty/inovovany-statny-vzdelavaci-program/svp_pv_2015.pdf).
- ŠPÚ (2015b), *Štátny vzdelávací program Nižšie stredné vzdelávanie – 2.stupeň ZŠ (State Educational Program – Lower secondary education)* (https://www.statpedu.sk/files/articles/dokumenty/inovovany-statny-vzdelavaci-program/svp_nsv_6_2_2015.pdf).
- ŠPÚ (n.d.), *Systémový model tvorby kurikulárnych dokumentov vrátane procesu ich periodickej inovácie 2019–2025 (System model of creating curricular documents, including the process of their periodic innovation 2019–2025)* (<https://www.statpedu.sk/files/sk/vzdelavanie/vyskumne-ulohy/systemovy-model-kurikularnej-transformacie/systemovy-model.pdf>).
- Štefková, M., Kundrátová, B., Belica I., Rajčániová, E., Hulín, M., Dočkal, V. and Perašínová, D. (2018), 'Vybrané psychologické faktory ovplyvňujúce čitateľskú gramotnosť žiakov pri riešení úloh PISA' ('Selected psychological factors influencing students' reading literacy in solving PISA's tasks'), *Psychológia a patopsychológia dieťaťa*, Vol. 52, Nos 3–4, pp. 244–258.
- Turek, I. (1990), *Didaktika technických predmetov (Didactics of Technical Subjects)*, SPN, Bratislava.
- Vančíková, K. (2015), 'Inkluzívne vzdelávanie slová a činy' ('Inclusive education of words and actions'), in Porubský, Š. and Lynch, Z. (eds), *Kurikulárna reforma a perspektívy základnej školy*, proceedings of a scientific conference held on 23–24 September 2015 in Banská Bystrica, Belianum.
- Zajac, P. (2018), 'Reforma školstva sa nikdy nezačala' ('Education reform has never begun'), *týždeň*, 1 September (<https://www.tyzden.sk/temy/50247/reforma-sa-nikdy-nezacala/>).

- Zamborová, K. (2021), 'Perspectives of native and non-native teachers of English', in Seresova K. (ed.), *Foreign Language Perspectives in University Education*, Ekonóm, Bratislava, pp. 143–144.
- Zápotočná, O. (2007), 'Prečo a ako v predškolskom vzdelávaní: Teoretické východiská súčasných prístupov a koncepcií rozvoja gramotnosti v predškolskom vzdelávaní' ('Why and how in preschool education: theoretical basis of current approaches and concepts of literacy development in pre-school education'), in Kaščák, O. and Žodošová, K. (eds), *Námety na reformu počiatočného vzdelávania*, Renesans, Bratislava, pp. 121–150.
- Zápotočná, O. (2010), 'Rozvoj počiatočnej literárnej gramotnosti' ('Development of initial literary literacy'), in Kollariková, Z. and Pupala, B. (eds), *Předškolní a primární pedagogika*, Portál, Prague, pp. 271–305.
- Zápotočná, O. (2021). 'Tradície a problematické stránky vyučovania čítania a písania v sloveskom vzdelávacom prostredí' ('Traditions and problematic aspects of teaching reading and writing in the Slovak educational environment'), in Zápotočná, O., Petrová, Z. and Urban, K. (eds), *Čitateľská gramotnosť – Výsledky slovenských žiakov v štúdiu PISA 2018 – Tematická správa (Reading Literacy – Results of Slovak pupils in the PISA 2018 study – Thematic report)*, National Institute for Certified Educational Measurements, Bratislava, pp. 43–51 (https://www.nucem.sk/dl/4867/tematicka_sprava_citateľska_gramotnost_PISA_2018.pdf).
- Zápotočná, O., Petrová, Z. and Urban, K. (eds) (2021), *Čitateľská gramotnosť – Výsledky slovenských žiakov v štúdiu PISA 2018 – Tematická správa (Reading Literacy – Results of Slovak pupils in the PISA 2018 study – Thematic report)*, National Institute for Certified Educational Measurements, Bratislava (https://www.nucem.sk/dl/4867/tematicka_sprava_citateľska_gramotnost_PISA_2018.pdf).
- Zelina, M. (1996), *Stratégie a metódy rozvoja osobnosti dieťaťa (Strategies and Methods of Development of Child Personality)*, IRIS, Bratislava.
- Zelina, M. (2000), *Alternatívne školstvo (Alternative Education)*, IRIS, Bratislava.
- Zelina, M. (2019), 'Taxonómie pre učenie, vyučovanie a meranie', *Studia Scientifica Facultatis Paedagogicae*, Vol. 1, No 18, pp. 11–19.
- Zelina, M. (2020), 'Interviews with teachers about inclusive education', *Acta Educationis Generalis*, Vol. 10, No 2, pp. 95–111.

Part III

CONCLUSION

CHAPTER 12

Cross-national achievement surveys and educational outcomes: a future path

Don A. Klinger, Sylke V. Schnepf and Louis Volante

Abstract

In providing a summary and analysis of the preceding 11 chapters, this chapter illustrates the diversity of responses within European Union Member States to the educational challenges and opportunities that the Member States have encountered as a result of cross-national achievement surveys. As described in this chapter, concerns regarding cross-national achievement surveys remain. Nevertheless, the brief summaries in this chapter also highlight the potential of these surveys to positively affect and influence education policy within the EU. One consistent theme is that these surveys have resulted in a much greater focus on educational equity for specific subpopulations of students within Member States. The analysis of the ongoing implementation and use of cross-national achievement surveys in the EU highlights that future editions of such surveys will probably need to find ways to be even more responsive to national efforts to positively affect the education and well-being of young people throughout the EU.

Introduction

There is little doubt that cross-national achievement surveys (international large-scale assessments) have become part of the educational ecosystem and that organisations such as the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) will continue to be dominant sources of such measures. Even the fiercest critics (see, for instance, Komatsu and Rappleye, 2021) have revised their focus, paying less attention to the oversimplification of the interpretation of international results and more to exploring ways to improve the use of such surveys to inform educational policy and practice.

Due to its multinational structure, the European Union provides a perfect setting to examine the relationship between policies, practices and political reforms, and cross-national achievement surveys. Given the principle of subsidiarity, EU Member States have the primary responsibility for education and training. Our coverage of different Member States therefore allows us to examine a wide variety of education systems embedded in diverse cultural and institutional settings. At the same time, Member States aim to achieve similar goals on educational outcomes and inclusion through different education system designs, as demonstrated by concerted actions within the European Education Area framework, which aims to enrich the quality and inclusiveness of education systems. It is the combination of similar goals on education and different ways of implementing education policy to achieve them that justifies a close look at how the results of education surveys shape policy design in EU Member States.

This volume covers a sample of 10 EU Member States whose educational policies and practices reflect variability not only in terms of achievement results in surveys such as the Programme for the International Assessment of Adult Competencies (PIAAC), the Progress in International Reading Literacy Study (PIRLS), the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) but also with respect to educational policy decisions, given the growing international influence of these surveys. Our intention in this final chapter is to first summarise the educational reform trends and trajectories in relation to these cross-national surveys. We then offer valuable insights into past and present policy discourses that have influenced educational policies and governance within the EU and surmise what future policy discourses may look like, nationally and across the EU.

Cross-national achievement surveys: a cautionary tale

In their introduction to this volume, Volante, Schnepf and Klinger (Chapter 1) provide a brief historical account of the rise of cross-national achievement surveys and their influence on educational policy. While acknowledging the past and ongoing criticism of these assessments (see, for example, Andrews et al., 2014; Goldstein, 2014; Komatsu and Rappleye, 2021), they highlight the work of the developers to address these criticisms from a measurement perspective and efforts to expand the educational domains of

focus. Their introduction to the 10 European examples covered in the chapters that follow illustrates the broad impact of these surveys, noting that the previous strategic framework for European cooperation in education and training and the current strategic framework towards a European Education Area include an educational target directly tied to PISA. The variability of educational policies enacted in the presence of these cross-national achievement surveys further demonstrates the importance of this volume and the examples set out in it for identifying the challenges to be addressed and informing future educational policies and practices.

We, and probably many other measurement and policy researchers, have often found ourselves serving as hopeful critics of large-scale achievement testing at both national and international levels. These tests and surveys garner much attention, especially given the international desire to compare the quality of education. 'Although survey results generally attract substantial attention, considerably less effort is spent on reflection on what kind of culturally comparable education skills are actually measured and how the data are generated' (Volante, Schnepf and Klinger, Chapter 1, p. 21). Such criticisms are hardly new (see, for example, Delandshere, 2001; Klinger and Rogers, 2011; Wilbrink, 1997) and – given the growing desire to measure increasingly vague educational constructs such as creative thinking, well-being and digital literacy – will probably not subside. The economic drivers, and arguably problematic economic modelling, underlying the implementation of PISA have fuelled much of the ongoing criticism of PISA and other cross-national assessments. As a simplistic example, based on some of the underlying economic predictive models, Estonia, with its increasing status in terms of cross-national achievement survey results (Tire and Värä, Chapter 10), should be an economic powerhouse in Europe. Admittedly, the growth in Estonia's gross domestic product is very positive, but the relationship between its PISA results and its economic growth remains tenuous (Komatsu and Rappleye, 2021).

Our criticisms of cross-national achievement surveys tend to be most aligned with those concerned about (i) the overuse of broad (blunt) achievement surveys to justify structural changes to educational systems or programmes, (ii) the ranking of educational systems, which leads to policy borrowing, and (iii) the use of these surveys to measure the effectiveness of educational initiatives, with little evidence that the surveys, being consistently cross-sectional in nature, are sensitive to such changes. Many of the authors of this volume have raised concerns with respect to the political impact of cross-national surveys, which can result in the redirection of educational resources or debate. Nevertheless, they also describe how their nations are supplementing data obtained from these broad cross-national surveys with more localised national data and information in order to tailor policy decisions to their specific national educational contexts, which are shaped by varying syllabi and cultural values relating to learning progress.

The chapters in this volume also highlight the reasons for each of the three concerns listed above. As one example, Finland – with its history of very strong cross-national achievement survey results – has not been a strong proponent of national testing, and many have argued for similar policies to support educational improvements in other countries. Yet Germany, Estonia, Italy and Sweden, to name a few examples, have

increased their use of national testing and, as Türe and Värä (Chapter 10) claim, these assessments have had very positive effects. Similarly, tracking has generally been considered a problematic educational practice that exacerbates inequalities (see, for example, Volante and Klinger, 2021), and it was central to the earlier criticisms of the German education system. Yet, as Timmermans, Naayer, Meelissen, Gubbels, Scheerens and van der Werf (Chapter 5) explain, the Netherlands has a relatively similar system of tracking, and student performance on cross-national assessments has been much higher, with less inequity. Collectively, these types of concerns underscore the difficulties that countries inevitably encounter when they attempt to draw general lessons from other high-performing nations.

Cross-national achievement surveys as a driver of change in education within the EU

The chapters in this volume provide a strong reminder that EU Member States differ greatly in their choices on education policies. While the variability in educational structures and systems is the most obvious example of substantial cross-national differences, the relative results on cross-national achievement surveys, and the variation in attention and responses to these surveys both initially and in subsequent years, further demonstrate these differences. Neighbouring countries (i.e. Finland and Sweden, Germany and the Netherlands) and those emerging from communism in the East (i.e. Estonia and Slovakia) have very different educational stories to tell.

While the IEA and its mathematics and science assessments (e.g. the First International Mathematics Study in 1964, the First International Science Study in 1971 and TIMSS 1995) have a longer history, it was the first administration by the OECD of PISA that seems to have greatly increased the influence of cross-national achievement surveys, especially in Europe. Subsequently, interest in other cross-national achievement surveys grew, resulting in an even greater number of such instruments to measure the educational outcomes of people at different stages of education and even beyond schooling (e.g. PIRLS, PIAAC).

Perhaps no other European country has experienced as strong a degree of influence of cross-national achievement surveys as Germany, the only country covered in this volume where a clear causal link between PISA results and policy design has been observed. The initial PISA 2000 results received an immense amount of attention in Germany, leading to the term 'PISA shock'. Prior to this time, there was a strong belief that the German educational system provided high-quality educational outcomes for its children. It was a belief that was shattered with the release of the 2000 PISA results. Subsequently, Germany has implemented a wide range of education policies successfully, improving its international achievement rank and decreasing educational inequalities, according to the results of later educational achievement survey rounds. This has made Germany an important country example for analyses of educational reform in the EU (see, for example, Sälzer and Prenzel, 2014; Volante, 2015; Volante and Ritzen, 2016). McElvany (Chapter 2) acknowledges that, while PISA and other cross-

national surveys now have much less direct influence in Germany, they continue to serve as an educational bellwether. Even with this declining influence, past and present efforts to address educational disparities, ongoing reforms to secondary schooling, and greater funds for educational research related to educational justice, societal cohesion and educational quality can all be easily traced to the results and findings of cross-national achievement surveys (Entorf and Davoli, 2019).

In contrast to the situation in Germany, the 2000 PISA results put Finland's educational system on a pedestal. Finnish students finished at or near the top across the major reading, mathematics and science literacy domains surveyed by PISA, results that were consistent in subsequent editions. Perhaps more importantly, small achievement gaps associated with family background led to the conclusion that education in Finland provided a high level of equity. The Finnish education system received international attention and became a model for jurisdictions looking at educational reform; it was made even more intriguing by some of its unique features, including a lack of standardised testing, later entry into school and the high level of education required to become an educator. Yet, even in the presence of such positive results, the attention on Finland identified underlying challenges. For example, underlying issues of equity were identified. Subsequent cross-national achievement results have shown a declining achievement trend (Harju-Luukkainen, Sulkunen and Maunula, Chapter 7) and, while the socioeconomic gap remains below the EU average (it is increasing), there are larger gender and immigration gaps in Finland. While the early PISA results provided a sound defence for the Finnish education system, Harju-Luukkainen, Sulkunen and Maunula (Chapter 7) state that the unexpectedly positive PISA results led to a loss of momentum for a planned policy development in the area of literacy education. More recent results from cross-national surveys in combination with internal research have given rise to renewed efforts to implement a series of reforms to reverse these negative achievement trends.

While Germany and Finland reflect two extreme early responses to cross-national achievement surveys, the other countries covered in this volume have had similarities in their responses, along with notable differences resulting from their contexts. As the influence of cross-national achievement surveys grew, the PISA shock first experienced in Germany reverberated across other EU Member States. Looney, O'Leary, Scully and Shiel (Chapter 8) argue that the relatively poor 2009 PISA results in Ireland probably accelerated the implementation of the national literacy and numeracy strategy and also resulted in a specific goal focused on improved performance in cross-national assessments. Similarly, Rönnerberg and Wikström (Chapter 6) describe the PISA shock that hit Sweden when the declining results for TIMSS and PIRLS in 2011 and, more importantly, the 2012 PISA results seem to have accelerated national efforts to improve the quality and equity of Sweden's education system.

In contrast, Tire and Värä (Chapter 10) describe a positive PISA shock experienced in Estonia, the effects of which continue today. The country has become a model of the potential of educational reform to improve the learning outcomes of children and, as some argue, the envy of the EU (Butrymowicz, 2016; Volante and Klinger, 2021). Estonia was a latecomer to participation in cross-national achievement surveys, and its initial

results were very positive, providing support for educational reforms that had occurred in the 1990s. The results of these surveys also found a very high level of equity in terms of achievement, with the gaps being much smaller than those found in other OECD countries and EU Member States. Interestingly, Estonia seems to be one of the few examples of a country that has fully embraced the use of these surveys and it now also uses the Teaching and Learning International Survey and PIAAC to identify and respond to ongoing teaching and learning challenges. Estonia continues to be an educational leader and has introduced a 4-year strategy to address educational inequalities caused by COVID-19.

While many of the chapters in this volume describe extreme examples of the impact of cross-national achievement surveys, each of the chapters highlights the attention that the survey results have received in different phases of the policy design process. For example, Slovakia has used the results from these surveys to improve teacher education and to challenge long-held beliefs about learning, beginning with early childhood education. The social inequities that affect Roma in particular were a key focus of Zamborová and Jursová Zacharová in Chapter 11, where they describe recent reforms that endeavour to encourage lifelong learning, 'with an emphasis on individualism and inclusive education and attention given to socially disadvantaged pupils' (p. 220). Even in countries with relatively stable cross-national achievement survey results, such as the Netherlands, small but steady declines have given rise to new initiatives to improve literacy, mathematics and science knowledge and skills (Timmermans, Naayer, Meelissen, Gubbels, Scheerens and van der Werf, Chapter 5), even though it is very difficult to measure the success of implementation at school level given the decentralised Dutch education system. The challenge for education reform drawing on education survey results, as highlighted by many of the chapter authors and in line with our earlier critique of these surveys, is that educational achievement surveys measure educational outcomes within broadly defined core learning areas. At one extreme, this can lead to general scepticism about the importance of educational achievement survey results given their limited potential to reflect knowledge of national curricula. On the other hand, educational achievement surveys can increase the focus at national level on educational outcomes in core areas, for example through the introduction of national tests or by using educational achievement survey results as national benchmarks, as discussed in the chapters on Ireland and Poland. While such policies can effectively draw attention to these core areas, they can also lead to 'teaching to the test' practices that may undermine the desired educational focus.

Each of the chapters in this volume describes past and present debates (likely to continue in future) regarding the accuracy of cross-national achievement surveys. At the same time, the analyses highlight a key contribution of such surveys. EU Member States are being forced to address issues and challenges relating to educational inequity, not solely in terms of future economic benefits but increasingly in terms of supporting children regardless of their socioeconomic background, migration status or ethnicity. These challenges have led to increased funding to support teaching and learning initiatives, and the redirection of educational resources to where they can have a greater positive impact. Perhaps even more valuably, EU nations are reviewing long-established educational structures to identify potential negative effects of their systems

on learning or on specific subpopulations of students. Among its initial responses to PISA, Germany introduced a national curriculum to be followed, developed nationwide monitoring of educational outcomes and revised its three-tier model of education, providing later entry into more academic pathways in some *Länder*. Poland decreased the age of entry into school, while also increasing the duration of mandatory schooling. Sweden revised its schooling structure to provide students with greater independence (Rönnerberg and Wikström, Chapter 6). Along with revisions to its three secondary school tracks, Italy introduced educational policies to enforce regulations on the duration of mandatory education to reduce dropout rates, albeit with only relatively minor success (Giancola and Salmieri, Chapter 4).

The use of cross-national achievement surveys in a political world

Throughout this volume, the authors have illustrated how the impacts of cross-national achievement surveys are moderated by changing political circumstances. As an example, French students' performance in cross-national achievement surveys has been close to or below OECD and EU averages, with mathematics being a particular concern, along with a substantial performance gap reflecting family background. Given the strong centralised governance model in France, led by the Ministry of National Education, targeted educational reforms would seem to have broader national appeal and a greater likelihood of being enacted. Yet this has not been the case (Volante and Klinger, 2021). As Normand and Gendron state in Chapter 3 (p. 63), 'the complexity of the French governance system does not help lead to the implementation of effective reforms'.

Clearly, France is not alone in terms of political barriers. Zamborová and Jursová Zacharová argue in Chapter 11 that, in spite of substantial educational reforms and policies enacted in Slovakia, political changes (20 education ministers in 30 years) and a lack of sustainability have made these reforms ineffectual. Poland serves as an unfortunate, albeit perhaps extreme, example of how a politically charged context can challenge and disrupt effective educational policy implementation (Jakubowski and Gajderowicz, Chapter 9). Jakubowski and Gajderowicz describe the substantial post-communist changes to education that began in 1999 and resulted in structural changes to schools that led to longer compulsory schooling, extensive curriculum changes and increased local autonomy. Further reforms in 2008 were intended to reduce educational inequalities. In spite of the continual positive trends in cross-national achievement survey results subsequent to these new reforms, Jakubowski and Gajderowicz describe a culture of resistance that has now resulted in a return to the educational structures of 1999. Poland provides an important cautionary tale regarding the power of communication and interest groups to shape or prevent educational reform in spite of international and national evidence supporting such reform.

The increasing use of national assessment programmes is in no small part a reaction to the inability of cross-national assessments to provide sufficient detail to inform national

policy directions. This in turn may be shaping some of the decisions of organisations such as the IEA and the OECD to explore the measurement of other educational constructs. The national testing programme in Sweden is now administered in years 3, 6 and 9, and has become compulsory, with higher stakes; the results from these tests have an explicit role in educational monitoring (Rönnerberg and Wikström, Chapter 6). Italy has witnessed a full roll-out of its national assessment programme over the past 10 years, after many years of implementation challenges (Giancola and Salmieri, Chapter 4). Giancola and Salmieri argue that this national assessment programme is the most consequential change arising in Italy from cross-national achievement surveys. As in Sweden, this national programme is closely linked to school evaluation and data-based accountability, with ‘the latter reflecting the pressure exerted by international surveys based on large-scale assessment’ (p. 84). Estonia, a country that is generally very rigorous in terms of administrative data collection across policy fields, had implemented national assessments long before the country became involved in cross-national achievement surveys, and Tire and Värä (Chapter 10) argue that these assessments have been used very positively to monitor school effectiveness and enhance educational practice. Slovakia has similarly embraced the use of national testing to provide evidence-based accountability and monitoring of schools, and also to provide feedback to individual students (Zamborová and Jursová Zacharová Chapter 11).

In contrast to the cross-sectional design of cross-national achievement surveys, national surveys can also make a very valuable contribution if conducted for different age cohorts. If student data are linked, longitudinal sample data on students are available. This makes it possible to follow up students over time and therefore allow an investigation of which education policies and teacher practices lead to greater learning progress. The need for longitudinal data for evidence-based policymaking has often been emphasised (see, for example, Araujo et al., 2017; Goldstein, 2017). Cross-national educational achievement surveys are just not fit for drawing specific policy implications on improvements to education systems, due to their cross-sectional design.

A future path

In spite of strong reservations about measuring cross-national achievement from educational achievement surveys, their results have identified important issues of inequity and achievement gaps in all EU Member States and beyond. The European benchmark aiming to decrease the share of low-performing PISA students stresses the importance of overcoming low achievement. Achievement inequities due to socioeconomic differences and immigration or cultural background continue to be a challenge for the EU, as in the rest of the world. There is little disagreement that successfully tackling inequalities has significant benefits. Countries with low inequalities tend to be those with, on average, higher-achieving children (Granato et al., 2022).

While the focus on inequalities is a very positive outcome of cross-national surveys, lessons that can be drawn from the results on ‘what works’ to decrease inequalities are less clear. This volume has illustrated the challenges and complexities that policymakers

face in using monitoring results from cross-national achievement surveys to instigate policy reform within specific national institutional settings. As highlighted in the volume, while these inequalities often manifest themselves as socioeconomic inequalities, they may also have cultural factors (e.g. Roma people in Slovakia), regional factors (e.g. southern Italy) or linguistic factors (e.g. Russian speakers in Estonia). In many situations, a complex combination of factors is likely to be at work. Given these complexities, and the time it takes to resolve these challenges, national governments are often placed in an untenable position as they struggle to defend substantial educational reforms while the intentions of those reforms are not fully realised in what the public perceives as a timely manner. These challenges become election issues, and new governments may be inclined to abandon policies of previous governments, citing the lack of progress to discredit the reforms and those who enacted them.

The future of national assessments and cross-national educational achievement surveys

It is important to note that the availability of educational achievement survey results and rankings of countries on their overall mean achievement has shifted the focus from the inputs to be provided for successful learning to the outcomes of the learning process, as discussed in the chapters on Germany, Ireland and Poland. It might well be argued that what matters is not how education systems perform on the average educational outcome as such. Rather, from a learning and teaching perspective, more importance might be attributed to how education systems manage to foster the learning progress that students have achieved. The focus on outcomes has the potential to incentivise the exclusion of students who are recent immigrants from participation in cross-national surveys, thereby decreasing the representativeness of the sample and creating positive bias in a country's educational achievement results (see Chapter 6 on Sweden). However, if it is learning progress that we are concerned about, then it is exactly a focus on lower achievers and how they are supported to perform better that we need. Exclusion of low performers from participation in education surveys prevents a focus on learning progress and the factors that can foster it.

The IEA and the OECD continue to work to further enhance their cross-national relevance, and it would seem that any positive future for educational achievement surveys would entail the collection of longitudinal data. The previous point links to the often stated reflection that without longitudinal data causal statements about 'what works' cannot be made. As Schnepf, Volante and Klinger argue (Chapter 1): 'As long as the existing educational achievement surveys do not employ a longitudinal design, that is, a design that follows the learning progress of students over time and hence includes the same students in every survey wave, education policy evaluation with educational achievement surveys, such as PISA, TIMSS and PIRLS, will remain limited at cross-national level' (p. 26). Such surveys not only make it possible to compare the added value provided by education systems across different countries but also allow the causal evaluation of policy implementation. Certainly, the long-term operational challenges of longitudinal designs cannot be easily resolved. The financial side might be easier to tackle if data collection

were well coordinated: one longitudinal survey can provide much more information than, as currently in many European countries, the results of three expensive cross-sectional educational achievement measures (PISA, TIMSS and PIRLS) for different cohorts and a similar point in time.

Furthermore, the pandemic has changed perceptions about the timeliness of cross-national achievement surveys. In the past, a survey conducted every 3 or 4 years seemed a reasonable choice, since education systems change slowly. As a consequence, changes in students' outcomes were unlikely to happen quickly either. The pandemic and physical school closures have had an unprecedented impact on children's learning outcomes in Europe (see, for example, Blasko et al., 2021; Engzell et al., 2020) within a very short period. The writing of this book occurred during the COVID-19 pandemic, which had been affecting global societies for close to 2 years at the time of publication. Yet we still do not have internationally comparable data available on children's educational outcomes during the pandemic (Blasko et al., 2021), and some national achievement surveys are likely to have made statistical or administrative adjustments, making it difficult to measure the educational impact of the crisis. PISA 2021, a survey that could have shed some light on such a critical policy-relevant problem, was postponed for a year. PIRLS 2021 is currently being conducted but covering only a much smaller sample of countries. As a result, policymakers will lack valuable data on the situation during the height of the pandemic, which could have helped in understanding the extent of learning loss and, more positively, which strategies were successful in maintaining previous performance levels for some children. This lack of data collection is an opportunity lost for the future increase in digital learning that Europe is currently planning for in its digital education action plan. Digital learning strategies, which will be implemented in European classrooms on a more regular basis in the future, should be accompanied by survey instruments to measure not only pupils' achievement but also their well-being and mental health. Some of these survey instruments could be coordinated cross-nationally, thus allowing timelier data collection if need be.

In addition to revising sample collection and ensuring timely cross-national data availability, future pathways include different approaches to operationalising educational outcomes. The chapters in this volume discuss cross-national surveys' focus on core measures of educational outcomes and the problems that this entails for national policy negotiation and policymaking (see, for example, Chapter 7 on Finland). This criticism is far from new. Probably as a response, both the IEA and the OECD have focused on expanding the definition of learning outcomes. Most notably, the IEA has recently introduced the International Computer and Information Literacy Study, and it is preparing to administer its Responses to Educational Disruption Survey and its 21st Century Skills Mapping Survey. The OECD is revising its 2024 PISA science assessment to include the topics 'socio-environmental systems and sustainability', 'the development of scientific knowledge and its misuse', and 'informatics' (OECD, 2020). These areas will include some knowledge previously encompassed in disciplinary subject areas. Finally, the OECD has long been interested in health and well-being outcomes, and it previously explored the possibility of incorporating the Health Behaviour of School-aged Children Survey into its survey programme. This could provide a greater focus on children's mental health, a topic that has been stigmatised and neglected over recent decades (UNICEF, 2021) but which could be very fruitful.

Given that all measures of educational outcomes can easily be contested and criticised, the bright future of any educational achievement survey includes a socially inclusive survey design process conducted with education stakeholders to ensure common consent on measurements implemented within surveys. On a continuum between survey organisers' and societies' interests in relation to measuring school performance, this could clearly lead to a shift closer to the societal side, thereby increasing not only surveys' legitimacy and transparency but also their importance in the context of national policy design.

REFERENCES

- Andrews, P., Atkinson, L., Ball, S. J., Barber, M., Beckett, L., Berardi, J. et al. (2014), 'OECD and PISA tests are damaging education worldwide – academics', *The Guardian*, 6 May (<https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics>).
- Araujo, L., Saltelli, A. and Schnepf, S. V. (2017), 'Do PISA data justify PISA-based education policy?', *International Journal of Comparative Education and Development*, Vol. 19, No 1, pp. 20–34.
- Blasko, Z., Costa, P. and Schnepf, S. V. (2021), 'Learning loss and educational inequalities in Europe: mapping the potential consequences of the COVID-19 crisis', *IZA Discussion Paper Series*, No 14298 (<https://ftp.iza.org/dp14298.pdf>).
- Butrymowicz, S. (2016), 'Is Estonia the new Finland?', *The Atlantic*, 23 June (<https://www.theatlantic.com/education/archive/2016/06/is-estonia-the-new-finland/488351/>).
- Delandshere, G. (2001), 'Implicit theories, unexamined assumptions and the status quo of educational assessment', *Assessment in Education*, Vol. 8, No 2, pp. 113–133.
- Engzell, P., Frey, A. and Verhagen, M. D. (2020), 'Learning loss due to school closures during the COVID-19 pandemic' (<https://osf.io/preprints/socarxiv/ve4z7/>).
- Entorf, H. and Davoli, M. (2019), 'Socioeconomic inequality and student outcomes in German schools', in Volante, L., Schnepf, S. V., Jerrim, J. and Klinger, D. (eds), *Socioeconomic Inequality and Student Outcomes – National trends, policies, and practices*, Springer, Singapore, pp. 63–80.
- Goldstein, H. (2014), 'Responses to Andreas Schleicher's reply to open letter', *Policy Futures in Education*, Vol. 12, No 7, pp. 880–882.
- Goldstein, H. (2017), 'Measurement and evaluation issues with PISA', in Volante, L. (ed.), *The PISA Effect on Global Educational Governance*, Routledge, Abingdon-on-Thames UK, pp. 49–58.
- Granato, S., Havari, E. and Schnepf, S. V. (2022), 'Learning and human development – educational achievement: its determinants and long-term effects', in Choi, Á. (ed.), *Economic Foundations of Education*, Bloomsbury, London.
- Klinger, D. A. and Rogers, W. T. (2011), 'Teachers' perceptions of large-scale assessment programs within low-stakes accountability frameworks', *International Journal of Testing*, Vol. 11, No 2, pp. 122–143.
- Komatsu, H. and Rapple, J. (2021), 'Rearticulating PISA', *Globalisation, Societies and Education*, Vol. 19, No 2, pp. 245–258.
- OECD (Organisation for Economic Co-operation and Development) (2020), *PISA 2024 – Strategic vision and direction for science* (<https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf>).
- Sälzer, C. and Prenzel, M. (2014), 'Looking back at five rounds of PISA: impacts on teaching and learning mathematics in Germany', *Solsko Polje (School Field Journal)*, Vol. 25, No 5/6, pp. 53–72.
- UNICEF (United Nations Children's Fund) (2021), *The State of the World's Children 2021 – On my mind: Promoting, protecting and caring for children's mental health*, UNICEF, New York (<https://www.unicef.org/reports/state-worlds-children-2021>).

- Volante, L. (2015), 'The impact of PISA on education governance: some insights from highly reactive policy contexts', *International Studies in Educational Administration*, Vol. 43, No 2, pp. 103–117.
- Volante, L. and Ritzen, J. (2016), 'The European Union, education governance, and international education surveys', *Policy Futures in Education*, Vol. 14, No 7, pp. 988–1004.
- Volante, L. and Klinger, D. A. (2021), 'PISA and education reform in Europe: cases of policy inertia, avoidance, and refraction', *European Education*, Vol. 53, No 1, pp. 45–56.
- Wilbrink, B. (1997), 'Assessment in historical perspective', *Studies in Educational Evaluation*, Vol. 23, No 1, pp. 31–48.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from: <https://op.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The official portal for European data (<https://data.europa.eu/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

